



HAL
open science

Sharing resources for enhanced distributed hypothesis testing

Mustapha Hamad

► **To cite this version:**

Mustapha Hamad. Sharing resources for enhanced distributed hypothesis testing. Information Theory [cs.IT]. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAT029 . tel-03827464

HAL Id: tel-03827464

<https://theses.hal.science/tel-03827464>

Submitted on 24 Oct 2022

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 : 2022IPPAT029

Thèse de doctorat



Sharing Resources for Enhanced Distributed Hypothesis Testing

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

École doctorale n^o626 Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat: Réseaux, Information et Communications

Thèse présentée et soutenue à Palaiseau, le 07 Juillet 2022, par

MUSTAPHA HAMAD

Composition du Jury :

M. Sheng YANG Professeur, CentraleSupélec, Université Paris-Saclay	Président
M. Michel KIEFFER Professeur, CentraleSupélec, Université Paris-Saclay	Rapporteur
M. Tobias KOCH Professeur des Universités, Universidad Carlos III de Madrid	Rapporteur
M. Chadi ABOU-RJEILY Professeur, Lebanese American University	Examineur
M. Samir PERLAZA Chargé de Recherche, INRIA	Examineur
M. Abdellatif ZAIDI Professeur Agrégé, Université Paris-Est Marne la Vallée	Examineur
Mme Michèle WIGGER Professeure, Télécom Paris	Directrice de thèse
Mme Mireille SARKISS Professeure Agrégée, Télécom SudParis	Co-directrice de thèse

“The greatest challenge to any thinker is stating the problem in a way that will allow a solution.”

Bertrand Russell

Acknowledgments

First of all, I want to express my sincere gratitude and appreciation for my forever supporting supervisors Professors Michèle Wigger and Mireille Sarkiss. Their support, help, and guidance on all the levels, from technical enlightening to personal encouragement, was outstanding. They inspired and motivated me to enrich my research experience and to achieve what I thought at the beginning impossible. I am really very honored and grateful that I had the opportunity to work under their supervision.

I want also to thank the jury members for their delightful and insightful comments, questions, and suggestions. They made my thesis defense a memorable experience and helped me in improving my work. In particular, I am thankful to Prof. Sheng Yang for chairing the defense and to Professors Michel Kieffer and Tobias Koch for reviewing and evaluating my work. In addition, I am thankful to Professors Chadi Abou-Rjeily and Abdellatif Zaidi, and to Mr. Samir Perlaza for examining my thesis. The presence of Prof. Chadi meant a lot to me as he is the main reason behind my entrance to the research world.

I would like to thank my best friends Mohamad Ibrahim and Hussein Basma who have been true brothers, and my colleagues and friends Nour, Ibrahim, Homa, Aymen, Mehrasa, Akram, Abdelaziz, Chadi, Lina, Natasha as well as many others who all added fun and companionship to my PhD journey.

My ultimate and infinite thanks go to my precious and phenomenal wife, soulmate, and partner in everything, Jihane. She has been tirelessly there with me through all ups and downs, and was never tired of providing all the love, care, and support. She fully deserves a "co-doctorate" with me and I could not imagine this journey and my whole life without her.

Above-all, I would like to specially thank my amazing parents Ismail and Samia and brothers Omar and Mohamad for all their love and support. To my parents, your extraordinary love, sacrifices, and prayers to me are the main reason for my happiness and success. To Omar, you have been a unique person in my life. Finally, I want to dedicate this achievement to a very special person who would have been very happy and proud to see it comes true, to the soul of my grandma Rasmiya.

Résumé

Les tests d'hypothèses distribués ont de nombreuses applications dans la sécurité, la surveillance de la santé, le contrôle automobile, la détection d'anomalies, où les centres de décision, à l'aide des capteurs distribués dans le système, doivent distinguer entre la situation normale (hypothèse nulle) et la situation d'alerte (hypothèse alternative). A chaque centre de décision, on distingue deux types d'erreur : l'erreur de type-I (correspondant à une fausse alarme) et l'erreur de type-II (correspondant à une détection manquée). L'objectif de ces problèmes est de maximiser la décroissance exponentielle des probabilités d'erreur de type-II tout en préservant les probabilités d'erreur de type-I en dessous de seuils donnés. Plusieurs travaux théoriques ont étudié les limites de performance de tels systèmes des tests d'hypothèses distribués, en se focalisant sur les tests contre l'indépendance où les mesures aux différents terminaux sont censées être corrélées en situation normale et indépendantes en situation d'alerte. Dans la plupart des travaux précédents, chaque système est supposé avoir des ressources réseau dédiées représentées par des contraintes de taux maximum sur les liens de communication. Dans cette thèse, nous supposons que différents systèmes ou applications partagent les ressources limitées du réseau et nous imposons des contraintes de taux plus flexibles (taux prévu).

Notre travail établit la première extension du problème de tests d'hypothèses binaires distribués sous des contraintes de taux prévu à des capteurs multiples et/ou des centres de décision multiples. Nous caractérisons de nouvelles limites fondamentales sur les exposants d'erreur de type-II aux différents centres de décision, sous réserve de contraintes fixes sur les probabilités d'erreur de type-I et de contraintes de taux prévu sur la communication entre les différents terminaux, pour diverses configurations de réseau. Nous considérons principalement un réseau avec deux capteurs coopérant entre eux et communiquant avec un seul centre de décision, un réseau avec plusieurs capteurs communiquant de manière multi-saut avec plusieurs centres de décision, et un réseau avec un seul capteur communiquant sur un lien commun et deux liens individuels avec deux centres de décision. Nous renforçons également certains résultats précédents sur les

configurations étudiées sous des contraintes de taux maximal en prouvant des résultats de converse forte.

Plus précisément, nous étendons le gain de taux $(1 - \epsilon)^{-1}$, qui a été démontré pour la première fois par Salehkalaibar et Wigger pour un seul lien, à tous les liens de communication lorsqu'un seul centre de décision est présent ou que plusieurs centres de décision ont des probabilités d'erreur de type-I admissibles égales, chaque fois qu'un seul terminal peut communiquer avec tous les autres terminaux. Nous montrons que ce gain est réalisable en multiplexant deux sous-schémas : un schéma dégénéré qui déclenche toujours une alerte en signalant des messages à un seul bit et un schéma de codage optimal pour la configuration sous des contraintes de taux maximal et des probabilités d'erreur de type-I qui disparaissent. Sous l'hypothèse mentionnée de contraintes d'erreur de type-I égales sur tous les centres de décision, nous montrons que les exposants optimaux d'erreur de type-II à tous les centres de décision peuvent être atteints en même temps.

Pour les configurations avec plusieurs centres de décision qui ont des probabilités d'erreur de type-I différentes, nous observons et caractérisons un compromis entre les exposants d'erreur de type-II réalisables aux différents centres de décision qui découle des différentes marges à exploiter sous les contraintes de taux prévu données par les différents seuils d'erreur de type-I aux divers centres de décision. Nous caractérisons l'ensemble de tous les exposants d'erreur de type-II réalisables aux différents centres de décision, que nous appelons la région des exposants d'erreur fondamentaux, pour ces configurations. Nous proposons également une nouvelle stratégie de multiplexage et de partage de taux qui permet d'atteindre les régions d'exposants fondamentaux dérivés. Cette stratégie n'est pas spécifique aux configurations étudiées et se généralise à n'importe quelle configuration avec des contraintes de taux prévu avec des gains prometteurs par rapport aux résultats sur la même configuration sous des contraintes de taux maximum. Pour de nombreux cas dans les configurations étudiées, nous simplifions davantage la caractérisation de la région des exposants fondamentaux, ce qui permet de réduire la complexité du schéma de codage optimal pour y parvenir.

Dans cette thèse, nous fournissons de nouvelles méthodes de preuve de converse que nous utilisons pour prouver de nouveaux résultats de converse pour la caractérisation de la région des exposants fondamentaux sous des contraintes de taux prévu et pour établir des résultats de converse forte pour les configurations sous des contraintes de taux maximum. Alors que certaines de nos méthodes de preuve sont basées sur des arguments de changement de mesure, des chaînes de Markov asymptotiques et le lemme du blowing-up, nous proposons une méthode de preuve plus simple utilisant uniquement des arguments de changement de mesure et des chaînes de Markov asymptotiques. Cette méthode s'avère être applicable à des problèmes plus larges que les tests d'hypothèses, tels que les converses fortes pour la compression avec et sans perte, avec information additionnelle au niveau du décodeur.

Abstract

Distributed hypothesis testing have many applications in security, health monitoring, automotive car control, anomaly detection, where the decision centers with the help of the distributed sensors in the system need to distinguish between the normal situation (null hypothesis) and the alert situation (alternative hypothesis). At each decision center (DC), two types of error are distinguished: type-I error (corresponding to a false alarm) and type-II error (corresponding to a missed detection). The objective of such problems is to maximize the exponential decay of the type-II error probabilities while preserving the type-I error probabilities below given thresholds. Various information theoretic works studied the performance limits of such distributed hypothesis testing systems with a focus on testing against independence where the measurements at different terminals are expected to be correlated under normal situation and to be independent under alert situation. In most of the previous works, each system is assumed to have dedicated network resources represented by maximum-rate constraints on the communication links. In this thesis, we assume that different systems or applications share the limited network resources and impose more flexible *expected-rate* constraints.

Our work establishes the first extension of the distributed binary hypothesis testing problem under expected-rate constraints to multiple sensors and/or multiple decision centers. We characterize new fundamental limits on the type-II error exponents at different decision centers subject to fixed constraints on the type-I error probabilities and to expected-rate constraints on the communication between the different terminals, for various network setups. Mainly, we consider a network with two cooperating sensors communicating with a single DC, a network with multiple sensors communicating in a multi-hop way with multiple DCs, and a network with a single sensor communicating over a common link and two individual links with two DCs. We further strengthen some previous results on the studied setups under maximum-rate constraints by proving strong converse results.

More specifically, we extend the $(1-\epsilon)^{-1}$ rate-boost, that was first shown by Salehkalaibar and Wigger for

a single link, to all communication links when a single DC is present or multiple DCs have equal permissible type-I error probabilities, whenever a single terminal can communicate with all the other terminals. We show that this gain is achievable by multiplexing two subschemes: a degenerate scheme that always raises an alert by signaling single-bit messages and an optimal coding scheme for the setup under maximum-rate constraints and vanishing type-I error probabilities. Under the mentioned assumption of equal type-I error constraints on all DCs, we show that the optimal type-II error exponents at all DCs can be achieved at the same time.

For setups with multiple DCs that have different permissible type-I error probabilities, we observe and characterize a tradeoff between the achievable type-II error exponents at the different DCs that stems from different margins to exploit under expected-rate constraints given by the different type-I error thresholds at the various DCs. We characterize the set of all achievable type-II error exponents at the different DCs, which we call fundamental error exponents region, for these setups. We further propose a new multiplexing and rate-sharing strategy that achieves the derived fundamental exponents regions. This strategy is not specific to the studied setups and generalizes to any setup with expected-rate constraints with promising gains compared to the results on the same setup under maximum-rate constraints. For many cases in the studied setups, we further simplify the characterization of the fundamental exponents region which yields reduced complexity in the optimal coding scheme to achieve it.

In this thesis, we provide new converse proof methods that we use to prove new converse results to the characterizations of the fundamental exponents region under expected-rate constraints and to establish strong converse results for the setups under maximum-rate constraints. While some of our proof methods are based on change of measure arguments, asymptotic Markov chains, and the blowing-up lemma, we propose a simpler proof method using change of measure arguments and asymptotic Markov chains only. This method turns out to be applicable for wider problems than hypothesis testing such as strong converses for lossless and lossy compression with side-information at the decoder.

Contents

Glossary	xiv
1 Introduction	1
1.1 Background and Motivation	1
1.2 Main Contributions	5
1.3 Organization of Thesis	7
1.4 Notation	8
2 Preliminaries: Single-Sensor Single-DC Networks	9
2.1 Introduction	9
2.2 State of the Art	9
2.3 The Setup with Noiseless Links of positive Rates	11
2.4 Maximum-Rate Constraint	13
2.4.1 Optimal Error Exponent	13
2.4.2 Optimal Coding Scheme	14
2.5 Expected-Rate Constraint	15
2.5.1 Optimal Error Exponent	15
2.5.2 Optimal Coding Scheme	15
2.6 Numerical Comparison	16
2.7 Summary and Discussion	17
3 Multi-Sensor Single-DC Networks	18
3.1 Introduction	18
3.2 Related Works	19

3.2.1	MAC	19
3.2.2	Two Hops	21
3.3	The Cooperative MAC Model	21
3.3.1	The Setup	21
3.3.2	Maximum-Rate Constraints	23
3.3.3	Expected-Rate Constraints	26
3.3.4	Numerical Comparison	28
3.4	Converse Proof to Theorem 5	30
3.4.1	Outline of the Converse Proof	30
3.4.2	Equivalent Characterization	30
3.4.3	Change of Measure, Blowing Up, and Asymptotic Markov Chain	32
3.4.4	Bounds on Optimal Error Exponent and Rate Constraints	33
3.5	The Two-Hop Model	36
3.5.1	The Setup	36
3.5.2	Maximum-Rate Constraints	38
3.5.3	Expected-Rate Constraints	41
3.6	Summary and Discussion	43
4	Multi-Hop Networks with Multiple DCs	44
4.1	Introduction	44
4.2	Related Works	44
4.3	The Two-Hop Model	45
4.3.1	The Setup	45
4.3.2	Maximum-Rate Constraints	47
4.3.3	Expected-Rate Constraints	47
4.3.4	Comparison between Maximum- and Expected-Rate Constraints	54
4.3.5	Converse Proof to Theorem 9	63
4.4	A system with K -hops	76
4.4.1	The Setup	77
4.4.2	Maximum-Rate Constraints	79
4.4.3	Expected-Rate Constraints	81

4.4.4	The Special Case $K = 3$	91
4.5	Summary and Discussion	96
5	Single Sensor to Multiple DCs	98
5.1	Introduction	98
5.2	Related Works	99
5.3	The Setup	100
5.4	Maximum-Rate Constraints	102
5.4.1	Optimal Coding Scheme under Maximum-Rate Constraints	102
5.4.2	The Exponents Region	104
5.4.3	Strong Converse Result	105
5.5	Expected-Rate Constraints	105
5.5.1	Optimal Coding Scheme under Expected Rate Constraints	106
5.5.2	The Exponents Region	108
5.6	Comparison between Maximum and Expected Rate Constraints	110
5.6.1	Numerical Simulation	111
5.7	Converse Proof to Theorem 16	111
5.8	Summary and Discussion	116
6	Summary and Outlook	118
6.1	Summary	118
6.2	Outlook on Interactive Communication	119
6.3	Outlook on Non-Cooperative MAC	119
6.4	Extensions to Noisy Communication Channel	120
6.5	Extensions to General Hypothesis Testing	120
6.6	Setups with Privacy and Security Constraints	121
6.7	Finite-Blocklength and Machine Learning	121
A	Proofs for Chapters 2 and 3	123
A.1	Proof of Lemma 1: Concavity and Monotonicity of the Function η_1	123
A.2	Converse to Theorem 3	124
A.3	Strong Converse Proof to Theorem 4	125

A.4	Analysis of the coding scheme in Subsection 3.3.3.1	127
A.5	Proof of Lemma 2	128
B	Proofs for Chapter 4	132
B.1	Alternative Strong Converse Proof to Theorem 8	132
B.2	Analysis of the coding scheme in Subsection 4.3.3.1.1 for $\epsilon_1 = \epsilon_2 = \epsilon$	134
B.3	Analysis of the coding scheme in Subsection 4.3.3.1.2 for $\epsilon_2 > \epsilon_1$	135
B.4	Analysis of the coding scheme in Subsection 4.3.3.1.3 for $\epsilon_1 > \epsilon_2$	137
B.5	Proof of Lemma 4	139
B.6	Strong Converse Proof of Theorem 11	143
B.7	Analysis of the coding scheme in Section 4.4.3.1	145
B.8	Proof of Lemma 6	146
B.9	Proof of Lemma 7	150
C	Proofs for Chapter 5	157
C.1	Analysis of the coding scheme in Subsection 5.5.1	157
C.2	Proof of Lemma 9	159
C.3	Strong Converse Proof to Theorem 15	161
C.4	Proof of Corollary 1	164

List of Figures

1.1	Different network structure with multiple sensors and/or multiple DCs.	4
2.1	Single sensor node and single DC node.	11
2.2	Point-to-point setup under maximum-rate constraint [1].	13
2.3	Structure of the optimal coding scheme for point-to-point setup under expected-rate constraint.	16
2.4	Optimal exponents under expected-rate and maximum-rate constraints for Example 1 when $\epsilon = 0.1$	17
3.1	Cooperative MAC with two sensors and a single decision center.	19
3.2	Two-hop setup with a single decision center.	19
3.3	Non-cooperative MAC setup with two sensors and a single decision center.	20
3.4	Superposition construction of the codebooks C_{U_1} and $C_{U_2}(m_1)$ with rates R_1 and R_2 respectively.	24
3.5	Optimal coding scheme for Cooperative MAC setup under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on all links.	27
3.6	Optimal exponents under expected-rate and maximum-rate constraints for Example 2 when $p_T = 0.8$ and $\epsilon = 0.05$	29
3.7	Optimal exponents under expected-rate constraints for Example 2 when $Y_1 = Y_0$, $p_T = 1$, and $\epsilon = 0.05$	30
3.8	Optimal coding scheme for two hops with a single decision center under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on all links.	42
4.1	Cascaded two-hop setup with two decision centers.	45

4.2	Optimal coding scheme for the two-hop network with two DCs of equal type-I error thresholds, $\epsilon_1 = \epsilon_2 = \epsilon$, under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on both links.	49
4.3	Optimal coding scheme for the two-hop network with two DCs of smaller type-I error threshold at the relay, $\epsilon_1 < \epsilon_2$, under expected-rate constraints using Han's scheme and the optimal two-hop scheme under maximum-rate constraints.	50
4.4	Optimal coding scheme for the two-hop network with two DCs of larger type-I error threshold at the relay, $\epsilon_1 > \epsilon_2$, under expected-rate constraints using two versions of the optimal two-hop scheme under maximum-rate constraints.	52
4.5	Exponents regions for Example 3 when $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$ and $R_1 = R_2 = 0.5$	58
4.6	Exponents regions under expected- and maximum-rate constraints for Example 3 when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R_1 = R_2 = 0.5$	59
4.7	Exponents regions under symmetric and asymmetric expected-rate constraints for Example 3 when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R = 1$	59
4.8	Functions $\eta_1(r)$ and $\eta_2(r)$ for Example 3.	61
4.9	Functions $\eta_1(r)$ and $\eta_2(r)$ for Example 3 when $p_{Y_0} = 0.6, p_S = 0.75$, and $p_T = 0.8$	61
4.10	Exponents regions under expected-rate constraints subject to different distributions of the total rate $R = R_1 + R_2 = 1$ for Example 3 when $p_{Y_0} = 0.6, p_S = 0.75, p_T = 0.8$ and $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$	62
4.11	Sketch of partitioning $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1})$ and applying parallel change of measure arguments in the converse proof to Theorem 9.	63
4.12	Sketch of the simplification in the converse proof to Theorem 9 when $\epsilon_1 > \epsilon_2$	64
4.13	Cascaded K -hop setup with K DCs.	77
4.14	Illustration of the general optimal coding scheme for the K -hop setup with K DCs under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.	82
4.15	Illustration of the conjectured simplified optimal coding scheme for the K-hop setup with K DCs under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.	87
4.16	Sketch of partitioning $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2})$ and applying parallel change of measure arguments in the converse proof to Theorem 12 when $K = 2$	88

5.1	Distributed hypothesis testing with a single sensor and two remote decision centers with integrated sensors.	101
5.2	Optimal coding scheme for the setup with a single-sensor and two decision centers under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.	107
5.3	Optimal error exponents regions of the BC setup under expected- and maximum-rate constraints for $R_0 = 0.1, \epsilon_1 = 0.15, \epsilon_2 = 0.05$	112

List of Abbreviations

IoT Internet of Things

DC Decision Center

pmf Probability mass function

DSBS Doubly Symmetric Binary Sources

MAC Multi Access Channel

DMC Discrete Memoryless Channel

BC Broadcast Channel

SHA Shimokawa-Han-Amari

LHS Left-hand side

RHS Right-hand side

FME Fourier-Motzkin Elimination

Chapter 1

Introduction

1.1 Background and Motivation

We live in a widely connected world where advanced technologies are integrating with all types of human daily life needs such as health, security, mobility, household tasks, and work-related tasks. In this world, future Internet of Things (IoT) systems are expected to be present ubiquitously with more than 10 billion already installed IoT devices and a total of 25 billion devices estimated to be IoT-connected by 2030, according to Statista. Accurate decision systems for health monitoring, security alerting, anomaly or incident detection, automotive car control, and other highly sensitive applications are among the main use cases of widely emerging IoT and sensor networks. With this tremendous number of connected IoT devices and sensors which have limited network and energy resources, the various decision systems are required to efficiently use and share the available resources to achieve the best possible accuracy in their decisions. This explains the interest of information theoretic works in deriving fundamental limits of the performance of these decision systems and to suggest optimal strategies for achieving these limits. These theoretical findings also serve as a benchmark to assess the performance of practical implementations and their margins to the fundamental limits.

Therefore, such future systems aim at maximizing the performance of decision systems with limited resources. These decision systems are equipped with sensors collecting data and decision centers wishing to accurately identify the occurring events based on the received information from the sensors. In Information theory, such a problem is known as hypothesis testing where the decision center has to decide on the joint distribution underlying the data observed at all terminals. Each event, called hypothesis in the following, induces a different underlying joint distribution of the observations. In this thesis, we focus on binary

hypothesis testing where the decision centers have to distinguish between two possible hypotheses (called null hypothesis and alternative hypothesis). Our main objective is to characterize the performance of the decision centers in terms of the error probabilities under the two hypotheses.

In alerting applications such as anomaly detection, security breach detection, health monitoring systems, and automotive car control systems, the null hypothesis usually corresponds to a normal situation, while the alternative hypothesis corresponds to an alert situation. There are two types of error to distinguish here: *type-I* error and *type-II* error. Type-I error corresponds to a false alarm where the decision center decides on the alternative hypothesis when the true hypothesis is the null hypothesis. Type-II error corresponds to a missed detection where the decision center decides on the null hypothesis when the true one is the alternative hypothesis. Since our interest is in alert systems where a missed detection is more critical than a false alarm, we require the type-II error probability to decay to 0 exponentially fast, while the type-I error probability is only required to stay below a given threshold. We aim at maximizing the decay-rate of the type-II error probability with the number of observed samples.

Most of the information theoretic works studied the binary hypothesis testing problem when there is a single sensor and a single decision center [1–22]. When the sensor and the decision center (DC) are co-located, the setup is called centralized hypothesis testing and the decision center has full access to all measurements in the system. In a centralized setup, the Neyman-Pearson Lemma [23] characterizes the fundamental tradeoff between the two types of error probabilities. It is achieved via a likelihood ratio test with an appropriately chosen threshold, where the threshold increases with the desired type-I error probability and decreases with the type-II error probability. An exact characterization of the set of type-I and type-II error probabilities that are simultaneously achievable for finite number of samples (which in this thesis we denote by n) is not yet available. The Chernoff-Stein Lemma [23, Theorem 11.8.3] however exactly characterizes the optimal exponential decay-rate of the more critical type-II error probability when the type-I error probability stays below a given threshold $\epsilon > 0$. It is given by the Kullback-Leiber Divergence between the probability mass functions (pmfs) of the observations under the two hypotheses and does not depend on the allowed type-I error probability ϵ .

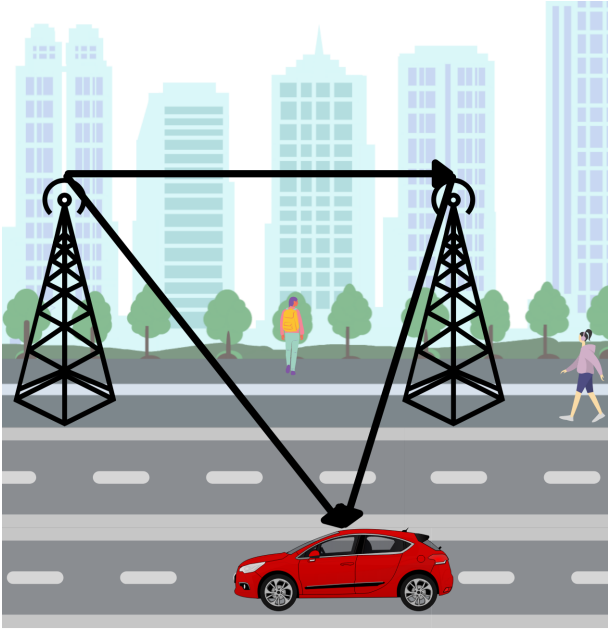
When the sensor and the DC are not co-located, we call the setup distributed hypothesis testing. In this case, the sensor communicates with the DC over a communication link, which is often modeled as noise-free but rate-limited link. The communication link can be subject either to a *maximum-rate* constraint or to an *expected-rate* constraint. Under maximum-rate constraint, which is the focus of most of the previous

works, the *maximum* length of *any* message sent over the communication link is limited. However, under expected-rate constraint, which was introduced recently in [20], the *expected* length of the message sent over the communication link is constrained. Extensions of the distributed hypothesis testing setup with a single sensor and a single decision center were considered under maximum-rate constraints to setups with either multiple sensors [6, 18, 24–27], multiple decision centers [28–32], or both of them [33–35]. In this thesis, we provide *the first extensions* of the distributed hypothesis testing setup under expected-rate constraints to setups with multiple sensors and/or multiple decision centers.

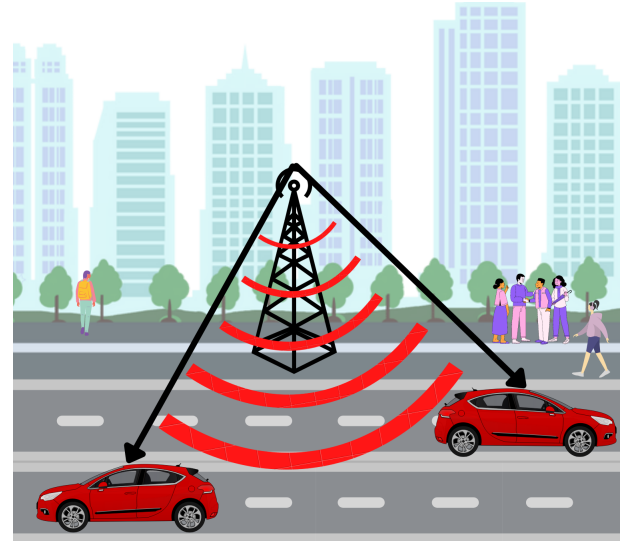
Specifically, in this thesis we focus on the three network structures in Figure 1.1, namely on: a) A cooperative two-sensor single-DC multi-access-channel (MAC) setup (Figure 1.1a) where the sensors are close to each other and thus can cooperate in their transmissions to the decision center. For simplicity, cooperation is assumed to be unidirectional from the first sensor to the second sensor. b) A ($K \geq 2$)-hop setup with K sensors and K DCs (Figure 1.1b), where the communication is short-range and thus only between neighbouring sensors. Such a scenario is motivated by the stringent battery constraints of IoT devices. c) A single-sensor two-DC broadcast-channel (BC) setup (Figure 1.1c) as relevant for systems where sensors are reused to serve different DCs such as common traffic sensors broadcasting to multiple automotive cars. Though simple, this scenario is interesting because it reveals the tradeoff in the communication from the sensor to the DCs over a common BC that stems from the different needed information to decide (due to the different local observations) at the two DCs.

All these setups have previously been studied, and the set of possible type-II error exponents that are simultaneously achievable at the various DCs when *testing against independence* or *testing against conditional independence* has been characterized under constraints on the maximum communication rates. In the special case of testing against independence, under the alternative hypothesis, it is either considered that the observations at the different terminals are independent of each other or that the observations at the sensors are correlated together but are independent of the observations at the DC, with same marginals as under the null hypothesis in both cases. For testing against conditional independence, the joint distribution of the observations under the alternative hypothesis is equal to the product of the conditional marginal distributions conditioned on a common random variable. Moreover, these works mostly rely on specific Markov properties between the observations at the various terminals, which can be motivated by the assumption that observations of sensors that are farther away are less correlated than observations of neighbouring sensors. While in almost all these previous works, except for the two-hop setup [35], the

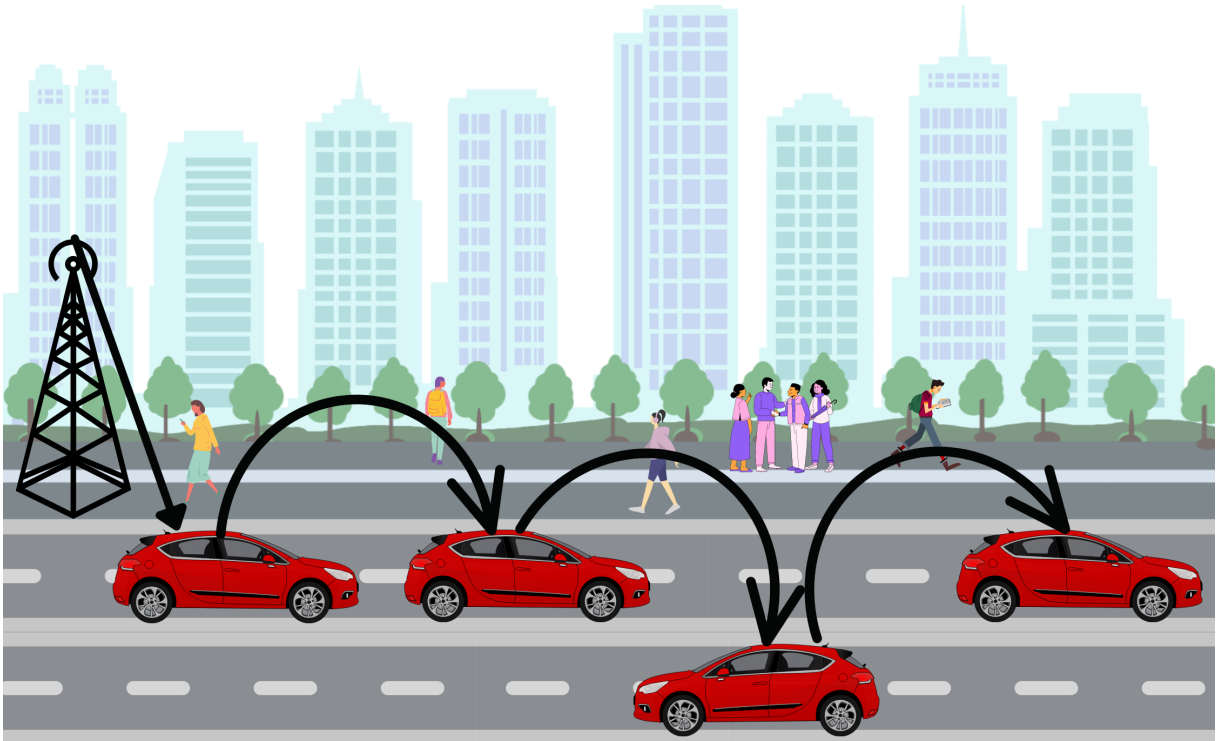
type-I error probability is required to vanish asymptotically, in this thesis we derive the optimal type-II error exponents for arbitrary type-I error probability thresholds $\epsilon \in [0, 1)$. We consider both maximum- and expected-rate constraints.



(a) Two sensors cooperating and communicating with a DC (automotive car).



(c) Single sensor communicating over BC (communication in red) and two individual links to two DCs (automotive cars).



(b) Multi-hop network with multiple sensors (including sensors at cars) and multiple DCs (all cars).

Figure 1.1: Different network structure with multiple sensors and/or multiple DCs.

For maximum-rate constraints, our results show that the set of achievable type-II error exponents does not depend on the permissible type-I error probability thresholds. Such a result is generally referred to as a *strong converse* result in information theory. In contrast, for expected-rate constraints, the set of achievable type-II error exponents depends on the various type-I error thresholds. Moreover, depending on the values of the different thresholds at the various DCs in the network, our results illustrate tradeoffs between the type-II error exponents that can simultaneously be achieved at the different DCs. The resulting tradeoff when different DCs have different type-I error thresholds is *first of its kind* and we *exactly characterize* it for the various setups we study.

To prove the achievability results under expected-rate constraints, we propose *a new rate-sharing* strategy that generalizes the degenerate rate-splitting scheme in [20] to apply to more interesting sets of rates. That means, we choose different sets of rates for the network and we apply an optimal coding and testing scheme under maximum-rate constraints for each of the chosen rate-tuples with a certain probability. In most of our setups, we manage to provide explicit characterizations of the probabilities that should be used for the rate-sharing: they are determined by the type-I error probability thresholds at the various DCs.

The converse proof methods that we use in this thesis rely on a similar change of measure argument as in [36–38], where we also restrict to jointly typical source sequences as [37]. No variational characterizations, or hypercontractivity arguments [39] are required. Instead, we rely on arguments showing that certain Markov chains hold in an asymptotic regime of infinite blocklengths and in some steps we resort also to the blowing-up lemma [40]. Our method to circumvent variational characterizations, or hypercontractivity, or blowing-up arguments seems to extend also to other converse proofs, see for example the simplified proof of the well-known strong converses for lossless and lossy compression with side-information at the decoder [41, 42] presented in [43].

1.2 Main Contributions

We present in the sequel the main contributions of this thesis:

- For the Cooperative MAC setup with two sensors and a single decision center:
 - We characterize the optimal error exponent of the setup under *expected-rate constraints* and under a specific Markov chain. To achieve this error exponent, we propose an optimal coding scheme using the technique of Salehkalaibar and Wigger [20] that chooses a subset of observations

- at the first sensor, under which the decision center always raises an alert by communicating a single-bit over all links, and otherwise applies the optimal coding scheme under maximum-rate constraints with vanishing type-I error probability.
- For the setup under *maximum-rate constraints*, we re-establish the weak converse proof in general and we show a *strong* converse result under the Markov chain. Under this result, the optimal error exponent is shown to be independent of the type-I error probability.
 - We show that our results extend the rate-gain factor of $(1 - \epsilon)^{-1}$ in [20] to all communication links in the cooperative MAC setup. This implies a boost in the optimal type-II error exponent under expected-rate constraints for any type-I error threshold $\epsilon > 0$.
- For a network with K hops and K decision centers:
 - We provide an exact characterization of the general fundamental exponents region under expected-rate constraints showing a tradeoff between the exponents at all decision centers with different type-I error thresholds. The proposed optimal coding scheme to achieve this region, based on our *multiplexing and rate-sharing strategy*, combines 2^K subschemes each with a certain probability, and with given rates. To determine the optimal scheme, one needs to optimize over the probabilities and the chosen rates subject to the available rate budgets.
 - We prove that our results simplify for the special cases of $K = 2$ or $K = 3$ hops, where the simplified optimal coding scheme is formed of $K + 1$ subschemes only. In this simplified scheme, the probabilities of the subschemes are determined by the type-I error thresholds and no optimization over them is needed. A conjecture for the K hops is proposed.
 - We further provide a strong converse result showing that the fundamental exponents region is of a rectangular form under maximum-rate constraints, which means that all DCs can achieve their optimal error exponents at the same time. For the converse proofs, we propose an alternative method using change of measure arguments and asymptotic Markov chains.
 - For $K = 2$, we also analyze the case when a more flexible sum-rate constraint (i.e. a constraint on the total rate in the system $R = R_1 + R_2$) is considered instead of the strict individual rate constraints.
 - For a BC with a single sensor and two decision centers:

- We provide an exact characterization of the general exponents region under expected-rate constraints. This region shows an additional tradeoff element, stemming from differences in the admissible type-I error thresholds at the DCs, compared to the tradeoff under maximum-rate constraints, stemming from the competition on the resources of the common BC link to the two DCs. The proposed optimal coding scheme to achieve this region consists of four subschemes, where the available rate at each link is distributed among these subschemes.
- We further provide a strong converse result for the setup under maximum-rate constraints showing that the fundamental exponents region is independent of the type-I error thresholds.

To summarize, in this thesis, we present the first results on the extended distributed binary hypothesis testing setup under expected-rate constraints with multiple sensors and/or multiple DCs. We propose a new rate-sharing strategy that multiplexes different subschemes with vanishing type-I error probabilities. Each subscheme is applied with a given probability and a rate choice, subject to the constraints on type-I error probabilities and available rates, to serve a subset of the DCs and maximize their type-II error exponents. To the best of our knowledge, we are the first to observe a tradeoff between the type-II error exponents at the different DCs when they have different type-I error thresholds. We show that our strategy is optimal when the applied subschemes resemble the optimal coding schemes for the targeted subset of the DCs under maximum-rate constraints. However, our proposed strategy is not specific to any setup, nor to the case when optimal coding schemes exist. As a result, this rate-sharing strategy can be *generalized* to other setups under expected-rate constraints and can be applied with given *generic* coding schemes of vanishing type-I error probabilities under maximum-rate constraints, and it is still expected to introduce the same comparable gains. Furthermore, we propose new proof methods for converse and strong converse results which use change of measure arguments and asymptotic Markov chains with or without the blowing-up lemma. Although these methods require certain setup-specific steps to establish our converse proofs, they seem to be extendable to other setups, not necessarily related to hypothesis testing [43].

1.3 Organization of Thesis

In Chapter 2, we present the state of the art for the distributed binary hypothesis testing problem with a single sensor and a single decision center (DC), and we review the main results and tools that exist in the literature in the special case of testing against independence on which we focus in this thesis. In Chapter 3,

we study the first extension of the problem under expected-rate constraint to multiple sensors and derive the optimal coding schemes and error exponents for two variations of the multi-sensor setup with single DC: cooperative MAC and two-hop networks. In Chapter 4, we introduce multiple DCs into the problem, first for the two-hop network and then for the extended K -hop network. A new rate-sharing strategy is proposed for optimal coding schemes under expected-rate constraints. The fundamental exponents region is exactly characterized for both models, and simplified for the special cases of two hops and three hops. In addition, new converse and strong converse proof techniques are presented. In Chapter 5, we study the setup with a single sensor communicating with two DCs over a common BC channel and two individual links. The optimal coding schemes and the fundamental exponents region are characterized for the general case and simplified for the special case of having only the common BC channel.

1.4 Notation

The set of all real numbers is denoted by \mathbb{R} and the set of nonnegative real numbers is denoted by \mathbb{R}_0^+ . For other sets, we use calligraphic letters, e.g. \mathcal{Y} . Random variables are denoted by uppercase letters, e.g. Y and their observations by lowercase letters, e.g. y . For the tuple of random variables (Y_1, Y_2, \dots, Y_n) we write Y^n and y^n for the tuple of n -samples of the observations (y_1, y_2, \dots, y_n) . We further use T_0 and T_1 for Transmitters, R_1, R_2, R_3 , and R_K for Receivers or Relays.

We use sans serif font for the bit-strings: \mathbf{M} for a random and \mathbf{m} for a deterministic bit-string. We let $\text{len}(\mathbf{m})$ and $\text{dec}(\mathbf{m})$ denote the length and the corresponding positive integer of the bit-string \mathbf{m} . For any positive integer, we let $\text{bin}(m)$ denote the length $\lceil \log_2(m) \rceil$ binary (bit-string) representation of the index m .

In addition, $\mathcal{T}_\mu^{(n)}$ denotes the strongly typical set given by [44, Definition 2.8], $H(\cdot)$ denotes the entropy function, $I(\cdot)$ the mutual information quantity, $h_b(\cdot)$ the binary entropy function, and $D(P\|Q)$ the Kullback-Leibler divergence between two probability mass functions (pmfs) on the same alphabet. For any positive integer K , we denote by $\mathcal{P}(K)$ the power set of all subset of $\{1, \dots, K\}$ excluding the emptyset which is denoted by \emptyset .

Chapter 2

Preliminaries: Single-Sensor Single-DC Networks

2.1 Introduction

In this chapter, we review the previous results on binary hypothesis testing system with a single sensor and a single decision center (DC), which are related to our work. We first present a general overview of the state of the art for different setups. Then we describe the distributed hypothesis testing setup, when the goal is to maximize the type-II error exponent for a given type-I error threshold and when the communication is over a noiseless link. In Sections 2.4 and 2.5, we review the *existing* optimal type-II error exponents and coding schemes for the special case of testing against independence under maximum-rate constraints and under expected-rate constraints, respectively.

2.2 State of the Art

Several information theoretic works investigated the problem of distributed binary hypothesis testing for a single sensor and a single DC under various assumptions [1–22]. In this section, we focus on the works maximizing the type-II error exponent, in noiseless or noisy communication environments, and under various rate constraints (single-bit, zero-rate, and positive-rate).

For the special case of *testing against independence* under *maximum-rate* constraint over a *noiseless* link, Ahlswede and Csiszàr [1] characterized the optimal type-II error exponent for testing against independence. They also showed that the optimal error exponent is independent of the type-I error threshold $\epsilon > 0$ for

general hypothesis testing, i.e., for arbitrary pmfs underlying the two hypotheses, and provided a lower bound to the optimal error exponent. Later, Han [2] and Shimokawa, Han, and Amari (SHA) [5] improved their lower bound for general hypothesis testing. The coding scheme in [5] uses quantization and unstructured binning at the encoder and a minimum empirical entropy decoder. The SHA scheme was shown to be optimal in the special case of testing against conditional independence [6].

The problem of general binary hypothesis testing for a single sensor and a single DC has been an open problem for long time and the SHA scheme was known to be the best achievable scheme for this problem. However, the very recent works of Haim and Kochman [9] and Watanabe [22] proved the suboptimality of the SHA scheme in some cases. The work of Haim and Kochman [9] extended the results of the SHA scheme and showed that using nested codes for the quantize and bin strategy in the special case of doubly-symmetric binary sources (DSBS) at least recovers the performance of the SHA scheme and slightly improves over it for specific values of correlation between the sources. Furthermore, Watanabe [22] presented a sequential scheme that clearly improves over the SHA scheme in the special case of having sources as a product of DSBS. This sequential scheme is a modified version of the SHA scheme that no longer uses naïve random binning.

For *zero-rate* constraints in the general hypothesis testing problem, Han [2] characterized the optimal type-II error exponent when the sensor communicates a single-bit to the decision center. Shalaby and Pappas [4] extended the work of Han to messages consisting of a sublinear (in the number of observations) number of bits. They proved that the optimal error exponent coincides with that obtained by Han [2], that is if the sensor is subject to asymptotically zero-rate constraint, then single-bit messages are enough to achieve the optimal exponent at the DC.

Moreover, for a *positive-rate* communication over a *noisy* discrete memoryless channel (DMC), Sreenuvar and Gündüz [12, 14, 17, 26] and Salehkalaibar and Wigger [18] studied the setup for general hypothesis testing and derived optimality results in the special cases of testing against independence, testing against conditional independence, and generalized testing against conditional independence. The results are based on joint source-channel coding and show that for generalized testing against conditional independence, the optimal type-II error exponent depends only on the capacity of the channel but not on its transition law. This seems to be different for general testing.

Recently, Salehkalaibar and Wigger [20] introduced the distributed hypothesis testing setup under an *expected-rate* constraint. They showed that this flexible rate constraint leads to a boost in the optimal error

exponent as a function of the type-I error threshold ϵ , in contrast to the result of [1] under maximum-rate constraints where the optimal error exponent is independent of ϵ . They also provided optimality results in the special case of testing against independence when the communication is over a noiseless link or over a noisy channel with a stop-feedback that informs the sensor to stop transmission. Their proposed scheme uses variable-length coding in a simple manner — it either sends single-bit messages or uses the fixed-length coding scheme as in the optimal coding scheme under maximum-rate constraint — which is proved to be enough and optimal.

Related are also the variable-length coding works by Salehkalaibar and Tan [21] and by Inan et al. [45]. Salehkalaibar and Tan considered a sequential hypothesis testing problem with zero-rate communication and showed that limiting the expected number of transmission requests from the DC does not increase the error exponent compared to a fixed-length hypothesis testing setup with zero-rate compression as in the work of Shalaby and Papamarcou [4]. Inan et al. derived a new lower bound on the feasible error exponents region for a non-cooperative MAC setup with memoryless quantization. They showed that the bound can be approached closely at high rates by lattice quantization.

In the following sections we present the system models, optimal coding schemes, and main results, under maximum-rate and expected-rate constraints, on which we build our work in this thesis.

2.3 The Setup with Noiseless Links of positive Rates

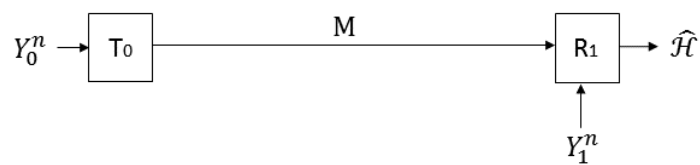


Figure 2.1: Single sensor node and single DC node.

Consider the distributed hypothesis testing problem in Figure 2.1. In the general binary hypothesis testing problem, depending on the binary hypothesis $\mathcal{H} \in \{0, 1\}$, the pair of measurements (Y_0^n, Y_1^n) is distributed as:

$$\text{under } \mathcal{H} = 0 : (Y_0^n, Y_1^n) \text{ i.i.d. } \sim P_{Y_0 Y_1}; \quad (2.1a)$$

$$\text{under } \mathcal{H} = 1 : (Y_0^n, Y_1^n) \text{ i.i.d. } \sim Q_{Y_0 Y_1}; \quad (2.1b)$$

for given probability mass functions (pmfs) $P_{Y_0 Y_1}$ and $Q_{Y_0 Y_1}$.

In the remainder of this thesis, we focus on testing against independence where

$$Q_{Y_0 Y_1} = P_{Y_0} P_{Y_1} \quad (2.2)$$

for P_{Y_0} and P_{Y_1} denoting the marginals of the pmf $P_{Y_0 Y_1}$.

The sensor node T_0 observes Y_0^n , while the decision node R_1 observes Y_1^n . The sensor node informs the decision node about its observations over a positive rate-limited communication link by sending the *bit-string* message $\mathbf{M} = \phi^{(n)}(Y_0^n)$ to R_1 , where the encoding function is of the form $\phi^{(n)} : \mathcal{Y}_0^n \rightarrow \{0, 1\}^*$. Two types of rate constraint are distinguished: maximum-rate constraint and expected-rate constraint. For a maximum-rate constraint, the maximum length of any communicated bit-string message \mathbf{M} is upper-bounded as

$$\text{len}(\mathbf{M}) \leq nR, \quad (2.3)$$

where R is the maximum available rate of communication. For an expected-rate constraint, as introduced by Salehkalaibar and Wigger in [20], only the expected length of the communicated bit-string message \mathbf{M} is upper-bounded:

$$\mathbb{E}[\text{len}(\mathbf{M})] \leq nR. \quad (2.4)$$

Receiver R_1 guesses hypothesis \mathcal{H} using its observation Y_1^n and the received bit-string message \mathbf{M} . I.e., using a decision function $g^{(n)} : \mathcal{Y}_1^n \times \{0, 1\}^* \rightarrow \{0, 1\}$, it produces the guess:

$$\hat{\mathcal{H}} = g^{(n)}(Y_1^n, \mathbf{M}) \in \{0, 1\}. \quad (2.5)$$

The type-I error probability, which corresponds to a false alarm, is given by

$$\alpha_n \triangleq \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0]. \quad (2.6)$$

The type-II error probability, which corresponds to a missed detection, is given by

$$\beta_n \triangleq \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1]. \quad (2.7)$$

The goal is to design encoding and decision functions such that their type-I error probability stays below a given threshold $\epsilon > 0$ and the type-II error probability decays to 0 with largest possible exponential decay while respecting a given rate constraint on the bit-string message \mathbf{M} .

Definition 1. Fix a maximum type-I error probability $\epsilon \in [0, 1]$ and rate $R \geq 0$. The type-II error exponent θ is called ϵ -*achievable* if there exists a sequence of encoding and decision functions $\{\phi^{(n)}, g^{(n)}\}_{n \geq 1}$ satisfying:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \triangleq \limsup_{n \rightarrow \infty} \alpha_n \leq \epsilon, \quad (2.8a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_n} \triangleq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta, \quad (2.8b)$$

and

$$\text{len}(\mathbf{M}) \leq nR, \quad (2.8c)$$

for the setup under maximum-rate constraint, or

$$\mathbb{E}[\text{len}(\mathbf{M})] \leq nR, \quad (2.8d)$$

for the setup under expected-rate constraint.

Definition 2. The supremum over all ϵ -achievable type-II error exponents θ is called the ϵ -optimal error exponent and is denoted $\theta_{\max}^*(R, \epsilon)$ for a maximum-rate constraint setup and $\theta^*(R, \epsilon)$ for an expected-rate constraint setup.

2.4 Maximum-Rate Constraint

In this section, we present the optimal coding scheme [2] and the characterization of the optimal error exponent for the distributed binary hypothesis testing problem of Figure 2.2 under maximum-rate constraint [1].

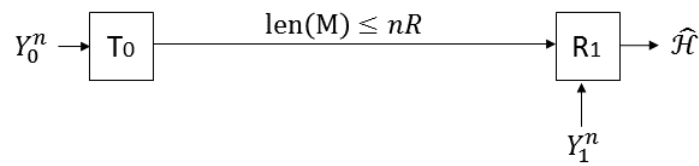


Figure 2.2: Point-to-point setup under maximum-rate constraint [1].

2.4.1 Optimal Error Exponent

Theorem 1 (Theorems 2&3 in [1]). *The ϵ -optimal error exponent in the special case of testing against independence under maximum-rate constraint $\theta_{\max}^*(R, \epsilon)$ is independent of ϵ and is given by*

$$\theta_{\max}^*(R, \epsilon) = \eta_1(R) := \max_{\substack{P_{U|Y_0}: \\ R \geq I(U; Y_0) \\ U \rightarrow Y_0 \rightarrow Y_1 \\ |\mathcal{U}| \leq |\mathcal{Y}_0| + 1}} I(U; Y_1). \quad (2.9)$$

Lemma 1. *The function $\eta_1(\cdot)$ is continuous, concave and monotonically non-decreasing on its entire domain \mathbb{R}_0^+ .*

Proof: See Appendix A.1. ■

2.4.2 Optimal Coding Scheme

We present this coding scheme, which we refer to as “*Han’s scheme*” in the rest of this manuscript. Han’s scheme is based on quantization and typicality checks. The set of observations at the transmitter are first quantized and compressed into a codebook of rate R . The transmitter selects the entry of the codebook that best represents its observation and sends the binary representation of this entry to the receiver. The receiver performs a typicality check between its local observations and the corresponding entry of the received message in the codebook. Details of the coding scheme are as follows.

Fix a blocklength n and choose the following parameters:

- a small positive number $\mu > 0$; and
- a conditional pmf $P_{U|Y_0}$ to generate $P_{Y_0Y_1U} = P_{Y_0Y_1}P_{U|Y_0}$,

where all mutual information quantities are evaluated according to the joint pmf $P_{Y_0Y_1U}$.

Randomly generate the codebook

$$\mathcal{C}_U \triangleq \left\{ u^n(m) : m \in \left\{ 1, \dots, 2^{n(I(U;Y_0)+\mu)} \right\} \right\}, \quad (2.10)$$

by drawing all entries i.i.d. according to the marginal pmf P_U .

Transmitter: Assume it observes $Y_0^n = y_0^n$. Then, it looks for indices m satisfying the condition that the pair $(u^n(m), y_0^n)$ lies in the strongly jointly typical set $\mathcal{T}_\mu^{(n)}(P_{UY_0})$, randomly picks one of these indices, and sends its corresponding bit-string

$$\mathbf{M} = [\text{bin}(m)]. \quad (2.11)$$

If no such index m exists, then T_0 sends

$$\mathbf{M} = [0]. \quad (2.12)$$

Receiver: Assume it observes $Y_1^n = y_1^n$ and receives the bit-string message $\mathbf{M} = \mathbf{m}$.

If $\mathbf{m} = [0]$, then it declares

$$\hat{\mathcal{H}} = 1. \quad (2.13)$$

Else it sets $m = \text{dec}(\mathbf{m})$, and checks if $(u^n(m), y_1^n) \in \mathcal{T}_\mu^{(n)}(P_{UY_1})$. It declares $\hat{\mathcal{H}} = 0$ if the check succeeds, and $\hat{\mathcal{H}} = 1$ otherwise.

2.5 Expected-Rate Constraint

2.5.1 Optimal Error Exponent

As mentioned earlier in this chapter, expected-rate constraints were first considered in the framework of distributed hypothesis testing by Salehkalaibar and Wigger [20]. In their work, the authors derived the ϵ -optimal error exponents for testing against independence in the setup of a single-sensor node with a single-decision node. They obtained a direct rate-boost by the factor $(1 - \epsilon)^{-1}$ showing thus that the type-II error exponent is dependent on the type-I error threshold ϵ under expected-rate constraint, in contrast to the case under maximum-rate constraint. Their main result is stated in the following theorem.

Theorem 2 (Theorem 1 in [20]). *The ϵ -optimal error exponent $\theta^*(R, \epsilon)$ in the special case of testing against independence under expected-rate constraint is given by*

$$\theta^*(R, \epsilon) = \eta_1 \left(\frac{R}{(1 - \epsilon)} \right) := \max_{\substack{P_{U|Y_0}: \\ R \geq (1 - \epsilon)I(U; Y_0) \\ U \rightarrow Y_0 \rightarrow Y_1 \\ |\mathcal{U}| \leq |\mathcal{Y}_0| + 1}} I(U; Y_1), \quad (2.14)$$

where $\eta_1(R)$ is defined in (2.9).

2.5.2 Optimal Coding Scheme

The optimal coding scheme achieving the optimal exponent in Theorem 2 multiplexes two subschemes, depending on the observations Y_0^n of the transmitter T_0 . In one subscheme, the decision center at R_1 tries to correctly guess the hypothesis \mathcal{H} by applying Han's scheme, which is reported in Subsection 2.4.2, while in the other subscheme, it directly declares $\hat{\mathcal{H}} = 1$.

The structure of this coding scheme is illustrated in Figure 2.3. The main idea is to choose a subset \mathcal{D}_\emptyset of the set of observations \mathcal{Y}_0^n with the following probability in the limit as $n \rightarrow \infty$

$$\overline{\lim}_{n \rightarrow \infty} \Pr[Y_0^n \in \mathcal{D}_\emptyset] = \epsilon. \quad (2.15)$$

Then whenever the observation Y_0^n lies in the set \mathcal{D}_\emptyset , a degenerate subscheme is applied with

$$\mathbf{M} = [0] \quad \text{and} \quad \hat{\mathcal{H}} = 1. \quad (2.16)$$

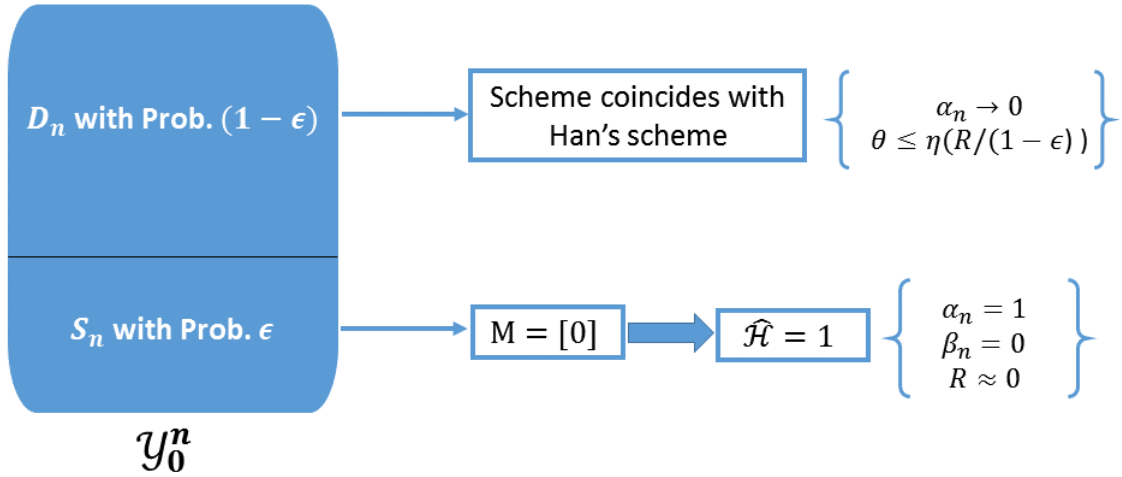


Figure 2.3: Structure of the optimal coding scheme for point-to-point setup under expected-rate constraint.

Otherwise, Han's scheme is applied with a maximum-rate constraint $\tilde{R} = R/(1 - \epsilon)$. Since in Han's scheme, the type-I error probability tends to 0 as $n \rightarrow \infty$, the type-I error probability of the overall scheme is upper-bounded by ϵ in the limit as $n \rightarrow \infty$. The expected-rate constraint is satisfied where a single-bit does not change the communication rate as the blocklength n grows large. Note that the type-II error probability in the degenerate subscheme is equal to 0, and thus the type-II error exponent of the overall scheme is determined by that of Han's scheme for a boosted maximum-rate $\tilde{R} = R/(1 - \epsilon)$. This achieves the desired error exponent in Theorem 2.

2.6 Numerical Comparison

The gain resulting from an expected-rate constraint compared to a maximum-rate constraint is illustrated in the following example.

Example 1. Let Y_0 and T be independent Bernoulli random variables of parameters $p_{Y_0} = 0.5$ and $p_T = 0.9$ and set $Y_1 = Y_0 \oplus T$.

For this example, Figure 2.4 shows the optimal error exponents under expected-rate and maximum-rate constraints, given by functions $\eta_1\left(\frac{R}{1-\epsilon}\right)$ and $\eta_1(R)$, when $\epsilon = 0.1$ as a function of the rate R . The figure illustrates the benefits of expected-rate constraints and the concavity and monotonicity of the function $\eta_1(\cdot)$.

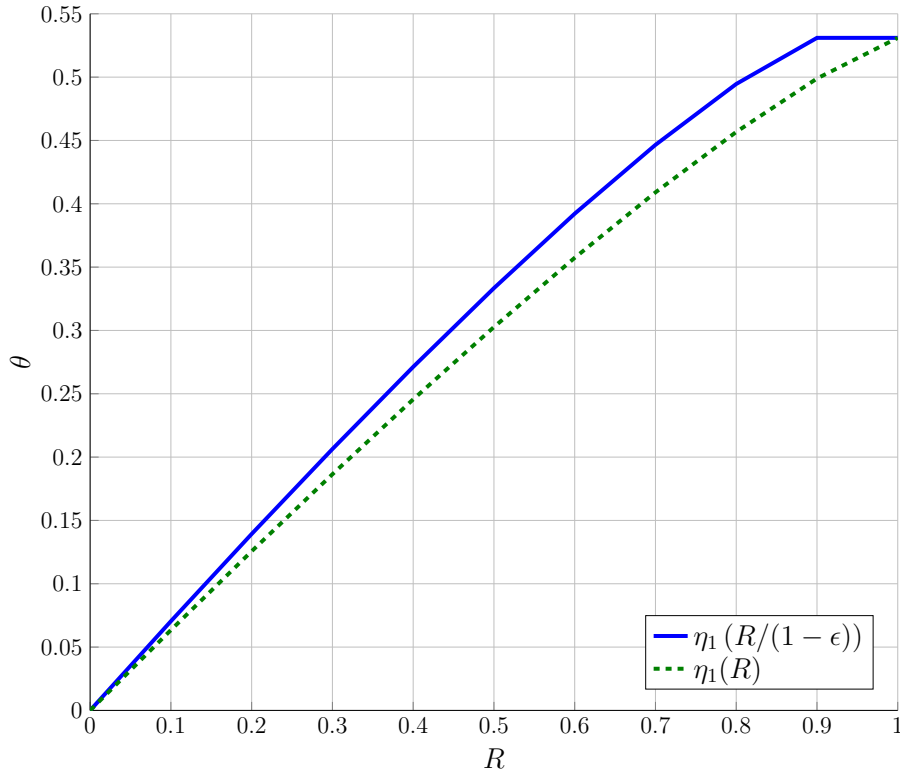


Figure 2.4: Optimal exponents under expected-rate and maximum-rate constraints for Example 1 when $\epsilon = 0.1$.

2.7 Summary and Discussion

In this chapter, we surveyed the existing works in the literature on the problem of distributed binary hypothesis testing for a single sensor and a single DC. We focused on testing against independence when the communication is assumed noise-free, and presented the main existing results on the problem either under maximum-rate constraint or under expected-rate constraint. We further highlighted the main gain of expected-rate constraints for positive rate and type-I error threshold ϵ is the rate-gain of the factor $(1-\epsilon)^{-1}$ as proved in [20]. We are thus interested in exploring if this rate-gain extends to other setups with multiple sensors as we investigate in the next chapters.

Chapter 3

Multi-Sensor Single-DC Networks

3.1 Introduction

In this chapter, we address the testing-against-independence problem, for example to detect if the measurements in a smart vehicle are independent from other measurements in the environment which might indicate that the vehicle is failing or being hacked. For this problem, we consider two setups with two sensors and a single decision center under expected-rate constraints which represent the first extensions of the work by Salehkalaibar and Wigger in [20] to multi-sensor networks. In the first setup, illustrated in Figure 3.1, the first sensor communicates over a shared link to both the second sensor and the DC, and after receiving its message, the second sensor communicates with the DC. We refer to this setup as “Cooperative MAC”. The two sensors observe the sequences Y_0^n and Y_1^n , respectively, and the DC observes Y_2^n , where we assume that the Markov chain

$$Y_0^n \rightarrow Y_1^n \rightarrow Y_2^n \quad (3.1)$$

holds under the null hypothesis $\mathcal{H} = 0$ whereas under the alternative hypothesis $\mathcal{H} = 1$, Y_2^n is independent of the pair (Y_0^n, Y_1^n) . In the second setup, illustrated in Figure 3.2, the first sensor cannot communicate directly with the distant DC (due to low energy constraints for instance) and the second sensor is also acting as a relay to convey the first sensor’s message. The observations at the transmitter Y_0^n , the relay Y_1^n , and the DC Y_2^n , form again the Markov chain (3.1) under $\mathcal{H} = 0$, while all observations are independent from each other under $\mathcal{H} = 1$.

For both setups, we characterize the optimal type-II error exponents under expected-rate constraints. Similarly to the single-sensor result in [20], we show that the optimal error exponent depends on the accepted type-I error threshold ϵ . Indeed, we show that the variable-length coding allows to boost the rates over all

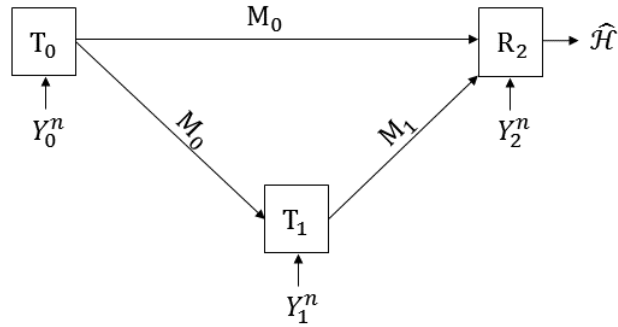


Figure 3.1: Cooperative MAC with two sensors and a single decision center.

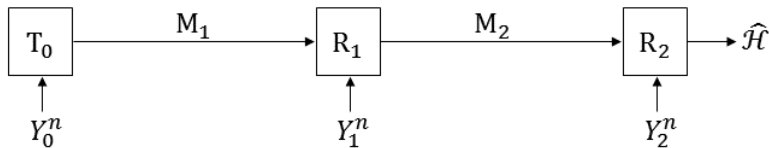


Figure 3.2: Two-hop setup with a single decision center.

the communication links in the network by the factor $(1 - \epsilon)^{-1}$ over an optimal fixed-length coding under maximum-rate constraints. Furthermore, under maximum-rate constraints, we prove a strong converse results for the two setups, i.e., we show that under maximum-rate constraints, the optimal type-II error exponents are independent of the accepted type-I error threshold ϵ , and thus the best exponents can be achieved with vanishing type-I error probabilities. The strong converse for the two-hop network was already proved in [35]; our proof is shorter and seems simpler.

3.2 Related Works

Distributed binary hypothesis testing has been studied for various network scenarios. We review in this section the prior works related to MACs and multi-hop networks.

3.2.1 MAC

In MAC networks, multiple sensors communicate directly with the single DC. In our cooperative MAC, there is an additional shared link from one sensor to the other. In the sequel, we refer to the MAC setups without this shared link as “Non-cooperative MAC” (Figure 3.3). This setup is the most widely studied MAC setup in the literature.

The non-cooperative MAC, with or without local observations at the DC, was studied in [2, 4, 6, 9, 24] under maximum-rate constraints. In [2], Han first established a lower and an upper bound on the

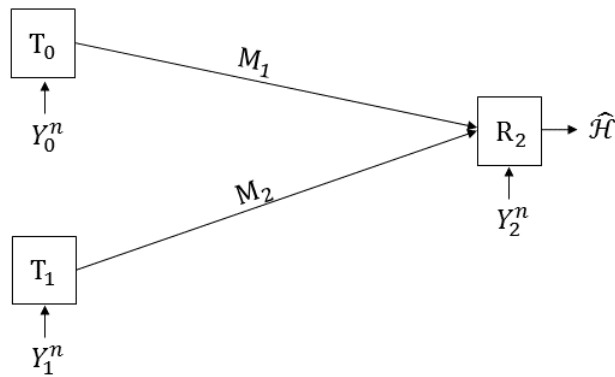


Figure 3.3: Non-cooperative MAC setup with two sensors and a single decision center.

optimal error exponent for general binary hypothesis testing. Rahman and Wagner [6] further highlighted the benefits of using binning and improved the lower bound obtained by Han using a quantize-bin-test scheme. In the special case of testing against conditional independence, they characterized the optimal error exponent. Haim and Kochman [9] studied the setup for symmetric-rate constraints, where both messages have equal rates, using Körner-Marton coding [46]. Zhao and Lai [24] studied the special case of testing against independence and provided a lower and an upper bound on the optimal error exponent.

In the special case of complete data compression, where communication is of a single bit, Han [2] characterized the optimal error exponent. Shalabi and Papamarcou [4] extended this result of Han to the more general case of asymptotically zero-rate compression. In this latter case, the messages sent by either one or both of the two sensors do not necessarily need to be of a single-bit only, but their communication rates tend to zero in the limit as the number of samples grows to infinity. Shalabi and Papamarcou proved that the optimal error exponent region coincides with that of the complete data compression obtained by Han [2], i.e., if one of the two sensors is subject to zero-rate data compression, then single-bit messages are enough for both sensors to achieve the optimal error exponent at the decision center. Zhao and Lai [24] extended the study to an arbitrary number of sensors with zero-rate compression under an exponential constraint, instead of a constant constraint, on the type-I error, i.e., when the type-I error is required to decrease exponentially fast to 0, and characterized the optimal type-II error exponent under such constraint.

For the non-cooperative MAC with noisy communication, Sreekumar and Gündüz [26] studied the setup when communication takes place over orthogonal DMCs and characterized the optimal error exponent in the special case of testing against conditional independence. In addition, Salehkalaibar and Wigger characterized in [27] the optimal error exponent in the special case when the observations of the sensors are independent under both hypotheses for testing against independence and where the communication occurs

over a general memoryless MAC. In [18], Salehkalaibar and Wigger derived similar results for generalized testing against conditional independence when the communication occurs over a MAC that decomposes into two DMCs. The general problem of testing against independence over memoryless MACs remains open even for noiseless communication.

Our main interest in this thesis is on MAC setups with an additional communication link from one sensor to the other. For such a cooperative MAC setup, Zhao and Lai [25] studied the problem of distributed binary hypothesis testing under maximum-rate constraints. They proposed a coding scheme using superposition coding and showed that it is optimal in the special case of testing against independence for vanishing type-I error probability. We present the coding scheme and main result of [25] in Section 3.3.2. In our work on the cooperative MAC setup, we prove a strong converse to this result under maximum-rate constraints and derive the optimal error exponent under *expected-rate* constraints [47]. We present our results in Sections 3.3.2.3, 3.3.3, 3.3.4, and 3.4.

3.2.2 Two Hops

Salehkalaibar, Wigger, and Wang [33] studied the two-hop relay network with two decision centers, where the relay takes a decision too. They proposed a “decision forwarding” strategy where the relay performs a joint typicality check between its observations and the decoded message received from the first sensor. Then it forwards this decision with its message to the decision center. This strategy is proved to be optimal in the special case of testing against independence under maximum-rate constraints and where a Markov chain is formed between the observations at the sensor, the relay, and the decision center. In this chapter, we consider a special case of the two-hop setup with only one decision center, see Figure 3.2. Thus, we present a special version of the optimality results of [33] under maximum-rate constraints and the corresponding strong converse result derived in [35] in Section 3.5.2. In this thesis, we simplify the proof of the strong converse result in [35]. We also provide the optimal coding scheme and the optimal type-II error exponent under expected-rate constraints. We present our results in Sections 3.5.3 and 3.5.2.2.

3.3 The Cooperative MAC Model

3.3.1 The Setup

Consider the distributed hypothesis testing problem in Figure 3.1 under the Markov chain (3.1) and in the special case of testing against independence between the observations at the sensors and the observations

at the decision center. Depending on the binary hypothesis $\mathcal{H} \in \{0, 1\}$, the tuple (Y_0^n, Y_1^n, Y_2^n) is thus distributed as:

$$\text{under } \mathcal{H} = 0 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0 Y_1} \cdot P_{Y_2 | Y_1}; \quad (3.2a)$$

$$\text{under } \mathcal{H} = 1 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0 Y_1} \cdot P_{Y_2} \quad (3.2b)$$

for given pmfs $P_{Y_0 Y_1}$ and $P_{Y_2 | Y_1}$ and where P_{Y_0} , P_{Y_1} , and P_{Y_2} denote the marginals of the joint pmf $P_{Y_0 Y_1 Y_2} := P_{Y_0 Y_1} P_{Y_2 | Y_1}$.

The system consists of two transmitters T_0 and T_1 , and a receiver R_2 . The transmitter T_0 observes the source sequence Y_0^n and sends its bit-string message $\mathbf{M}_1 = \phi_0^{(n)}(Y_0^n)$ to both T_1 and R_2 , where the encoding function is of the form $\phi_0^{(n)} : \mathcal{Y}_0^n \rightarrow \{0, 1\}^*$ and satisfies the rate constraint. For a maximum-rate constraint, the maximum length of \mathbf{M}_1 satisfies

$$\text{len}(\mathbf{M}_1) \leq nR_1, \quad (3.3)$$

whereas for an expected-rate constraint, the expected length of \mathbf{M}_1 satisfies

$$\mathbb{E}[\text{len}(\mathbf{M}_1)] \leq nR_1. \quad (3.4)$$

The second transmitter T_1 observes the source sequence Y_1^n and with the message \mathbf{M}_1 received from T_0 , it computes a bit-string message $\mathbf{M}_2 = \phi_1^{(n)}(Y_1^n, \mathbf{M}_1)$ using some encoding function $\phi_1^{(n)} : \mathcal{Y}_1^n \times \{0, 1\}^* \rightarrow \{0, 1\}^*$ that satisfies the rate constraint. For a maximum-rate constraint, the maximum length of \mathbf{M}_2 satisfies

$$\text{len}(\mathbf{M}_2) \leq nR_2, \quad (3.5)$$

whereas for an expected-rate constraint, the expected length of \mathbf{M}_2 satisfies

$$\mathbb{E}[\text{len}(\mathbf{M}_2)] \leq nR_2. \quad (3.6)$$

Then, T_1 sends \mathbf{M}_2 to the receiver R_2 , which guesses hypothesis \mathcal{H} using its observation Y_2^n and the received messages \mathbf{M}_1 and \mathbf{M}_2 , i.e., using a decision function $g^{(n)} : \mathcal{Y}_2^n \times \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}$, it produces the guess:

$$\hat{\mathcal{H}} = g^{(n)}(Y_2^n, \mathbf{M}_1, \mathbf{M}_2) \in \{0, 1\}. \quad (3.7)$$

The goal is to design encoding and decision functions such that their type-I error probability

$$\alpha_n \triangleq \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0] \quad (3.8)$$

stays below a given threshold $\epsilon > 0$ and the type-II error probability

$$\beta_n \triangleq \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1] \quad (3.9)$$

decays to 0 with largest possible exponential decay.

Definition 3. Fix a maximum type-I error probability $\epsilon \in [0, 1]$ and rates $R_1, R_2 \geq 0$. The type-II error exponent $\theta \geq 0$ is called ϵ -achievable if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)}, \phi_1^{(n)}, g^{(n)}\}_{n \geq 1}$ satisfying

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \leq \epsilon, \quad (3.10a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta, \quad (3.10b)$$

and

$$\text{len}(\mathbf{M}_i) \leq nR_i, \quad i \in \{1, 2\}, \quad (3.10c)$$

for the setup under maximum-rate constraints, or

$$\mathbb{E}[\text{len}(\mathbf{M}_i)] \leq nR_i, \quad i \in \{1, 2\}, \quad (3.10d)$$

for the setup under expected-rate constraints.

Definition 4. The supremum over all ϵ -achievable type-II error exponents θ is called the ϵ -optimal error exponent and is denoted $\theta_{\text{MAC, max}}^*(R_1, R_2, \epsilon)$ for the cooperative MAC setup under maximum-rate constraints and $\theta_{\text{MAC}}^*(R_1, R_2, \epsilon)$ for the cooperative MAC setup under expected-rate constraints.

3.3.2 Maximum-Rate Constraints

3.3.2.1 Optimal Coding Scheme under Maximum-Rate Constraints

The main feature in the optimal coding scheme proposed in [25] is the superposition coding where the codebook at T_1 is superposed on the message \mathbf{M}_1 received from T_0 . Since T_0 transmits its message \mathbf{M}_1 over the direct link to R_2 , the receiver can decode \mathbf{M}_1 using the adequate codebook. The coding scheme is described as follows.

Fix a large blocklength n , small numbers $\mu' < \mu'' < \mu'''$, and conditional pmfs $P_{U_1|Y_0}$ and $P_{U_2|U_1Y_1}$ satisfying the maximum-rate constraints

$$R_1 \geq I(U_1; Y_0) \quad (3.11)$$

$$R_2 \geq I(U_2; Y_1|U_1) \quad (3.12)$$

where mutual information quantities are calculated with respect to

$$P_{U_1U_2Y_0Y_1Y_2} \triangleq P_{U_1|Y_0}P_{U_2|U_1Y_1}P_{Y_0Y_1}P_{Y_2|Y_1}. \quad (3.13)$$

Randomly generate a codebook

$$C_{U_1} \triangleq \{u_1^n(m_1) : m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}\}, \quad (3.14)$$

by drawing all entries i.i.d. according to the marginal pmf P_{U_1} . For each codeword $u_1^n(m_1)$, generate a codebook

$$C_{U_2}(m_1) \triangleq \{u_2^n(m_2|m_1) : m_2 \in \{1, \dots, \lfloor 2^{nR_2} \rfloor\}\}, \quad (3.15)$$

by drawing the j -th entry of each codeword according to the conditional marginal pmf $P_{U_2|U_1}$.

This superposition construction of the codebooks is illustrated in Figure 3.4.

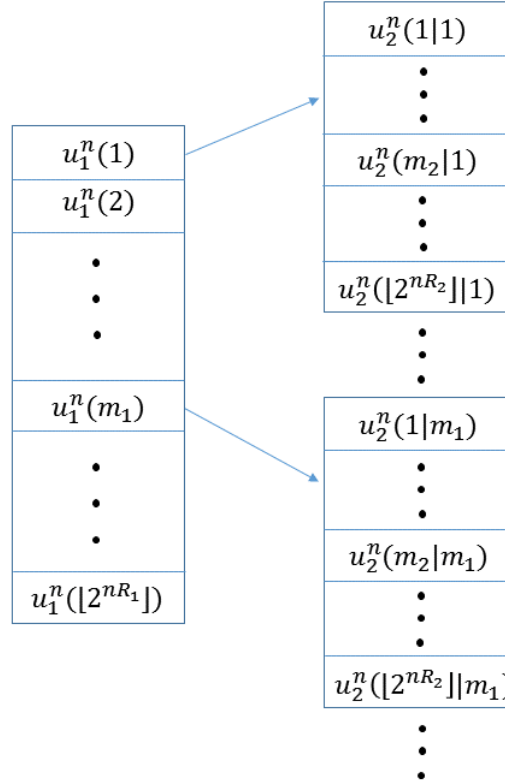


Figure 3.4: Superposition construction of the codebooks C_{U_1} and $C_{U_2}(m_1)$ with rates R_1 and R_2 respectively.

T₀: Assume it observes the sequence $Y_0^n = y_0^n$. Then it looks for indices $m_1 \geq 1$ satisfying

$$(u_1^n(m_1), y_0^n) \in \mathcal{T}_{\mu'}^n(P_{U_1 Y_0}),$$

randomly picks one of these indices, and sends its corresponding bit-string to both T₁ and R₂. Otherwise, it sends

$$\mathbf{M}_1 = [0]. \quad (3.16)$$

T₁: Assume it observes the sequence $Y_1^n = y_1^n$ and receives the message $\mathbf{M}_1 = m_1$ from T₀. If $m_1 = 0$, then

T_1 forwards the message

$$\mathbf{M}_2 = [0] \quad (3.17)$$

to R_2 . Else, it sets $m_1 = \text{dec}(\mathbf{m}_1)$ and looks for an index $m_2 \geq 1$ satisfying

$$(u_1^n(m_1), u_2^n(m_2|m_1), y_1^n) \in \mathcal{T}_{\mu''}^n(P_{U_1 U_2 Y_1}).$$

It randomly picks one of these indices and sends the bit-string message

$$\mathbf{M}_2 = [\text{bin}(m_2)] \quad (3.18)$$

to R_2 . Otherwise, it sends

$$\mathbf{M}_2 = [0]. \quad (3.19)$$

R_2 : Assume it observes the sequence $Y_2^n = y_2^n$ and receives messages $\mathbf{M}_1 = \mathbf{m}_1$ and $\mathbf{M}_2 = \mathbf{m}_2$. If any of the messages \mathbf{m}_1 or \mathbf{m}_2 equals $[0]$, it declares

$$\hat{\mathcal{H}} = 1. \quad (3.20)$$

Else, it sets $m_1 = \text{dec}(\mathbf{m}_1)$, $m_2 = \text{dec}(\mathbf{m}_2)$, and checks if

$$(u_1^n(m_1), u_2^n(m_2|m_1), y_2^n) \in \mathcal{T}_{\mu'''}^n(P_{U_1 U_2 Y_2})$$

then it declares

$$\hat{\mathcal{H}} = 0. \quad (3.21)$$

Otherwise, it declares

$$\hat{\mathcal{H}} = 1. \quad (3.22)$$

This coding scheme achieves the following optimal exponent.

3.3.2.2 Optimal Error Exponent

Theorem 3 (Theorems 1&2 in [25]). *The ϵ -optimal error exponent under the maximum-rate constraints (3.10c) and for vanishing type-I error probability is*

$$\lim_{\epsilon \downarrow 0} \theta_{\text{MAC}, \max}^*(R_1, R_2, \epsilon) = \eta(R_1, R_2) := \max_{\substack{P_{U_1|Y_0}, P_{U_2|U_1 Y_1}: \\ R_1 \geq I(U_1; Y_0) \\ R_2 \geq I(U_2; Y_1|U_1) \\ |\mathcal{U}_1| \leq |\mathcal{Y}_0| + 2 \\ |\mathcal{U}_2| \leq |\mathcal{U}_1| |\mathcal{Y}_1| + 1}} I(U_1 U_2; Y_2) \quad (3.23)$$

where the mutual information quantities are calculated with respect to the joint pmf $P_{U_1 U_2 Y_0 Y_1 Y_2} := P_{U_1|Y_0} P_{U_2|U_1 Y_1} P_{Y_0 Y_1 Y_2}$.

Proof: Achievability is proved in [25]. A converse proof, which corrects the converse in [25], is

presented in Appendix A.2.¹ The bounds on the cardinalities of U_1 and U_2 can be shown by standard applications of Carathéodory's theorem, see [48, Appendix C]. ■

3.3.2.3 Strong Converse Result

Theorem 3 shows the optimal error exponent for the general cooperative two-sensor single-DC setup without the Markov chain (3.1) under maximum-rate constraints and vanishing type-I error probability. Here, we strengthen that result in the special case when the Markov chain (3.1) holds.

This is the first result of this thesis.

Theorem 4. *The optimal error exponent under maximum-rate constraints (3.10c) and in the presence of the Markov chain (3.1) is independent of $\epsilon \in [0, 1)$ and given by:*

$$\theta_{\text{MAC,max}}^*(R_1, R_2, \epsilon) = \eta(R_1, R_2) \quad (3.24)$$

Proof: Achievability follows directly from Theorem 3. The strong converse is proved in Appendix A.3. It is similar to the converse under expected-rate constraints presented in Section 3.4 ahead but where inequality (3.72), i.e., $\Delta H(\tilde{M}_i) \leq nR_i \left(1 + h_b\left(\frac{\Delta}{nR_i}\right)\right)$, is replaced by the trivial inequality $H(\tilde{M}_i) \leq nR_i$. ■

Notice that $\theta_{\text{MAC,max}}^*(R_1, R_2, \epsilon)$ is independent of ϵ , and thus we can abbreviate it as $\theta_{\text{MAC,max}}^*(R_1, R_2)$.

In the following section, we present our optimal coding scheme under expected-rate constraints which builds on the optimal coding scheme under maximum-rate constraints described in Subsection 3.3.2.1.

3.3.3 Expected-Rate Constraints

3.3.3.1 Optimal Coding Scheme under Expected-Rate Constraints

In the optimal coding scheme under expected-rate constraints, the three terminals T_0, T_1, R_2 multiplex two different subschemes, and the choice of which subscheme to use depends on the transmitter T_0 's observations Y_0^n . In one subscheme, R_2 attempts to correctly guess the hypothesis \mathcal{H} , while in the other subscheme it simply declares $\hat{\mathcal{H}} = 1$. The main structure of the coding scheme is illustrated in Figure 3.5.

¹In the converse proof of [25, Theorem 2], the second line in the lower bound to R_1 relies on the identity $H(Y_{0,i}Y_{1,i}|M_1Y_0^{i-1}Y_1^{i-1}) = H(Y_{0,i}Y_{1,i}|M_1Y_0^{i-1}Y_{1,i+1}^n)$ which does not necessarily hold. In fact, the "=" has to be replaced with " \geq " and in some cases the inequality is strict, which poses problems in the converse. Note that we replace the notation of X_1 and X_2 in [25] by Y_0 and Y_1 in this thesis.

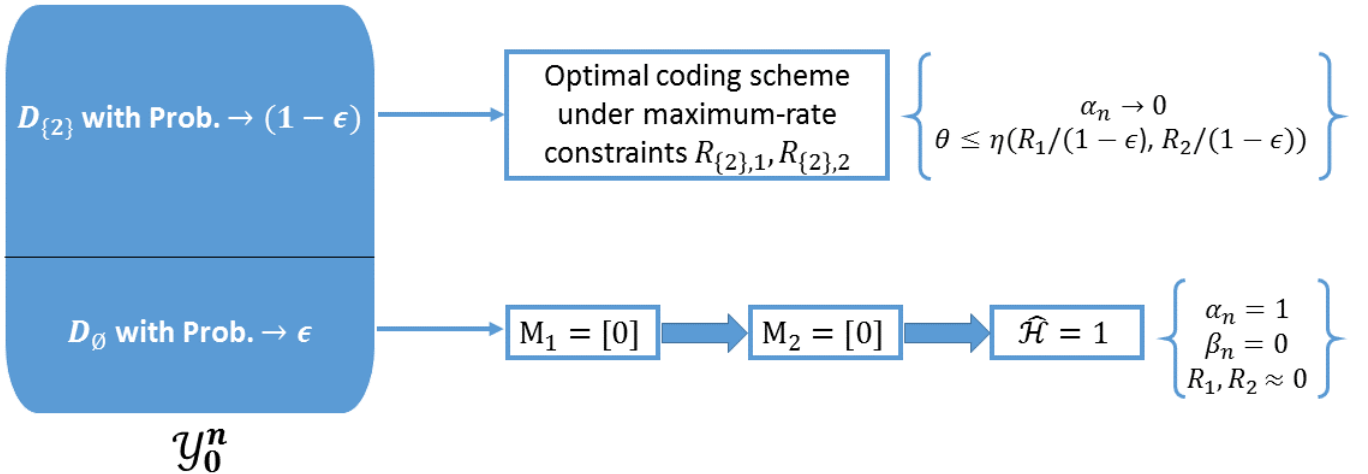


Figure 3.5: Optimal coding scheme for Cooperative MAC setup under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on all links.

We partition the set \mathcal{Y}_0^n into subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0}^n$ the probability of subset $\mathcal{D}_{\{2\}}$ is as large as possible but satisfies

$$\Pr [Y_0^n \in \mathcal{D}_{\{2\}}] \leq 1 - \epsilon. \quad (3.25)$$

As a consequence

$$\Pr [Y_0^n \in \mathcal{D}_\emptyset] \geq \epsilon. \quad (3.26)$$

Notice that as $n \rightarrow \infty$ the above inequalities turn into an equalities.

Remark 1. *The subscript of the subset \mathcal{D} indicates the set of decision centers trying to correctly guess the hypothesis while the others directly declare $\hat{\mathcal{H}} = 1$. Since receiver R_2 is the only decision center, the subscript could be either the empty set or the single-element set containing 2 as the index of R_2 .*

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset or $\mathcal{D}_{\{2\}}$, the three terminals follow a different subscheme.

If $Y_0^n \in \mathcal{D}_\emptyset$: In this case, T_0 and T_1 both send the single-bit string messages

$$M_1 = M_2 = [0] \quad (3.27)$$

and R_2 simply declares

$$\hat{\mathcal{H}} = 1. \quad (3.28)$$

This implies that the type-I and type-II error probabilities in this case are given by $\alpha_n = 1$ and $\beta_n = 0$.

If $Y_0^n \in \mathcal{D}_{\{2\}}$: In this case, R_2 attempts to correctly guess \mathcal{H} based on the transmitted messages. Specifically, T_0, T_1, R_2 all apply the encoding/decision functions of the optimal hypothesis testing scheme in Subsection 3.3.2.1 with vanishing type-I error probability and respecting maximum-rate constraints $R_{\{2\},1}$

and $R_{\{2\},2}$ on the two links, where these rates are chosen to satisfy

$$(1 - \epsilon)R_{\{2\},1} \leq R_1 \quad (3.29a)$$

$$(1 - \epsilon)R_{\{2\},2} \leq R_2. \quad (3.29b)$$

Notice that as per Remark 1, the first subscripts in $R_{\{2\},1}$ and $R_{\{2\},2}$ follow the subscript of the set $\mathcal{D}_{\{2\}}$.

Analysis: By (3.25), (3.27), (3.29), and since the transmission of a single bit hardly changes the rate for sufficiently large blocklengths, the overall scheme satisfies the expected-rate constraints R_1 and R_2 on the two links. Appendix A.4 proves that when the hypothesis testing scheme in Subsection 3.3.2.1 with vanishing type-I error probability [25] is employed for $Y_0^n \in \mathcal{D}_{\{2\}}$, then the overall scheme meets the permissible type-I error probability ϵ and achieves the error exponent given by Equation (3.30) of Theorem 5.

3.3.3.2 Optimal Error Exponent

Theorem 5. *For any $\epsilon \geq 0$, the ϵ -optimal error exponent under expected-rate constraints (3.4), (3.6), is*

$$\theta_{\text{MAC}}^*(R_1, R_2, \epsilon) = \eta(R_1/(1 - \epsilon), R_2/(1 - \epsilon)) \quad (3.30)$$

Proof: Achievability is based on the above scheme and is analyzed in Appendix A.4. The converse is proved in Section 3.4. ■

3.3.4 Numerical Comparison

By Theorems 4 and 5, we can observe the gain in the optimal error exponent induced by the expected-rate constraints:

$$\theta_{\text{MAC}}^*(R_1, R_2, \epsilon) = \theta_{\text{MAC,max}}^*(R_1/(1 - \epsilon), R_2/(1 - \epsilon)). \quad (3.31)$$

This result shows that the rate-boost under expected-rate constraints reported in [20] (see Theorem 2 in this thesis) for a single-link setup extends to networks with multiple links, where all the communication rates are boosted by the factor $(1 - \epsilon)^{-1}$ simultaneously.

We further examine this gain at considering example.

Example 2. *Let Y_0, S, T be independent Bernoulli random variables of parameters $p_{Y_0} = 0.4, p_S = 0.8, p_T$ and set $Y_1 = Y_0 \oplus T$ and $Y_2 = Y_1 \oplus S$.*

For this example, Figure 3.6 shows the optimal error exponents under expected-rate and maximum-rate constraints, $\theta_{\text{MAC}}^(R_1, R_2, \epsilon)$ and $\theta_{\text{MAC,max}}^*(R_1, R_2)$, when $p_T = 0.8, \epsilon = 0.05$ and both links are of same rates $R_1 = R_2$. Figure 3.6 also presents the optimal error exponent under expected-rate constraints*

$\theta_{\text{MAC}}^*(R_1 = 0, R_2, \epsilon)$ when $R_1 = 0$, i.e., when the first sensor is not present or cannot communicate. The figure thus illustrates the benefits of expected-rate constraints (the gap between the solid blue line and the dash-dotted red line) and of the first sensor T_0 (the gap between the dashed green line and the solid blue line).

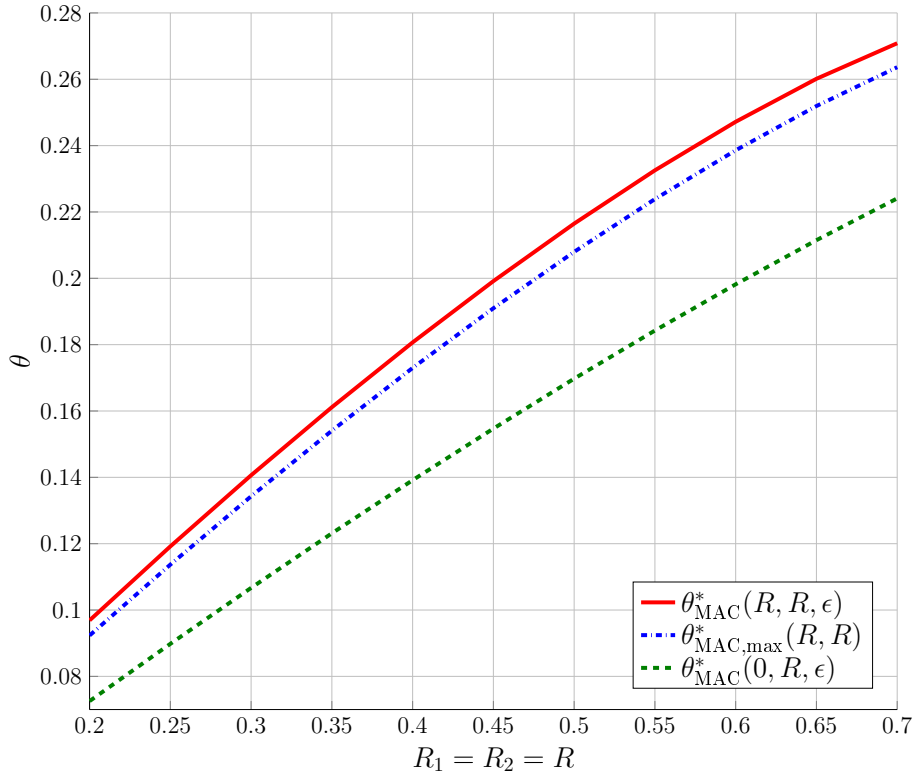


Figure 3.6: Optimal exponents under expected-rate and maximum-rate constraints for Example 2 when $p_T = 0.8$ and $\epsilon = 0.05$.

Remark 2. Even under the Markov chain (3.1), the direct link between the first terminal T_0 and the decision center R_2 is beneficial for the optimal error exponent. To verify this fact, consider the case when $Y_1 = Y_0$, then T_0 and T_1 can be seen as a combined terminal. In the absence of a direct link between T_0 and R_2 , the resulting optimal error exponent is identical to that of a single sensor with a single decision center and a communication rate R_2 , i.e. $\eta_1(R_2/(1-\epsilon))$. In the presence of the direct link between T_0 and R_2 , we can achieve an error exponent equal to $\eta_1\left(\frac{R_1+R_2}{1-\epsilon}\right)$ by having a unified codebook at the combined terminal with total communication rate $R_1 + R_2$. By the non-decreasing monotonicity of the function $\eta_1(\cdot)$ (see Lemma 1, Appendix A.1), we can deduce that

$$\eta_1\left(\frac{R_1 + R_2}{1 - \epsilon}\right) \geq \eta_1\left(\frac{R_2}{1 - \epsilon}\right). \quad (3.32)$$

For verification, a numerical simulation is presented in Figure 3.7 for Example 2 when $p_T = 1$, $\epsilon = 0.05$

and $R_1 = R_2$.

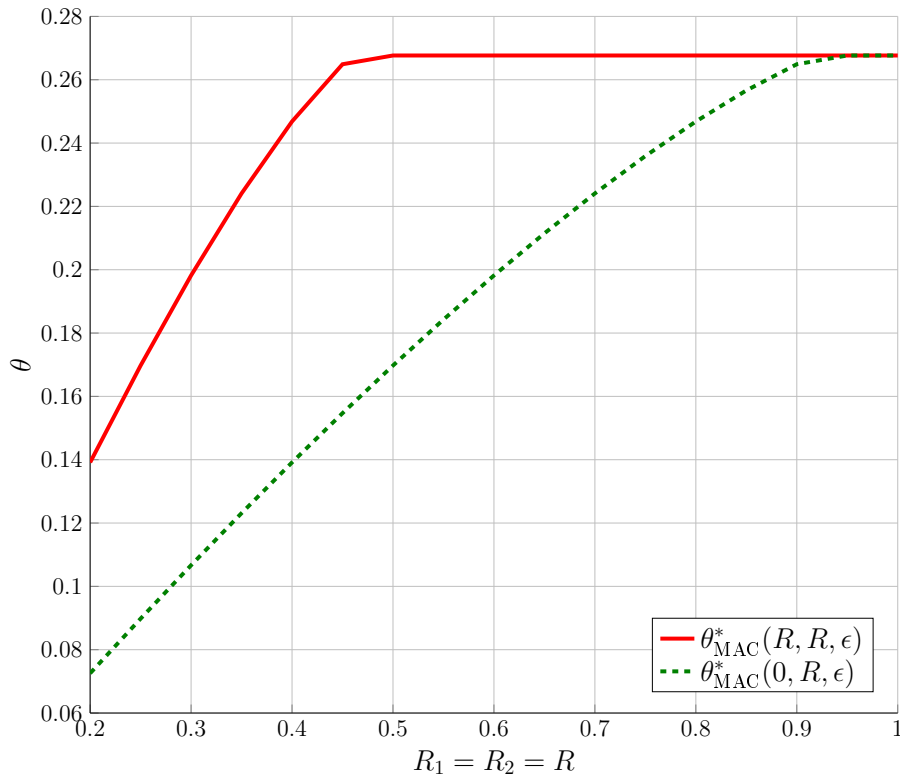


Figure 3.7: Optimal exponents under expected-rate constraints for Example 2 when $Y_1 = Y_0$, $p_T = 1$, and $\epsilon = 0.05$.

3.4 Converse Proof to Theorem 5

3.4.1 Outline of the Converse Proof

In this converse proof, we provide first an equivalent characterization of (3.23), then we propose a lemma that uses the change of measure arguments, the blowing-up lemma, and one asymptotic Markov chain. In addition, we use other arguments such as laws of probability on expectation and standard inequalities and steps to derive lower bounds on the rate constraints and an upper bound on the optimal error exponent. The proof is concluded with continuity and convergence arguments.

3.4.2 Equivalent Characterization

Notice first that the Markov chain $U_2 \rightarrow (U_1, Y_1) \rightarrow (Y_0, Y_2)$, used in Theorem 5, and defined in Theorem 3 for the characterization of the error-exponent (3.23), can be replaced by the weaker Markov chain $U_2 \rightarrow$

$(U_1, Y_1) \rightarrow Y_2$ because the right-hand side (RHS) of (3.23) does not depend on the joint pmf of U_2 and Y_0 . In other words, we have the equivalence

$$\bigcup_{\substack{U_1, U_2: \\ U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2) \\ U_2 \rightarrow (U_1, Y_1) \rightarrow (Y_2, Y_0)}} (I(U_1 U_2; Y_2), I(U_1; Y_0), I(U_2; Y_1 | U_1)) = \bigcup_{\substack{U_1, U_2: \\ U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2) \\ U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2}} (I(U_1 U_2; Y_2), I(U_1; Y_0), I(U_2; Y_1 | U_1)). \quad (3.33)$$

Since the two objective functions coincide and the constraints on the left-hand side (LHS) are more stringent, it suffices to show that the RHS is included in the LHS. To this end, fix U_1, U_2 satisfying the constraints on the LHS, i.e., the Markov chains $U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ and $U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2$. Then, construct \tilde{U}_1, \tilde{U}_2 such that

$$P_{\tilde{U}_1 | Y_0 Y_1 Y_2}(u_1 | y_0, y_1, y_2) = P_{U_1 | Y_0}(u_1 | y_0) \quad (3.34)$$

$$P_{\tilde{U}_2 | \tilde{U}_1 Y_0 Y_1 Y_2}(u_2 | u_1, y_0, y_1, y_2) = P_{U_2 | U_1 Y_1}(u_2 | u_1, y_1), \quad (3.35)$$

and thus satisfying the Markov chains on the RHS: $\tilde{U}_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ and $\tilde{U}_2 \rightarrow (\tilde{U}_1, Y_1) \rightarrow (Y_2, Y_0)$.

The proof is concluded by noting that

$$I(\tilde{U}_1; Y_0) = I(U_1; Y_0), \quad (3.36)$$

$$I(\tilde{U}_2; Y_1 | \tilde{U}_1) = I(U_2; Y_1 | U_1), \quad (3.37)$$

$$I(\tilde{U}_1 \tilde{U}_2; Y_2) = I(U_1 U_2; Y_2), \quad (3.38)$$

where equations (3.36) and (3.37) hold trivially by construction, and (3.38) holds because $P_{\tilde{U}_1 Y_1} = P_{U_1 Y_1}$ and $P_{\tilde{U}_2 Y_2 | \tilde{U}_1 Y_1} = P_{U_2 | U_1 Y_1} \cdot P_{Y_2 | Y_1} = P_{U_2 Y_2 | U_1 Y_1}$.

We proceed to show that

$$\theta \leq \tilde{\eta}(R_1/(1 - \epsilon), R_2/(1 - \epsilon)) \quad (3.39)$$

where $\tilde{\eta}(\cdot, \cdot)$ is defined as

$$\tilde{\eta}(R_1, R_2) := \max_{\substack{P_{U_1 | Y_0}, P_{U_2 | U_1 Y_1}: \\ R_1 \geq I(U_1; Y_0) \\ R_2 \geq I(U_2; Y_1 | U_1) \\ |\mathcal{U}_1| \leq |\mathcal{Y}_0| + 2 \\ |\mathcal{U}_2| \leq |\mathcal{U}_1| |\mathcal{Y}_1| + 1}} I(U_1 U_2; Y_2) \quad (3.40)$$

with the auxiliary random variables U_1 and U_2 satisfying the Markov chains $U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ and $U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2$.

3.4.3 Change of Measure, Blowing Up, and Asymptotic Markov Chain

Consider a sequence (in n) of encoding and decision functions $\{(\phi_1^{(n)}, \phi_2^{(n)}, g^{(n)})\}$ satisfying the constraints on the expected rates and error probabilities in (3.10). The following lemma is the heart of our converse proof.

Lemma 2. *Fix a small number $\eta > 0$, a blocklength n , and a set $\mathcal{D} \subseteq \mathcal{Y}_0^n \times \mathcal{Y}_1^n$. Let the tuple $(\tilde{\mathbf{M}}_1, \tilde{\mathbf{M}}_2, \tilde{Y}_0^n, \tilde{Y}_1^n, \tilde{Y}_2^n)$ follow the pmf*

$$P_{\tilde{\mathbf{M}}_1 \tilde{\mathbf{M}}_2 \tilde{Y}_0^n \tilde{Y}_1^n \tilde{Y}_2^n}(\mathbf{m}_1, \mathbf{m}_2, y_0^n, y_1^n, y_2^n) \triangleq P_{Y_0^n Y_1^n Y_2^n}(y_0^n, y_1^n, y_2^n) \cdot \frac{\mathbb{1}\{(y_0^n, y_1^n) \in \mathcal{D}\}}{P_{Y_0^n Y_1^n}(\mathcal{D})} \cdot \mathbb{1}\{\phi_1^{(n)}(y_0^n) = \mathbf{m}_1\} \cdot \mathbb{1}\{\phi_2^{(n)}(y_1^n, \phi_1^{(n)}(y_0^n)) = \mathbf{m}_2\}. \quad (3.41)$$

Further, define

$$U_1 \triangleq (\tilde{\mathbf{M}}_1, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, T), \quad (3.42)$$

$$U_2 \triangleq (\tilde{\mathbf{M}}_2, T), \quad (3.43)$$

$$\tilde{Y}_i \triangleq \tilde{Y}_{i,T}, \quad i \in \{0, 1, 2\}, \quad (3.44)$$

where T is uniform over $\{1, \dots, n\}$ and independent of all previously defined random variables. Notice the Markov chain $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$. The following (in)equalities hold

$$H(\tilde{\mathbf{M}}_1) \geq nI(U_1; \tilde{Y}_0) + \log P_{Y_0^n Y_1^n}(\mathcal{D}), \quad (3.45)$$

$$H(\tilde{\mathbf{M}}_2) \geq nI(U_2; \tilde{Y}_1|U_1), \quad (3.46)$$

$$I(U_1; \tilde{Y}_1 \tilde{Y}_2 | \tilde{Y}_0) = \phi_1(n), \quad (3.47)$$

where $\phi_1(n)$ is a function that tends to 0 as $n \rightarrow \infty$.

If

$$\Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, Y_1^n = y_1^n] \geq \eta, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (3.48)$$

then

$$-\frac{1}{n} \log \beta_n \leq I(U_1 U_2; \tilde{Y}_2) + \phi_2(n), \quad (3.49)$$

where $\phi_2(n)$ is a function that tends to 0 as $n \rightarrow \infty$.

Proof: See Appendix A.5. ■

To obtain the above lemma, we condition on a joint set of observations at the two sensors that yield a positive probability of avoiding type-I error at the DC. We intersect this set with the set of jointly typical sequences at the two sensors. For the intersection set, we apply change of measure arguments to

restrict to random variables in this set, on which the blowing-up lemma is then applied to show that the probability of avoiding type-I error tends to 1 under the blown-up set in the limit as $n \rightarrow \infty$. A sequence of standard inequalities is then used to upper-bound the type-II error exponent under this set. The lower-bound expressions on the rates are derived using the laws on expectation. For the desired Markov chains, the first one is proved asymptotically in the limit as $n \rightarrow 0$, while the second one holds by construction. With this lemma, we can prove the desired outer bound on the exponents region as follows.

3.4.4 Bounds on Optimal Error Exponent and Rate Constraints

Fix a positive $\eta > 0$. Set $\mu_n = n^{-1/3}$, and define the sets

$$\mathcal{B}(\eta) \triangleq \{(y_0^n, y_1^n) : \Pr[\hat{\mathcal{H}} = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \geq \eta\}, \quad (3.50)$$

$$\mathcal{D}(\eta) \triangleq \mathcal{B}(\eta) \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}). \quad (3.51)$$

Further define for each n the probability

$$\Delta \triangleq P_{Y_0^n Y_1^n}(\mathcal{D}(\eta)), \quad (3.52)$$

and notice that by the constraint (3.10a) on the type-I error probability

$$1 - \epsilon \leq \sum_{y_0^n, y_1^n} \Pr[\hat{\mathcal{H}} = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (3.53)$$

$$\leq \sum_{(y_0^n, y_1^n) \in \mathcal{B}(\eta)} P_{Y_0^n Y_1^n}(y_0^n, y_1^n) + \sum_{(y_0^n, y_1^n) \notin \mathcal{B}(\eta)} \Pr[\hat{\mathcal{H}} = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (3.54)$$

$$\leq P_{Y_0^n Y_1^n}(\mathcal{B}(\eta)) + \eta(1 - P_{Y_0^n Y_1^n}(\mathcal{B}(\eta))). \quad (3.55)$$

Thus we have

$$P_{Y_0^n Y_1^n}(\mathcal{B}(\eta)) \geq \frac{1 - \epsilon - \eta}{1 - \eta}. \quad (3.56)$$

Moreover, by [44, Remark to Lemma 2.12], the probability that the pair (Y_0^n, Y_1^n) lies in the strongly jointly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})$ satisfies

$$P_{Y_0 Y_1}^n \left(\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) \right) \geq 1 - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{4\mu_n^2 n}, \quad (3.57)$$

and since for any two events A and B ,

$$\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1, \quad (3.58)$$

then by (3.51), (3.52), (3.56), and (3.57), we obtain

$$\Delta \geq \frac{1 - \epsilon - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{4\mu_n^2 n}. \quad (3.59)$$

We thus conclude that in the limit $n \rightarrow \infty$ and $\eta \downarrow 0$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta \geq 1 - \epsilon, \quad (3.60a)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta \leq 1. \quad (3.60b)$$

We proceed by applying Lemma 2 to the set $\mathcal{D}(\eta)$ with $\Delta \geq \eta$. This allows to conclude that there exists a pair (U_1, U_2) satisfying the Markov chain $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$ and the (in)equalities

$$H(\tilde{M}_1) \geq nI(U_1; \tilde{Y}_0) + \log P_{Y_0^n Y_1^n}(\mathcal{D}), \quad (3.61)$$

$$H(\tilde{M}_2) \geq nI(U_2; \tilde{Y}_1 | U_1), \quad (3.62)$$

$$\phi_1(n) = I(U_1; \tilde{Y}_1 \tilde{Y}_2 | \tilde{Y}_0), \quad (3.63)$$

and

$$-\frac{1}{n} \log \beta_n \leq I(U_1 U_2; \tilde{Y}_2) + \phi_2(n), \quad (3.64)$$

where the functions $\phi_1(n), \phi_2(n) \downarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_0, \tilde{Y}_1, \tilde{Y}_2, \tilde{M}_1, \tilde{M}_2$ are defined as in the lemma, when applied to the set $\mathcal{D}(\eta)$.

To simplify exposition, we assume η very small and $\Delta \geq \eta$. Otherwise the proof is similar but omitted here.

Further, define the following random variables

$$\tilde{L}_i \triangleq \text{len}(\tilde{M}_i), \quad i \in \{1, 2\}. \quad (3.65)$$

By the rate constraints (3.4) and (3.6), and the definition of the random variables \tilde{L}_i , we obtain for $i \in \{1, 2\}$

$$nR_i \geq \mathbb{E}[L_i] \geq \mathbb{E}[\tilde{L}_i] \Delta. \quad (3.66)$$

Moreover,

$$H(\tilde{M}_i) = H(\tilde{M}_i, \tilde{L}_i) \quad (3.67)$$

$$= \sum_{l_i} \Pr[\tilde{L}_i = l_i] H(\tilde{M}_i | \tilde{L}_i = l_i) + H(\tilde{L}_i) \quad (3.68)$$

$$\leq \sum_{l_i} \Pr[\tilde{L}_i = l_i] l_i + H(\tilde{L}_i) \quad (3.69)$$

$$= \mathbb{E}[\tilde{L}_i] + H(\tilde{L}_i), \quad (3.70)$$

which combined with (3.66) establishes

$$\Delta H(\tilde{M}_i) \leq \Delta \mathbb{E}[\tilde{L}_i] + \Delta H(\tilde{L}_i) \quad (3.71)$$

$$\leq nR_i \left(1 + h_b \left(\frac{\Delta}{nR_i} \right) \right), \quad (3.72)$$

where (3.72) holds by (3.66) and because the entropy of the discrete and positive random variable \tilde{L}_i of

mean $\mathbb{E}[\tilde{L}_i] \leq \frac{nR_i}{\Delta}$ is bounded by $\frac{nR_i}{\Delta} \cdot h_b\left(\frac{\Delta}{nR_i}\right)$, see [23, Theorem 12.1.1].

Then by combining (3.72) for $i \in \{1, 2\}$ with (3.61) and (3.62), noting (3.59), and considering also (3.64), we have proved so far that for all $n \geq 1$ there exists joint pmf $P_{U_1 U_2 \tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2} = P_{U_1 | \tilde{Y}_0 \tilde{Y}_1} P_{U_2 | U_1 \tilde{Y}_1} P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2}$ (abbreviated as $P^{(n)}$) so that the following conditions hold (where I_P indicates that the mutual information should be calculated according to a pmf P):

$$R_1 \geq (I_{P^{(n)}}(U_1; \tilde{Y}_0) + g_1(n)) \cdot g_2(n, \eta), \quad (3.73a)$$

$$R_2 \geq (I_{P^{(n)}}(U_2; \tilde{Y}_1 | U_1)) \cdot g_2(n, \eta), \quad (3.73b)$$

$$\theta \leq I_{P^{(n)}}(U_1 U_2; \tilde{Y}_2) + g_3(n), \quad (3.73c)$$

$$g_4(n) = I_{P^{(n)}}(\tilde{Y}_1 \tilde{Y}_2; U_1 | \tilde{Y}_0), \quad (3.73d)$$

for some nonnegative functions $g_1(n)$, $g_2(n, \eta)$, $g_3(n)$, $g_4(n)$ with the following asymptotic behaviors

$$\lim_{n \rightarrow \infty} g_1(n) = 0, \quad (3.74)$$

$$\lim_{n \rightarrow \infty} g_3(n) = 0, \quad (3.75)$$

$$\lim_{n \rightarrow \infty} g_4(n) = 0, \quad (3.76)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} g_2(n, \eta) \geq 1 - \epsilon. \quad (3.77)$$

We next observe that by Carathéodory's theorem [48, Appendix C] for each n there must exist random variables U_1, U_2 satisfying (3.73) over alphabets of sizes

$$|\mathcal{U}_1| \leq |\mathcal{Y}_0| \cdot |\mathcal{Y}_1| + 3, \quad (3.78)$$

$$|\mathcal{U}_2| \leq |\mathcal{U}_1| \cdot |\mathcal{Y}_1| + 1. \quad (3.79)$$

Then we invoke the Bolzano-Weierstrass theorem and consider an increasing subsequence of positive numbers $\{n_k\}_{k=1}^{\infty}$ such that the following subsequences converge:

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2 U_1 U_2}^{(n_k)} = P_{Y_0 Y_1 Y_2 U_1 U_2}^*. \quad (3.80)$$

Considering further an appropriate sequence of diminishing η -values, we conclude by (3.73a)–(3.73c) and (3.60) that:

$$R_1 \geq (1 - \epsilon) I_{P^*}(U_1; Y_0), \quad (3.81)$$

$$R_2 \geq (1 - \epsilon) I_{P^*}(U_2; Y_1 | U_1), \quad (3.82)$$

$$\theta \leq I_{P^*}(U_1 U_2; Y_2). \quad (3.83)$$

Notice further that since for any k , the pair $(\tilde{Y}_0^{(n_k)}, \tilde{Y}_1^{(n_k)})$ lies in the jointly typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0 Y_1})$, we have $|P_{\tilde{Y}_0 \tilde{Y}_1} - P_{Y_0 Y_1}| \leq \mu_{n_k}$ and thus the limiting pmf satisfies $P_{\tilde{Y}_0 \tilde{Y}_1}^* = P_{Y_0 Y_1}$. Moreover, since for each n_k the random variable \tilde{Y}_2 is drawn according to $P_{Y_2|Y_1}$ given \tilde{Y}_1 , irrespective of \tilde{Y}_0 , the limiting pmf also satisfies $P_{\tilde{Y}_2|\tilde{Y}_0 \tilde{Y}_1}^* = P_{Y_2|Y_1}$. We also notice that under $P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2 U_1 U_2}^*$ the Markov chain

$$U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2, \quad (3.84)$$

holds because $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$ forms a Markov chain for any n_k . Finally, by continuity considerations and by (3.73d), the following Markov chain must hold under $P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2 U_1 U_2}^*$:

$$U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2). \quad (3.85)$$

Using the definition of the function $\tilde{\eta}(\cdot, \cdot)$ in (3.40), we thus proved that all achievable error exponents θ are upper-bounded by the error exponent given in (3.39). This concludes our converse proof.

3.5 The Two-Hop Model

3.5.1 The Setup

Consider the distributed hypothesis testing problem for the two-hop setup with a single decision center in Figure 3.2. It consists of a transmitter T_0 , a relay R_1 , and a receiver R_2 . Unlike the Cooperative MAC setup of Section 3.3, there is no direct link between T_0 and R_2 . The communication occurs over two hops from T_0 to R_1 , and from R_1 to R_2 . All the terminals T_0 , R_1 , and R_2 have their own observations Y_0^n, Y_1^n and Y_2^n , respectively, forming the Markov chain

$$Y_0^n \rightarrow Y_1^n \rightarrow Y_2^n \quad (3.86)$$

under both null and alternative hypotheses. In this case, the testing problem is a testing-against-independence of all observations. It differs from the previous Cooperative MAC setup where under the alternative hypothesis the observations at the sensors T_0, T_1 are correlated together but are independent of the observations at the receiver R_2 . This two-hop setup is motivated by systems in which terminals are correlated in a serial manner. In this case, if the observations at any two adjacent terminals are independent then all observations are independent from each other. In other words, in the two-hop setup, in the special case of testing against independence, depending on the binary hypothesis $\mathcal{H} \in \{0, 1\}$, the tuple (Y_0^n, Y_1^n, Y_2^n) is distributed as

$$\text{under } \mathcal{H} = 0 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0 Y_1} \cdot P_{Y_2|Y_1}, \quad (3.87a)$$

$$\text{under } \mathcal{H} = 1 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1} \cdot P_{Y_2}, \quad (3.87b)$$

for given pmfs $P_{Y_0 Y_1}$ and $P_{Y_2 | Y_1}$ and where P_{Y_0} , P_{Y_1} , and P_{Y_2} denote the marginals of the joint pmf $P_{Y_0 Y_1 Y_2} := P_{Y_0 Y_1} P_{Y_2 | Y_1}$.

The transmitter T_0 observes the source sequence Y_0^n and sends its bit-string message $\mathbf{M}_1 = \phi_0^{(n)}(Y_0^n)$ to R_1 , where the encoding function is of the form $\phi_0^{(n)} : \mathcal{Y}_0^n \rightarrow \{0, 1\}^*$ and satisfies the rate constraint. For a maximum-rate constraint, the maximum length of \mathbf{M}_1 satisfies

$$\text{len}(\mathbf{M}_1) \leq nR_1, \quad (3.88)$$

whereas for an expected-rate constraint, the expected length of \mathbf{M}_1 satisfies

$$\mathbb{E}[\text{len}(\mathbf{M}_1)] \leq nR_1. \quad (3.89)$$

The relay R_1 observes the source sequence Y_1^n and with the message \mathbf{M}_1 received from T_0 , it computes a bit-string message $\mathbf{M}_2 = \phi_1^{(n)}(Y_1^n, \mathbf{M}_1)$ using some encoding function $\phi_1^{(n)} : \mathcal{Y}_1^n \times \{0, 1\}^* \rightarrow \{0, 1\}^*$ that satisfies the second rate constraint. For a maximum-rate constraint, the maximum length of \mathbf{M}_2 satisfies

$$\text{len}(\mathbf{M}_2) \leq nR_2, \quad (3.90)$$

whereas for an expected-rate constraint, the expected length of \mathbf{M}_2 satisfies

$$\mathbb{E}[\text{len}(\mathbf{M}_2)] \leq nR_2. \quad (3.91)$$

Then R_1 sends \mathbf{M}_2 to the receiver R_2 , which guesses hypothesis \mathcal{H} using its observation Y_2^n and the received message \mathbf{M}_2 , i.e., using a decision function $g_2^{(n)} : \mathcal{Y}_2^n \times \{0, 1\}^* \rightarrow \{0, 1\}$, it produces the guess

$$\hat{\mathcal{H}}_2 = g_2^{(n)}(Y_2^n, \mathbf{M}_2) \in \{0, 1\}. \quad (3.92)$$

The goal is to design encoding and decision functions such that their type-I error probability

$$\alpha_{2,n} \triangleq \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0] \quad (3.93)$$

stays below a given threshold $\epsilon_2 > 0$ and the type-II error probability

$$\beta_{2,n} \triangleq \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1] \quad (3.94)$$

decays to 0 with largest possible exponential decay.

Definition 5. Fix a maximum type-I error probability $\epsilon_2 \in [0, 1]$ and rates $R_1, R_2 \geq 0$. The type-II error exponent θ_2 is called ϵ_2 -achievable if there exists a sequence of encoding and decision functions

$\{\phi_0^{(n)}, \phi_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$ satisfying

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{2,n} \leq \epsilon_2, \quad (3.95a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_{2,n}} \geq \theta_2, \quad (3.95b)$$

and

$$\text{len}(\mathbf{M}_i) \leq nR_i, \quad i \in \{1, 2\}, \quad (3.95c)$$

for the setup under maximum-rate constraints, or

$$\mathbb{E}[\text{len}(\mathbf{M}_i)] \leq nR_i, \quad i \in \{1, 2\}, \quad (3.95d)$$

for the setup under expected-rate constraints.

Definition 6. The supremum over all ϵ_2 -achievable type-II error exponents θ_2 is called the ϵ_2 -optimal error exponent and is denoted $\theta_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon_2)$ for the two-hop one-DC setup under maximum-rate constraints and $\theta_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_2)$ for the two-hop one-DC setup under expected-rate constraints.

In the rest of this chapter, let $\epsilon := \epsilon_2$.

3.5.2 Maximum-Rate Constraints

3.5.2.1 Optimal Coding Scheme under Maximum-Rate Constraints

In this section, we present a simplified version of the optimal coding scheme suggested by [33] for the two-hop setup. This coding scheme can achieve the optimal error exponent $\theta_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon)$ for any $\epsilon \geq 0$. The optimal coding scheme uses the Han's scheme, described in the previous chapter in Section 2.4.2, in the following manner.

Han's scheme is first applied between T_0 and R_1 . If R_1 decides locally on hypothesis $\mathcal{H} = 1$, then R_1 sends the single bit $\mathbf{M}_2 = [0]$ to R_2 , and R_2 declares $\hat{\mathcal{H}}_2 = 1$. If R_1 decides locally on hypothesis $\mathcal{H} = 0$, then R_1 and R_2 run a separate Han scheme, and R_2 produces the guess $\hat{\mathcal{H}}_2$ indicated by this second Han scheme. Details of the coding scheme are as follows.

Fix a blocklength n and choose the following parameters:

- a small positive number $\mu > 0$; and
- conditional pmfs $P_{U_1|Y_0}$ and $P_{U_2|Y_1}$ leading to $P_{Y_0Y_1U_1} := P_{Y_0Y_1}P_{U_1|Y_0}$ and $P_{Y_1Y_2U_2} := P_{Y_1Y_2}P_{U_2|Y_1}$,

where all mutual information quantities are evaluated according to the joint pmfs $P_{Y_0Y_1U_1}$ and $P_{Y_1Y_2U_2}$.

Randomly generate the codebooks

$$\mathcal{C}_{U_1} \triangleq \left\{ u_1^n(m_1) : m_1 \in \left\{ 1, \dots, 2^{n(I(U_1; Y_0) + \mu)} \right\} \right\} \quad (3.96)$$

$$\mathcal{C}_{U_2} \triangleq \left\{ u_2^n(m_2) : m_2 \in \left\{ 1, \dots, 2^{n(I(U_2; Y_1) + \mu)} \right\} \right\}, \quad (3.97)$$

by drawing all entries i.i.d. according to the marginal pmfs P_{U_1} and P_{U_2} .

T₀: Assume it observes $Y_0^n = y_0^n$. Then, it looks for indices m_1 satisfying

$$(u_1^n(m_1), y_0^n) \in \mathcal{T}_\mu^{(n)}(P_{U_1Y_0}),$$

randomly picks one of these indices, and sends its corresponding bit-string

$$\mathbf{M}_1 = [\text{bin}(m_1)]. \quad (3.98)$$

If no such index m_1 exists, then T₀ sends

$$\mathbf{M}_1 = [0]. \quad (3.99)$$

R₁: Assume it observes $Y_1^n = y_1^n$ and receives the bit-string message $\mathbf{M}_1 = \mathbf{m}_1$.

If $\mathbf{m}_1 = [0]$, then it forwards the single-bit message

$$\mathbf{M}_2 = [0]. \quad (3.100)$$

Else it checks if

$$(u_1^n(\mathbf{m}_1), y_1^n) \in \mathcal{T}_\mu^{(n)}(P_{U_1Y_1}). \quad (3.101)$$

If the check fails, it sends $\mathbf{M}_2 = [0]$. If the check is successful, R₁ next looks for indices m_2 satisfying $(u_2^n(m_2), y_1^n) \in \mathcal{T}_\mu^{(n)}(P_{U_2Y_1})$, randomly picks one of them and sends

$$\mathbf{M}_2 = [\text{bin}(m_2)] \quad (3.102)$$

to R₂.

If no such index m_2 exists, R₁ directly sends the single-bit message

$$\mathbf{M}_2 = [0]. \quad (3.103)$$

R₂: Assume it observes the sequence $Y_2^n = y_2^n$ and receives message $\mathbf{M}_2 = \mathbf{m}_2$.

If $\mathbf{m}_2 = [0]$, it declares $\hat{\mathcal{H}}_2 = 1$.

Else it sets $m_2 = \text{dec}(\mathbf{m}_2)$, and checks if

$$(u_2^n(m_2), y_2^n) \in \mathcal{T}_\mu^{(n)}(P_{U_2Y_2}).$$

It declares $\hat{\mathcal{H}}_2 = 0$ if the check succeeds, and $\hat{\mathcal{H}}_2 = 1$ otherwise.

3.5.2.2 Optimal Error Exponent

Recall the definition of the optimal error exponent of Han's scheme in Theorem 1 considering a point-to-point setup with a single T₀–R₁ link as

$$\eta_1(R_1) := \max_{\substack{P_{U_1|Y_0}: \\ R_1 \geq I(U_1; Y_0) \\ U_1 \rightarrow Y_0 \rightarrow Y_1 \\ |\mathcal{U}_1| \leq |\mathcal{Y}_0| + 1}} I(U_1; Y_1), \quad (3.104a)$$

and define similarly the optimal error exponent considering a point-to-point setup with a single R₁–R₂ link as

$$\eta_2(R_2) := \max_{\substack{P_{U_2|Y_1}: \\ R_2 \geq I(U_2; Y_1) \\ U_2 \rightarrow Y_1 \rightarrow Y_2 \\ |\mathcal{U}_2| \leq |\mathcal{Y}_1| + 1}} I(U_2; Y_2), \quad (3.104b)$$

where the mutual information quantities are calculated with respect to the joint pmfs $P_{Y_0 Y_1 U_1}$ and $P_{Y_1 Y_2 U_2}$, respectively. As stated in [1], in above maximization problems it suffices to consider auxiliary random variables U_1 and U_2 over alphabets of sizes $|\mathcal{Y}_0| + 1$ and $|\mathcal{Y}_1| + 1$ by standard applications of Carathéodory's theorem, see [48, Appendix C].

Lemma 3. *The functions η_1 and η_2 are continuous, concave and monotonically nondecreasing on their entire domain \mathbb{R}_0^+ .*

Proof: Analogous to the proof of Lemma 1 and thus omitted. ■

Theorem 6. *For any $\epsilon \geq 0$, the optimal error exponent under the maximum-rate constraints (3.95c) of the two-hop setup is*

$$\theta_{\text{Two-Hop, max}}^*(R_1, R_2, \epsilon) = \eta_1(R_1) + \eta_2(R_2). \quad (3.105)$$

Proof: The achievability proof can be obtained as a special case of the proof in [33] by disregarding the decision at the relay and thus removing the corresponding error probability analysis. The strong converse can be obtained also as a special case of the strong converse of [35], or in a simplified manner as a corollary of the simplified strong converse proof for the two-hop with two DCs setup (to be presented in the next chapter) in Appendix B.1. ■

Notice that the optimal error exponent $\theta_{\text{Two-Hop, max}}^*(R_1, R_2, \epsilon)$ does not depend on the permissible type-I error probability ϵ . Thus, we abbreviate it as $\theta_{\text{Two-Hop, max}}^*(R_1, R_2)$.

By definition, note that $\eta_1(R_1)$ determines the optimal exponent in a point-to-point system where R_2 is not present and R_1 is a decision center, and $\eta_2(R_2)$ determines the optimal exponent in a point-to-point system where T_0 is not present [1]. In the studied two-hop setup, the receiver R_2 thus accumulates the optimal exponents achieved over the two links. As we see next, the gains under expected-rate constraints, that were shown for the setups of single sensor and of cooperative two sensors with a single DC, are extended here for the two-hop setup with a single DC. First, we show the optimal coding scheme under expected-rate constraints for the two-hop setup with a single DC.

3.5.3 Expected-Rate Constraints

3.5.3.1 Optimal Coding Scheme under Expected-Rate Constraints

The main idea of this optimal coding scheme is similar to that of Section 3.3.3.1 where here the three terminals T_0, R_1, R_2 multiplex two different subschemes, and the choice of which subscheme to use depends on the transmitter T_0 's observations Y_0^n . In one subscheme, R_2 attempts to correctly guess the hypothesis \mathcal{H} , while in the other subscheme it simply declares $\hat{\mathcal{H}}_2 = 1$.

The structure of the coding scheme is illustrated in Figure 3.8. We partition the set \mathcal{Y}_0^n into subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0}^n$ the probability of subset $\mathcal{D}_{\{2\}}$ is as large as possible but satisfies

$$\Pr [Y_0^n \in \mathcal{D}_{\{2\}}] \leq 1 - \epsilon. \quad (3.106)$$

As a consequence

$$\Pr [Y_0^n \in \mathcal{D}_\emptyset] \geq \epsilon. \quad (3.107)$$

Notice that as $n \rightarrow \infty$ the above inequalities turn into equalities.

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset or $\mathcal{D}_{\{2\}}$, the three terminals follow a different subscheme.

If $Y_0^n \in \mathcal{D}_\emptyset$: In this case, T_0 and R_1 both send the single-bit string messages

$$M_1 = M_2 = [0] \quad (3.108)$$

and R_2 simply declares

$$\hat{\mathcal{H}} = 1. \quad (3.109)$$

Here, $\alpha_{2,n} = 1, \beta_{2,n} = 0$ by (3.109), and $R_1 = R_2 \approx 0$ by (3.108) since the transmission of single-bit messages can hardly change the communication rate for large values of n .

If $Y_0^n \in \mathcal{D}_{\{2\}}$: In this case, R_2 attempts to correctly guess \mathcal{H} based on the transmitted messages. Specif-

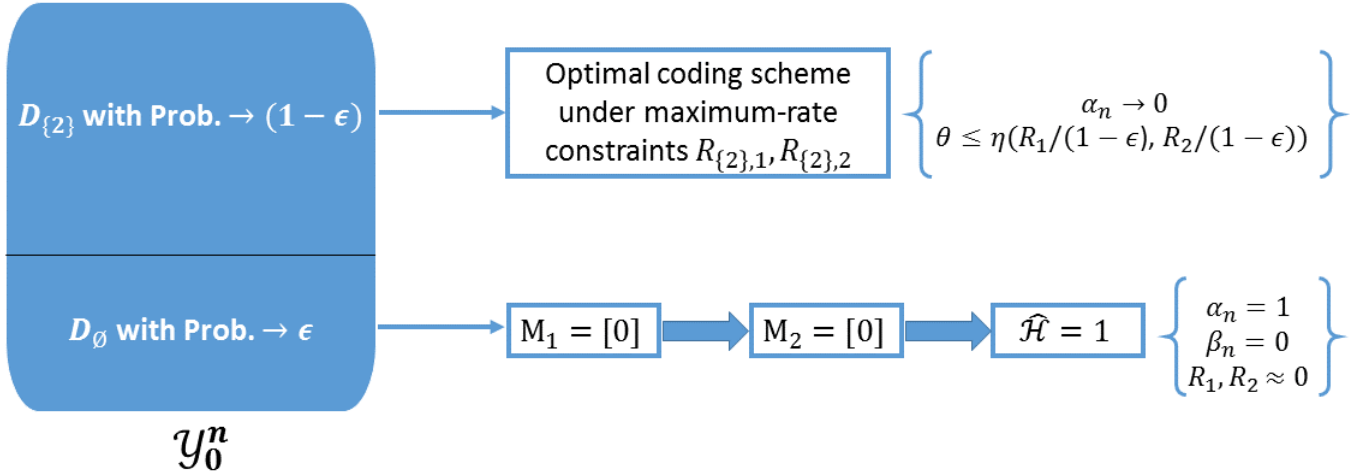


Figure 3.8: Optimal coding scheme for two hops with a single decision center under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on all links.

ically, T_0 , R_1 , R_2 all apply the encoding/decision functions of the optimal hypothesis testing scheme in Subsection 3.5.2.1 with vanishing type-I error probability and respecting maximum-rate constraints $R_{\{2\},1}$ and $R_{\{2\},2}$ on the two links, where these rates are chosen to satisfy

$$(1 - \epsilon)R_{\{2\},1} \leq R_1 \quad (3.110a)$$

$$(1 - \epsilon)R_{\{2\},2} \leq R_2. \quad (3.110b)$$

Analysis: By (3.106) and the choices of the rates in each of the two subschemes, the overall scheme satisfies the expected-rate constraints R_1 and R_2 on the two links for large values of n . Following similar steps as in Appendix A.4, we can prove that when the hypothesis testing scheme in Subsection 3.5.2.1 with vanishing type-I error probability is employed for $Y_0^n \in \mathcal{D}_{\{2\}}$, then the overall scheme meets the permissible type-I error probability ϵ and achieves the optimal error exponent given by Equation 3.111 of Theorem 7.

3.5.3.2 Optimal Error Exponent

Theorem 7. For any $\epsilon \geq 0$, the optimal error exponent under the expected-rate constraints (3.95d) is

$$\theta_{\text{Two-Hop}}^*(R_1, R_2, \epsilon) = \eta_1(R_1/(1 - \epsilon)) + \eta_2(R_2/(1 - \epsilon)). \quad (3.111)$$

Proof: Follows as a corollary from the more general result in Theorem 9 ahead. ■

3.6 Summary and Discussion

In this chapter, we studied testing against independence for two setups with two sensors and a single decision center: the cooperative MAC and the two-hop network. We derived the optimal error exponents under expected-rate constraints showing a rate-gain of $(1 - \epsilon)^{-1}$ on all communication links, thus extending the single-link result in [20]. To achieve this gain, we propose optimal coding schemes that multiplex two subschemes, one of them is a degenerate scheme where the communication is of zero-rate and the decision center always declares $\hat{\mathcal{H}} = 1$, and the other one is the optimal coding scheme for the setup under maximum-rate constraints and for vanishing type-I error probability.

Notice that for the overall scheme to respect the expected-rate constraints, it suffices that the second subscheme respects the rate constraints $R_{\{2\},1}$ and $R_{\{2\},2}$ on expectation. However, as a consequence of our main results in Theorem 5 and Theorem 7, under vanishing type-I error probabilities, the same type-II error exponents are achievable under both expected- and maximum-rate constraints. There is thus no benefit in multiplexing schemes with expected rates $R_{\{2\},1}$ and $R_{\{2\},2}$, but possibly larger maximum rates. This technical note holds also for the next two chapters.

The proposed multiplexing strategy can achieve a rate-boost of $(1 - \epsilon)^{-1}$ on any generic network with a single DC and where all terminals can coordinate the multiplexing, i.e., for example when one terminal can send a zero-rate message to all other terminals in the network. This implies that the suggested strategy is general, simple and practical with promising gains for setups with one DC, and it leads to optimal results when the optimal coding scheme under maximum-rate constraints and vanishing type-I error probability exists. For setups with more than one DC, as we show in next chapters, this strategy might not be sufficient and thus we propose new strategies based on a rate-sharing concept.

In our converse proofs, we used change of measure arguments, the blowing-up lemma, asymptotic Markov chains, and probability laws under expectation. Our proof methods can also be used to establish strong converse results for the cooperative MAC setup under maximum-rate constraints and under the Markov chain (3.1). Recall that strong converse results show that the optimal error exponent is independent of the type-I error probability threshold ϵ (Theorem 4). For the two-hop network, a similar strong converse result existed already in the literature [35] (Theorem 6 in this chapter). However as we show in the next chapter (see Sections 4.3.2.2, 4.4.2.3), our strong converse proof method is simpler.

Chapter 4

Multi-Hop Networks with Multiple DCs

4.1 Introduction

In this chapter, we consider multi-hop networks with multiple decision centers and derive the optimal type-II error exponents region that all DCs can simultaneously achieve under expected-rate constraints. We first characterize the exponents regions for the special case of the two-hop network with two DCs and show that the exponents region under expected-rate constraints is boosted compared to the region under maximum-rate constraints. The exponents region has a rectangular shape under maximum-rate constraints, where both DCs can achieve simultaneously their optimal error exponents. A similar exponents region shape is obtained under expected-rate constraints if the two DCs have equal permissible type-I error probabilities, extending thus the rate-gain observed in previous setups with one DC to multiple DCs with equal type-I error thresholds. However, when the DCs have different permissible type-I error probabilities, a tradeoff appears between the different exponents at the DCs because all the DCs are competing for the same communication resources. We generalize all these results to the K -hop network with K DCs. In the special cases of two and three hops, we further simplify the characterizations of the fundamental exponents region. This allows to propose optimal coding schemes with significantly reduced complexity.

4.2 Related Works

The problem of distributed hypothesis testing in a multi-hop network was considered in [33–35]. The authors in [33] focused on the special case of two-hop network with two DCs, while in [34] they presented some extensions to the general K -hop network with K DCs. In addition, the authors in [35] established a

strong converse proof to the optimality result in [33].

For the two-hop network, the authors in [33] proposed two coding schemes, one with binning and one without binning. In the general hypothesis testing setup, binning was shown to improve the characterization of the exponents region. Interestingly, the authors further showed that under a Markov chain between the measurements at the sensor, the relay, and the DC, $Y_0^n \rightarrow Y_1^n \rightarrow Y_2^n$, their coding schemes decouple into two point-to-point coding schemes. Moreover, in the special cases of testing against independence and testing against conditional independence, their schemes without and with binning, respectively, are optimal under maximum-rate constraints and for vanishing type-I error probabilities. In [35], it was further proved that the optimal error exponent for the two-hop network in the special case of testing against independence and under maximum-rate constraints, is independent of the type-I error probability. For the K -hop network, the authors in [34] extended the no-binning scheme to K hops and showed that it is optimal under the Markov chain $Y_0^n \rightarrow Y_1^n \rightarrow Y_2^n \rightarrow \dots \rightarrow Y_K^n$ in the special case of testing against independence under maximum-rate constraints and for vanishing type-I error probabilities.

In our work, we study testing against independence for the two-hop setup where the maximum-rate constraints are replaced by expected-rate constraints [49, 50]. For the K -hop setup, we derive a strong converse to the previous result under maximum-rate constraints [43]. In addition, we describe the optimal coding scheme and characterize the error exponents region under expected-rate constraints [51]. In the sequel, we present our main results on the two-hop setup in Section 4.3 and on the K -hop setup in Section 4.4.

4.3 The Two-Hop Model

4.3.1 The Setup

Consider the two-hop hypothesis testing setup of Section 3.5 with a transmitter T_0 , a relay R_1 , and a receiver R_2 . Here, R_1 also takes a decision $\hat{\mathcal{H}}_1$ as shown in Figure 4.1, i.e., R_1 observes the source sequence

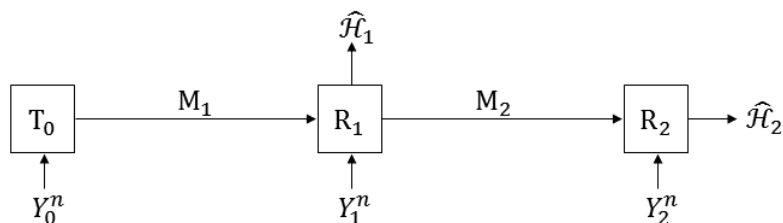


Figure 4.1: Cascaded two-hop setup with two decision centers.

Y_1^n and with the message M_1 received from T_0 , it produces a guess $\hat{\mathcal{H}}_1$ of the hypothesis \mathcal{H} using a decision

function $g_1^{(n)} : \mathcal{Y}_1^n \times \{0, 1\}^* \rightarrow \{0, 1\}$:

$$\hat{\mathcal{H}}_1 = g_1^{(n)}(Y_1^n, M_1) \in \{0, 1\}. \quad (4.1)$$

We define R_1 's type-I error probability as

$$\alpha_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \quad (4.2)$$

and its type-II error probability as

$$\beta_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1]. \quad (4.3)$$

Recall, from Section 3.5, that R_2 produces a guess $\hat{\mathcal{H}}_2$ of the hypothesis \mathcal{H} using a decision function $g_2^{(n)} : \mathcal{Y}_2^n \times \{0, 1\}^* \rightarrow \{0, 1\}$:

$$\hat{\mathcal{H}}_2 = g_2^{(n)}(Y_2^n, M_2) \in \{0, 1\}, \quad (4.4)$$

and its type-I and type-II error probabilities are defined as

$$\alpha_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0], \quad (4.5)$$

$$\beta_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1]. \quad (4.6)$$

Definition 7. Fix maximum type-I error probabilities $\epsilon_1, \epsilon_2 \in [0, 1]$ and rates $R_1, R_2 \geq 0$. The exponent pair (θ_1, θ_2) is called (ϵ_1, ϵ_2) -*achievable* if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)}, \phi_1^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$ satisfying $\forall i \in \{1, 2\}$:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{i,n} \leq \epsilon_i, \quad (4.7a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_{i,n}} \geq \theta_i, \quad (4.7b)$$

and

$$\text{len}(M_i) \leq nR_i, \quad (4.7c)$$

for the setup under maximum-rate constraints, or

$$\mathbb{E}[\text{len}(M_i)] \leq nR_i, \quad (4.7d)$$

for the setup under expected-rate constraints.

Definition 8. The closure of the set of all (ϵ_1, ϵ_2) -achievable exponent pairs (θ_1, θ_2) is called the fundamental (ϵ_1, ϵ_2) -*exponents region* and is denoted $\mathcal{E}_{\text{Two-Hop, max}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ for the setup under maximum-rate constraints and $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ for the setup under expected-rate constraints.

4.3.2 Maximum-Rate Constraints

In this section, we review the fundamental exponents region $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ derived in [33, 35] and the optimal coding scheme that can achieve it.

4.3.2.1 Optimal Coding Scheme under Maximum-Rate Constraints

Apply the same two-hop scheme as in Section 3.5.2.1, but where here, R_1 produces $\hat{\mathcal{H}}_1 = 1$ if $M_1 = [0]$ or the typicality check (3.101) fails, and it produces $\hat{\mathcal{H}}_1 = 0$ otherwise.

4.3.2.2 The Exponents Region

Theorem 8 (Theorem 2 in [35]). *Fix $\epsilon_1, \epsilon_2 \in [0, 1]$. The fundamental exponents region under the maximum-rate constraints (4.7c) is*

$$\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon_1, \epsilon_2) = \{(\theta_1, \theta_2) : \theta_1 \leq \eta_1(R_1), \theta_2 \leq \eta_1(R_1) + \eta_2(R_2)\}, \quad (4.8)$$

where the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$ were defined in Equations (3.104) in Section 7.

Proof: Achievability is proved in [33], and the strong converse in [35]. We present a simplified strong converse proof in Appendix B.1. ■

We notice that the fundamental exponents region does not depend on the permissible type-I error probabilities ϵ_1 and ϵ_2 . We will therefore abbreviate $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ by $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2)$.

In the studied two-hop setup, R_2 thus accumulates the optimal exponents achieved over the two links. Since the exponents region is a rectangle, each of the two DCS, R_1 and R_2 , can simultaneously achieve their optimal exponents, no tradeoff occurs between the two exponents. We shall see that this is not always the case under expected-rate constraints.

4.3.3 Expected-Rate Constraints

4.3.3.1 Optimal Coding Scheme under Expected-Rate Constraints

The optimal coding scheme under expected-rate constraints depends on whether $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$. The general idea of all the three schemes is that the three terminals T_0, R_1, R_2 multiplex two or three different subschemes, and the choice of which subscheme to use depends on the transmitter T_0 's observations y_0^n . To inform all terminals about the choice of the subscheme, T_0 adds one or two flag-bits to its message, which the relay R_1 forwards to the receiver R_2 .

The main distinguishing feature of the different subschemes is the choice of the subset of terminals—either only R_1 or only R_2 , both R_1 and R_2 , or neither of them—which exploit the information in the transmitted messages to produce a guess of hypothesis \mathcal{H} . The other terminals ignore this communication and simply declare $\hat{\mathcal{H}} = 1$. The different subschemes use different communication rates, and as we shall see in Section 4.3.3.2, the allocation of the rates has to be chosen in function of the desired tradeoff between the exponents θ_1 and θ_2 . In this section, we formulate the subschemes based on the optimal hypothesis testing schemes for the two-hop network [33] (see Section 4.3.2.1 in this thesis) and the single-hop network [2] (see Section 2.4.2 in this thesis) with vanishing type-I error probabilities and under maximum-rate constraints, which attains the optimal error exponents presented in Theorem 9 ahead. Note, however, that this multiplexing and rate-sharing strategy can be applied for generic schemes with vanishing type-I error probabilities and respecting given rate constraints and can still show improvement compared to applying one version of the generic schemes alone.

4.3.3.1.1 The case $\epsilon_1 = \epsilon_2 = \epsilon$

We combine two subschemes, where in one subscheme both R_1 and R_2 attempt to correctly guess the hypothesis \mathcal{H} and in the other subscheme both simply declare $\hat{\mathcal{H}} = 1$. To this end, we partition the set \mathcal{Y}_0^n into subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0}^n$ the probability of subset $\mathcal{D}_{\{1,2\}}$ is as large as possible but satisfies

$$\Pr [Y_0^n \in \mathcal{D}_{\{1,2\}}] \leq 1 - \epsilon. \quad (4.9)$$

Notice that as $n \rightarrow \infty$ the inequality turns into an equality.

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset or $\mathcal{D}_{\{1,2\}}$, the three terminals follow a different subscheme. The main structure of the optimal coding scheme is illustrated in Figure 4.2.

If $Y_0^n \in \mathcal{D}_\emptyset$: In this case, none of the terminals attempts to correctly guess the hypothesis \mathcal{H} . Specifically, T_0 and R_1 both send

$$M_1 = M_2 = [0] \quad (4.10)$$

and R_1 and R_2 simply declare

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1. \quad (4.11)$$

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: In this case, both R_1 and R_2 attempt to correctly guess \mathcal{H} based on the transmitted messages. Specifically, T_0, R_1, R_2 all apply the encoding/decision functions of the optimal two-hop hypothesis testing scheme with vanishing type-I error probabilities and respecting maximum-rate constraints

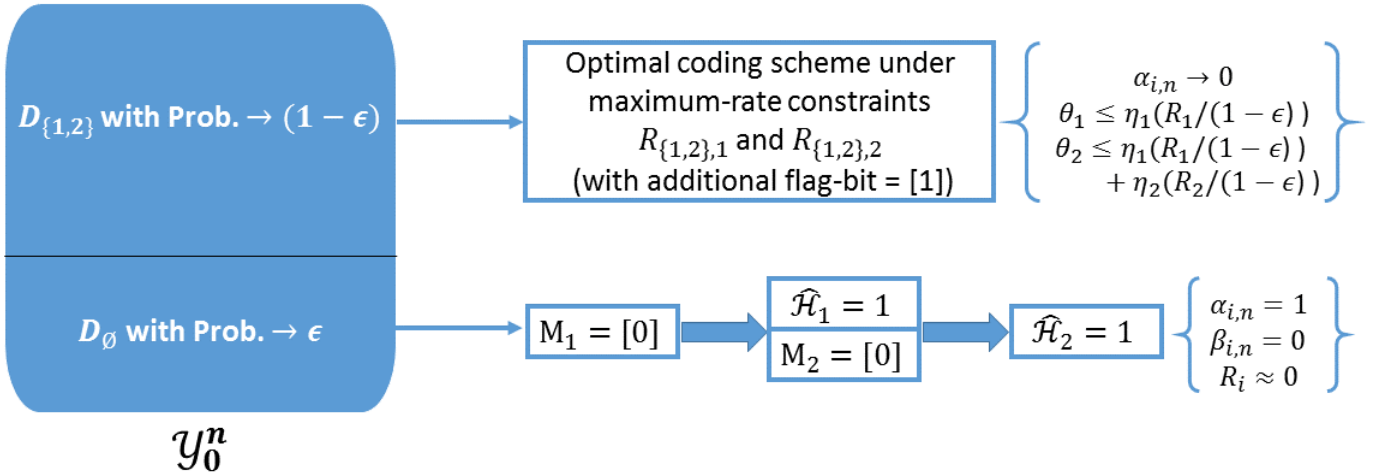


Figure 4.2: Optimal coding scheme for the two-hop network with two DCs of equal type-I error thresholds, $\epsilon_1 = \epsilon_2 = \epsilon$, under expected-rate constraints using the optimal coding scheme under maximum-rate constraints with boosted rates on both links.

$R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ on the two links, where these rates are chosen to satisfy

$$(1 - \epsilon)R_{\{1,2\},1} \leq R_1 \quad (4.12a)$$

$$(1 - \epsilon)R_{\{1,2\},2} \leq R_2. \quad (4.12b)$$

To inform all the terminals about the event $Y_0 \in \mathcal{D}_{\{1,2\}}$ and consequently about the employed scheme, T_0 and R_1 append the $[1]$ -flag at the beginning of their messages M_1 and M_2 .

Analysis: By (4.9) and (4.12), and because transmission of single bits hardly changes the communication rate for large blocklengths, the overall scheme satisfies the expected-rate constraints R_1 and R_2 on the two links. Appendix B.2 proves that when the optimal two-hop hypothesis testing scheme with vanishing type-I error probability [33] is employed for $Y_0^n \in \mathcal{D}_{\{1,2\}}$, then the overall scheme meets the permissible type-I error probability ϵ and achieves the error exponent given by Equation (4.23) of Theorem 9, page 53.

4.3.3.1.2 The case $\epsilon_1 < \epsilon_2$

We combine three subschemes, where in each subscheme either no terminal, only R_1 , or both R_1 and R_2 attempt to correctly guess \mathcal{H} . To this end, we partition the set \mathcal{Y}_0^n into three disjoint subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{1\}}, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under P_{Y_0} the two sets $\mathcal{D}_{\{1\}}$ and $\mathcal{D}_{\{1,2\}}$ have largest possible probabilities but limited by

$$\Pr [Y_0^n \in \mathcal{D}_{\{1\}}] \leq \epsilon_2 - \epsilon_1 \quad (4.13a)$$

$$\Pr [Y_0^n \in \mathcal{D}_{\{1,2\}}] \leq 1 - \epsilon_2. \quad (4.13b)$$

As a consequence,

$$\Pr [Y_0^n \in \mathcal{D}_\emptyset] \geq \epsilon_1. \quad (4.13c)$$

Notice that as $n \rightarrow \infty$, the three inequalities (4.13) can hold with equality.

Choose also nonnegative rates $R_{\{1\},1}$, $R_{\{1,2\},1}$, $R_{\{1,2\},2}$ satisfying

$$(\epsilon_2 - \epsilon_1)R_{\{1\},1} + (1 - \epsilon_2)R_{\{1,2\},1} \leq R_1 \quad (4.14a)$$

$$(1 - \epsilon_2)R_{\{1,2\},2} \leq R_2. \quad (4.14b)$$

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset , $\mathcal{D}_{\{1\}}$, or $\mathcal{D}_{\{1,2\}}$, the three terminals apply a different subscheme satisfying a different pair of maximum-rate constraints, where the subscript \mathcal{I} of set $\mathcal{D}_{\mathcal{I}}$ indicates the set of relays that attempt to correctly guess \mathcal{H} in the event $Y_0^n \in \mathcal{D}_{\mathcal{I}}$. To communicate which of the three subschemes is used, T_0 adds a two-bit flag at the beginning of its message M_1 to R_1 , which forwards this flag at the beginning of its message M_2 to inform R_2 . The main structure of the optimal coding scheme is illustrated in Figure 4.3.

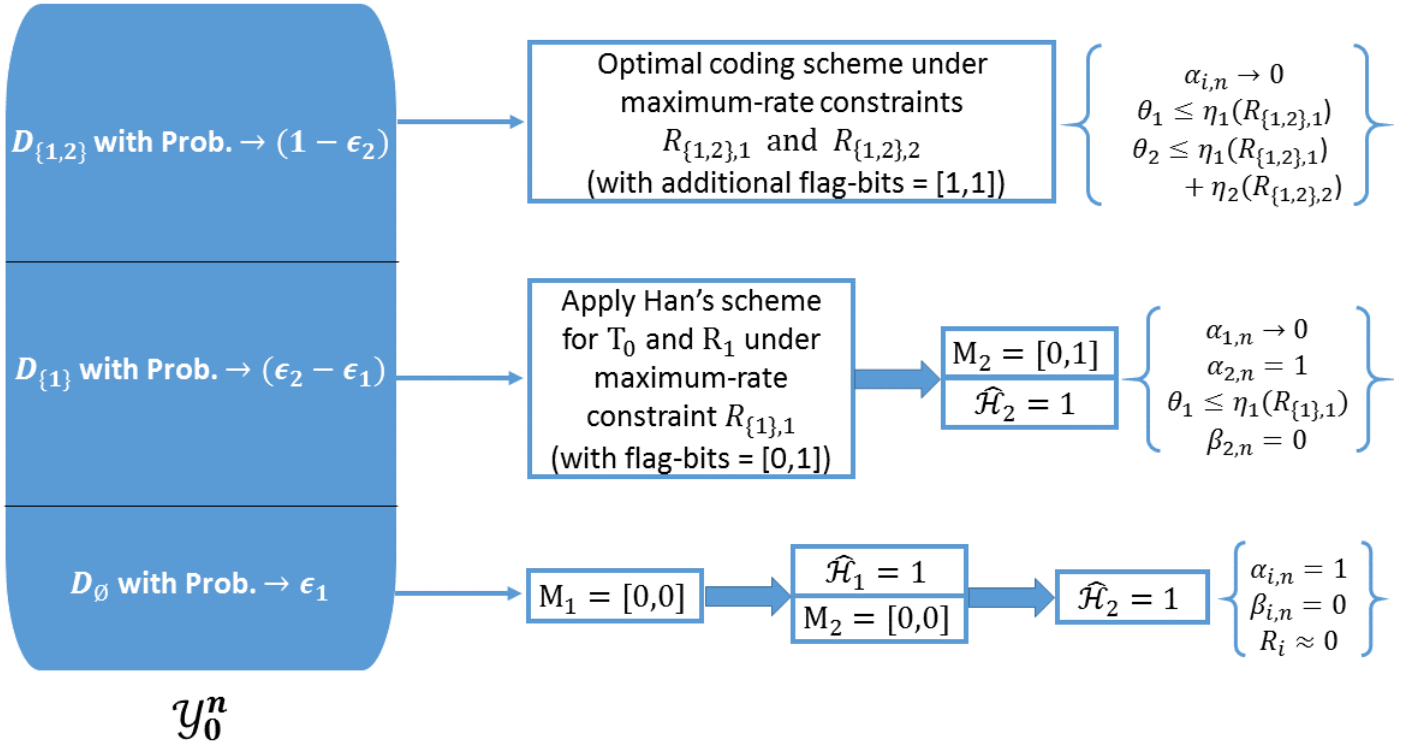


Figure 4.3: Optimal coding scheme for the two-hop network with two DCs of smaller type-I error threshold at the relay, $\epsilon_1 < \epsilon_2$, under expected-rate constraints using Han's scheme and the optimal two-hop scheme under maximum-rate constraints.

If $Y_0^n \in \mathcal{D}_\emptyset$: T_0 and R_1 send only the flag-bits

$$\mathbf{M}_1 = \mathbf{M}_2 = [0, 0] \quad (4.15)$$

and R_1 and R_2 decide on

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1. \quad (4.16)$$

If $Y_0^n \in \mathcal{D}_{\{1\}}$: T_0 and R_1 apply Han's scheme with vanishing type-I error probability and expected-rate constraint $R_{\{1\},1}$ for the bit-string message \mathbf{M}_1 . Moreover, \mathbf{M}_1 is preceded by flag-bits $[0, 1]$, and the relay R_1 forwards these flag-bits to R_2 :

$$\mathbf{M}_2 = [0, 1]. \quad (4.17)$$

Upon reception of these flag-bits, R_2 declares

$$\hat{\mathcal{H}}_2 = 1. \quad (4.18)$$

We observe that, as indicated by the subscript $\{1\}$ of set $\mathcal{D}_{\{1\}}$, only terminal R_1 attempts to correctly guess \mathcal{H} . Receiver R_2 produces the trivial guess in (4.18) because of its higher admissible type-I error probability $\epsilon_2 > \epsilon_1$. Notice also that no communication rate is required for message \mathbf{M}_2 in the limit as $n \rightarrow \infty$.

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: T_0 , R_1 , R_2 apply the optimal two-hop hypothesis testing scheme with vanishing type-I error probabilities and satisfying the expected-rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$.

Analysis: By (4.13) and (4.14), and because transmission of two bits hardly changes the rate for sufficiently large blocklengths, the proposed overall scheme respects the expected-rate constraints R_1 and R_2 for large values of n . Appendix B.3 proves that when the optimal single-hop and two-hop hypothesis testing schemes under maximum-rate constraints $R_{\{1\},1}$ and $(R_{\{1,2\},1}, R_{\{1,2\},2})$ with vanishing type-I error probability [2, 33] are used, then the overall scheme satisfies the type-I error constraints ϵ_1 and ϵ_2 and achieves the error exponents in Equation (4.24) of Theorem 9.

4.3.3.1.3 The case $\epsilon_1 > \epsilon_2$

We combine three subschemes, where in each subscheme either no terminal, only R_2 , or both R_1 and R_2 attempt to correctly guess \mathcal{H} . To this end, we partition the set \mathcal{Y}_0^n into three disjoint subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{2\}}, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0^n}$ the two sets $\mathcal{D}_{\{2\}}$ and $\mathcal{D}_{\{1,2\}}$ have largest possible probabilities but limited by

$$\Pr [Y_0^n \in \mathcal{D}_{\{2\}}] \leq \epsilon_1 - \epsilon_2 \quad (4.19a)$$

$$\Pr [Y_0^n \in \mathcal{D}_{\{1,2\}}] \leq 1 - \epsilon_1. \quad (4.19b)$$

As a consequence,

$$\Pr [Y_0^n \in \mathcal{D}_\emptyset] \geq \epsilon_2. \quad (4.19c)$$

Notice that as $n \rightarrow \infty$, the three inequalities (4.19) hold with equality.

Choose also nonnegative rates $R_{\{2\},1}$, $R_{\{1,2\},1}$, $R_{\{2\},2}$, and $R_{\{1,2\},2}$ satisfying

$$(\epsilon_1 - \epsilon_2)R_{\{1\},1} + (1 - \epsilon_1)R_{\{1,2\},1} \leq R_1 \quad (4.20)$$

$$(\epsilon_1 - \epsilon_2)R_{\{2\},2} + (1 - \epsilon_1)R_{\{1,2\},2} \leq R_2. \quad (4.21)$$

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset , $\mathcal{D}_{\{2\}}$, or $\mathcal{D}_{\{1,2\}}$, the three terminals apply a different subscheme. The subscript \mathcal{I} of set $\mathcal{D}_{\mathcal{I}}$ again indicates the set of terminals that attempt to correctly guess \mathcal{H} in the event $Y_0^n \in \mathcal{D}_{\mathcal{I}}$, and $R_{\mathcal{I},1}, R_{\mathcal{I},2}$ indicate the maximum rates of the subscheme employed under $Y_0^n \in \mathcal{D}_{\mathcal{I}}$. (An exception is the event $Y_0^n \in \mathcal{D}_\emptyset$, where both rates are 0.) Flag-bits are used at the beginning of the messages M_1 and M_2 to inform R_1 and R_2 about which of the subschemes is employed. The main structure of the optimal coding scheme is illustrated in Figure 4.4.

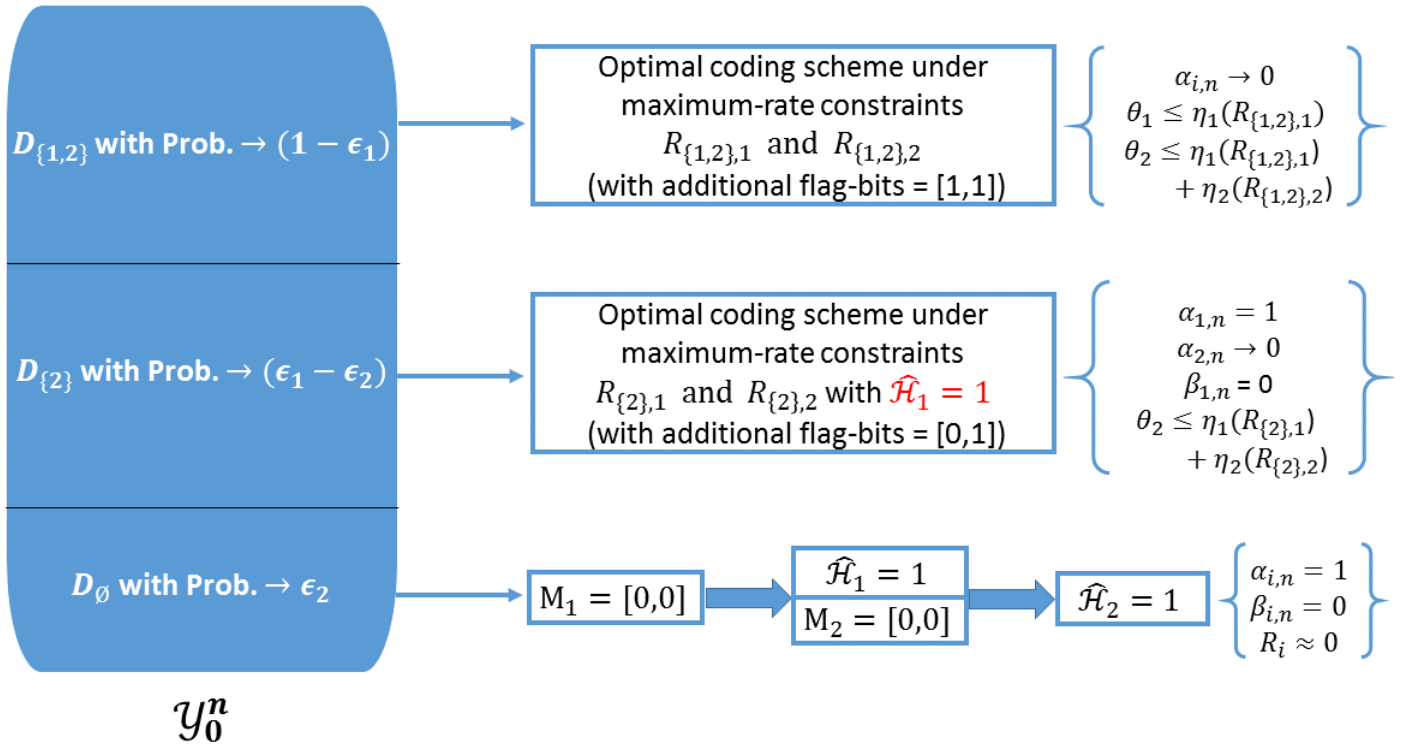


Figure 4.4: Optimal coding scheme for the two-hop network with two DCs of larger type-I error threshold at the relay, $\epsilon_1 > \epsilon_2$, under expected-rate constraints using two versions of the optimal two-hop scheme under maximum-rate constraints.

If $Y_0^n \in \mathcal{D}_\emptyset$: All three terminals, T_0 , R_1 , and R_2 apply the degenerate scheme in (4.15)–(4.16).

If $Y_0^n \in \mathcal{D}_{\{2\}}$: As indicated by the subscript of set $\mathcal{D}_{\{2\}}$, only R_2 makes a serious attempt to correctly guess \mathcal{H} , while R_1 always declares

$$\hat{\mathcal{H}}_1 = 1, \quad (4.22)$$

irrespective of the received message and its observations. This implies that under this subscheme, $\alpha_{1,n} = 1$ and $\beta_{1,n} = 0$. Besides this decision, T_0 , R_1 , and R_2 apply the optimal two-hop distributed hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\{2\},1}$ and $R_{\{2\},2}$ for messages M_1 and M_2 . Moreover, both T_0 and R_1 append the two-bit flag [0,1] at the beginning of these two messages to inform all the terminals about the employed scheme.

Notice that in the optimal two-hop hypothesis testing scheme (see Section 4.3.2.1), the relay R_1 computes a tentative decision based on M_1 and Y_1^n , which influences the message M_2 sent to R_2 and allows the latter to improve its type-I error probability. Here we propose that R_1 itself ignores its tentative decision, because the naive decision (4.22) is sufficient to satisfy the constraint ϵ_1 on its type-I error probability and is also the most-favorable decision to maximize the type-II error exponent.

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: Both DCs at R_1 and R_2 attempt to correctly guess \mathcal{H} . Specifically, T_0 , R_1 , and R_2 apply the optimal two-hop hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ for messages M_1 and M_2 . Moreover, both T_0 and R_1 append the two-bit flag [1,1] at the beginning of these two messages to inform all the terminals about the employed scheme.

Analysis: Similarly to the case $\epsilon_1 < \epsilon_2$, it can be shown that the described scheme respects the expected-rate constraints (4.7d) on both links. Appendix B.4 further shows that when the optimal two-hop scheme [33] is employed, then the described scheme achieves the error exponents in Equation (4.25) of Theorem 9.

4.3.3.2 The Exponents Region

The fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ has a different form, depending on the three cases $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$.

Theorem 9. *Given $\epsilon_1, \epsilon_2, R_1, R_2 \geq 0$.*

If $\epsilon_1 = \epsilon_2 = \epsilon$, then $\mathcal{E}_{\text{Two-Hop}}^(R_1, R_2, \epsilon, \epsilon)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying*

$$\theta_1 \leq \eta_1(R_1/(1 - \epsilon)) \quad (4.23a)$$

$$\theta_2 \leq \eta_1(R_1/(1 - \epsilon)) + \eta_2(R_2/(1 - \epsilon)). \quad (4.23b)$$

If $\epsilon_1 < \epsilon_2$, then $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \min \{ \eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1}) \} \quad (4.24a)$$

$$\theta_2 \leq \eta_1(R_{\{1,2\},1}) + \eta_2(R_2/(1 - \epsilon_2)), \quad (4.24b)$$

for some rates $R_{\{1\},1}, R_{\{1,2\},1} \geq 0$ so that

$$R_1 \geq (\epsilon_2 - \epsilon_1)R_{\{1\},1} + (1 - \epsilon_2)R_{\{1,2\},1}. \quad (4.24c)$$

If $\epsilon_1 > \epsilon_2$, then $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \eta_1(R_{\{1,2\},1}) \quad (4.25a)$$

$$\theta_2 \leq \min \{ \eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2}) \}, \quad (4.25b)$$

for some rates $R_{\{1,2\},1}, R_{\{2\},1}, R_{\{1,2\},2}, R_{\{2\},2} \geq 0$, so that

$$R_1 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},1} + (1 - \epsilon_1)R_{\{1,2\},1} \quad (4.25c)$$

$$R_2 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},2} + (1 - \epsilon_1)R_{\{1,2\},2}. \quad (4.25d)$$

Proof: Achievability is based on the schemes in Section 4.3.3.1, see Appendices B.2, B.3, B.4 for their analyses. The converse is proved in Section 4.3.5. ■

4.3.4 Comparison between Maximum- and Expected-Rate Constraints

We observe from Theorem 9, that for $\epsilon_1 = \epsilon_2 = \epsilon$, the fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon, \epsilon)$ is a rectangle. Also, compared to the fundamental exponents region under maximum-rate constraints, here the rates are boosted by a factor $(1 - \epsilon)^{-1}$:

$$\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon, \epsilon) = \mathcal{E}_{\text{Two-Hop,max}}^* \left(\frac{R_1}{(1 - \epsilon)}, \frac{R_2}{(1 - \epsilon)} \right). \quad (4.26)$$

In particular, for $\epsilon_1 = \epsilon_2 = 0$ the fundamental exponents regions under maximum- and expected-rates coincide:

$$\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, 0, 0) = \mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2). \quad (4.27)$$

For $\epsilon_1 \neq \epsilon_2$, the fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is not a rectangle, as can be verified by the numerical results in Figures 4.5, 4.6, and 4.7 in the next subsection. In fact, one observes a tradeoff between the two exponents θ_1 and θ_2 , which is driven by the choice of the rates $R_{\mathcal{I},1}, R_{\mathcal{I},2}$ for $\mathcal{I} \in \mathcal{P}(2)$ where $\mathcal{P}(2)$ is the power set of all subsets of $\{1, 2\}$ excluding the emptyset, i.e. $\mathcal{P}(2) =$

$\{\{1\}, \{2\}, \{1, 2\}\}$. More specifically, for $\epsilon_1 < \epsilon_2$ the choice

$$R_{\{1,2\},1} = R_1/(1 - \epsilon_2) \quad (4.28a)$$

$$R_{\{1\},1} = 0 \quad (4.28b)$$

maximizes exponent θ_2 , which then evaluates to

$$\theta_2 = \theta_{2,\max} := \eta_1(R_1/(1 - \epsilon_2)) + \eta_2(R_2/(1 - \epsilon_2)), \quad (4.29)$$

but completely degrades θ_1 to $\theta_1 = 0$. However, such a choice is not necessarily needed to maximize exponent θ_2 , if we can find a smaller rate $R_{\{1,2\},1}^*$ such that

$$R_{\{1,2\},1}^* := \min \{R : \eta_1(R) = \eta_{1,\max}\} < R_1/(1 - \epsilon_2), \quad (4.30)$$

where

$$\eta_{1,\max} := \max_R \eta_1(R).$$

In this case, we can still have a positive rate choice for $R_{\{1\},1}$ given by

$$R_{\{1\},1} = R'_{\{1\},1} := \frac{R_1 - (1 - \epsilon_2)R_{\{1,2\},1}^*}{\epsilon_2 - \epsilon_1}. \quad (4.31)$$

This leads to a positive degraded exponent θ_1 given by

$$\theta_1 = \theta_{1,\text{deg}} := \eta_1(R'_{\{1\},1}) > 0. \quad (4.32)$$

On the other hand, the choice

$$R_{\{1\},1} = R_{\{1,2\},1} = R_1/(1 - \epsilon_1) \quad (4.33)$$

maximizes exponent θ_1 , which then evaluates to

$$\theta_1 = \theta_{1,\max} := \eta_1(R_1/(1 - \epsilon_1)), \quad (4.34)$$

but it degrades θ_2 to

$$\theta_2 = \theta_{2,\text{deg}} := \eta_1(R_1/(1 - \epsilon_1)) + \eta_2(R_2/(1 - \epsilon_2)) < \theta_{2,\max}. \quad (4.35)$$

Varying the rate $R_{\{1,2\},1}$ between the choices in (4.28) and (4.33), (and accordingly varying also rate $R_{\{1\},1}$ to meet (4.24c)) achieves the entire Pareto-optimal boundary of the fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$.

For $\epsilon_1 > \epsilon_2$ the choice

$$R_{\{1,2\},1} = R_1/(1 - \epsilon_1) \quad (4.36a)$$

$$R_{\{2\},1} = 0 \quad (4.36b)$$

maximizes exponent θ_1 , which then evaluates to

$$\theta_1 = \theta_{1,\max}, \quad (4.37)$$

and degrades θ_2 to

$$\theta_2 = \theta'_{2,\text{deg}} := \min \{ \eta_1 (R_1/(1 - \epsilon_1)) + \eta_2(R_{\{1,2\},2}), \eta_2(R_{\{2\},2}) \}, \quad (4.38)$$

for $R_{\{2\},2}$ and $R_{\{1,2\},2}$ satisfying (4.25d). In a similar manner to the case $\epsilon_1 < \epsilon_2$, sometimes we might find a smaller rate $R_{\{1,2\},1}^*$ that maximizes exponent θ_1 , such that

$$R_{\{1,2\},1}^* := \min \{ R : \eta_1(R) = \eta_{1,\max} \} < R_1/(1 - \epsilon_1). \quad (4.39)$$

In this case, we can still have a positive rate for $R_{\{2\},1}$:

$$R_{\{2\},1} = R'_{\{2\},1} := \frac{R_1 - (1 - \epsilon_1)R_{\{1,2\},1}^*}{\epsilon_1 - \epsilon_2}. \quad (4.40)$$

This leads to a degraded but possibly better exponent θ_2 given by

$$\theta_2 = \theta''_{2,\text{deg}} := \min \{ \eta_1 (R_{\{1,2\},1}^*) + \eta_2(R_{\{1,2\},2}), \eta_1 (R'_{\{2\},1}) + \eta_2(R_{\{2\},2}) \}, \quad (4.41)$$

$$\geq \min \{ \eta_1 (R_1/(1 - \epsilon_1)) + \eta_2(R_{\{1,2\},2}), \eta_2(R_{\{2\},2}) \}, \quad (4.42)$$

$$= \theta'_{2,\text{deg}}, \quad (4.43)$$

for $R_{\{2\},2}$ and $R_{\{1,2\},2}$ satisfying (4.25d). Here, (4.42) holds by the monotonicity of the function $\eta_1(\cdot)$, and (4.43) holds by (4.38).

On the other hand, the choice

$$R_{\{2\},1} = R_{\{1,2\},1} = R_1/(1 - \epsilon_2) \quad (4.44a)$$

$$R_{\{2\},2} = R_{\{1,2\},2} = R_2/(1 - \epsilon_2) \quad (4.44b)$$

maximizes exponent θ_2 , which then evaluates to $\theta_2 = \theta_{2,\max}$, but it degrades θ_1 to

$$\theta_1 = \theta_{1,\text{deg}} := \eta_1 (R_1/(1 - \epsilon_2)). \quad (4.45)$$

Varying the rate $R_{\{1,2\},1}$ between the choices in (4.36) and (4.44) (and varying the rates $R_{\{1\},1}, R_{\{1,2\},2}, R_{\{1\},2}$ accordingly), achieves the entire Pareto-optimal boundary of the fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$.

Notice that in our two-hop system with expected-rate constraints, exponents $\theta_{1,\max}$ and $\theta_{2,\max}$ defined in (4.34) and (4.29), are the largest possible exponents achievable at the two DCs, irrespective of the ordering

of ϵ_1 and ϵ_2 . By Theorem 8, they coincide with the optimal exponents under maximum-rate constraints $R_1/(1 - \epsilon_1)$ and $R_2/(1 - \epsilon_1)$ for the two links in case of (4.34), and maximum-rate constraints $R_1/(1 - \epsilon_2)$ and $R_2/(1 - \epsilon_2)$ in case of (4.29). We thus observe that whenever $\epsilon_1 \neq \epsilon_2$, the rate-boosts that *expected-rate* constraints allow to obtain over *maximum-rate* constraints depend on the permissible type-I error probabilities and also on the tradeoff between the two exponents θ_1 and θ_2 . In this view, notice that when the focus is on maximizing θ_2 , then for $\epsilon_1 < \epsilon_2$ one might have to entirely sacrifice θ_1 , whereas for $\epsilon_1 > \epsilon_2$ positive θ_1 -exponents are possible but the rate-boost experienced by θ_1 is reduced from $(1 - \epsilon_1)^{-1}$, which is the boost experienced for its maximum $\theta_{1,\max}$, to the smaller factor $(1 - \epsilon_2)^{-1}$.

4.3.4.1 Numerical Simulations

We illustrate the benefits of exploiting the relaxed expected-rate constraints in (4.7d) compared to the more stringent maximum-rate constraints (4.7c) at hand of some examples. We also show for $\epsilon_1 < \epsilon_2$, the benefits of “Rate-sharing” on the first link and the corresponding tradeoff, where the rate R_1 is split into $(\epsilon_2 - \epsilon_1)R_{\{1\},1}$ and $(1 - \epsilon_2)R_{\{1,2\},1}$ as in (4.24), instead of restricting to a single rate choice for the communication on the first link $R_{\{1,2\},1} = R_1/(1 - \epsilon_1)$. For $\epsilon_1 < \epsilon_2$, “Rate-sharing” on the second link does not have any added value. However, for the case $\epsilon_1 > \epsilon_2$, we illustrate the benefits of “Rate-sharing” on both links and the resulting tradeoff from varying the choices of the rates $R_{\{1,2\},1}$, $R_{\{2\},1}$, $R_{\{1,2\},2}$ and $R_{\{2\},2}$ that satisfy (4.25), which stems from multiplexing three coding subschemes among which we have two full versions of the basic two-hop scheme and one degenerate subscheme as explained in Subsection 4.3.3.1.3.

Throughout this section we consider the following example.

Example 3. Let Y_0, S, T be independent Bernoulli random variables of parameters $p_{Y_0} = 0.4, p_S = 0.8, p_T = 0.8$ and set $Y_1 = Y_0 \oplus T$ and $Y_2 = Y_1 \oplus S$.

We first consider the case $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$, and plot the fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ in Figure 4.5 for symmetric rates $R_1 = R_2 = 0.5$. We note a tradeoff between the type-II error exponents θ_1 and θ_2 , which is not present neither for the case $\epsilon_1 = \epsilon_2$, nor for the same setup under maximum-rate constraints. (This tradeoff occurs because both exponents have to be optimized over the same choices of rates $R_{\{1\},1}, R_{\{1,2\},1}$.) The figure also shows a sub-optimal version of the exponents region in Theorem 9, where we set $R_{\{1\},1} = R_{\{1,2\},1}$ but still optimize over all choices of $R_{\{1,2\},1}$. We observe that using two different rates $R_{\{1\},1}$ and $R_{\{1,2\},1}$ (i.e., two different versions of the basic two-hop scheme) allows to obtain a better tradeoff between the two exponents. Finally, for comparison, Figure 4.5

also shows the exponents region $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2)$ under maximum-rate constraints, so as to illustrate the gain provided by having the relaxed {expected}-rate constraints instead of maximum-rate constraints.

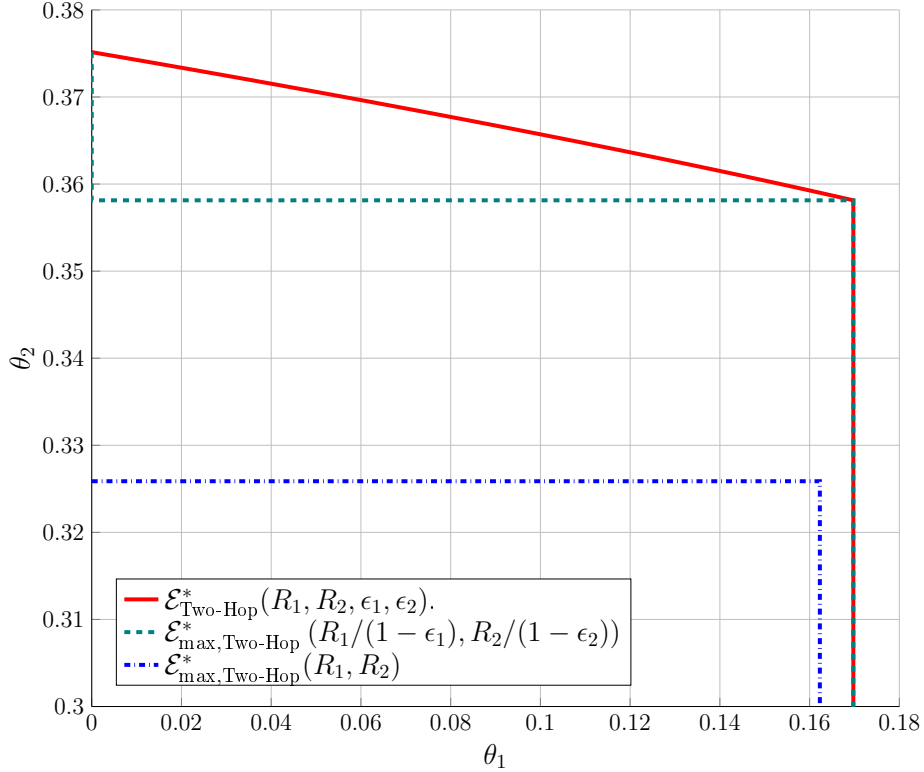


Figure 4.5: Exponents regions for Example 3 when $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$ and $R_1 = R_2 = 0.5$.

We then consider the case $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$. Here we distinguish two categories for the rates of the two links: symmetric rates $R_1 = R_2$, and asymmetric rates $R_1 \neq R_2$. For the first category we consider the sub-case $R_1 = R_2 = 0.5R$, while for the second category we consider a sub-case when R_1 has the larger portion of the total rate R (we consider $R_1 = 0.75R$ and $R_2 = 0.25R$), and a sub-case when R_2 has the larger portion of the total rate R (here we consider $R_1 = 0.25R$ and $R_2 = 0.75R$). In the three mentioned sub-cases, we set the total rate $R = R_1 + R_2 = 1$.

We plot the optimal exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ in Theorem 9, for the first sub-case, in Figure 4.6 and we compare it with the exponents region under maximum-rate constraints $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2)$ together with sub-optimal regions that are achievable under expected-rate constraints. In these sub-optimal regions, we either set $R_{\{1,2\},1} = R_{\{2\},1}$ for which the exponents region coincides with the rectangular region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_2, \epsilon_2) = \mathcal{E}_{\text{Two-Hop,max}}^*\left(\frac{R_1}{(1-\epsilon_2)}, \frac{R_2}{(1-\epsilon_2)}\right)$, or $R_{\{1,2\},2} = R_{\{2\},2}$ for which we have a tradeoff between the type-II error exponents due to rate-sharing on the first link. We observe that rate-sharing on both links (i.e., having two full versions of the basic two-hop scheme) allows to further improve the tradeoff between the two exponents.

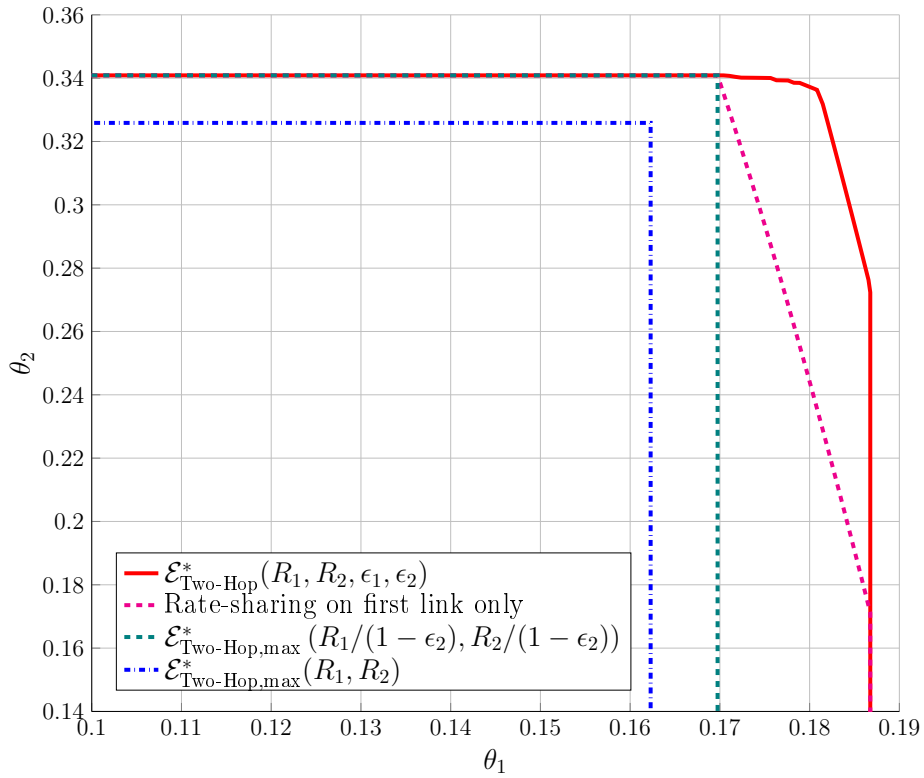


Figure 4.6: Exponents regions under expected- and maximum-rate constraints for Example 3 when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R_1 = R_2 = 0.5$.

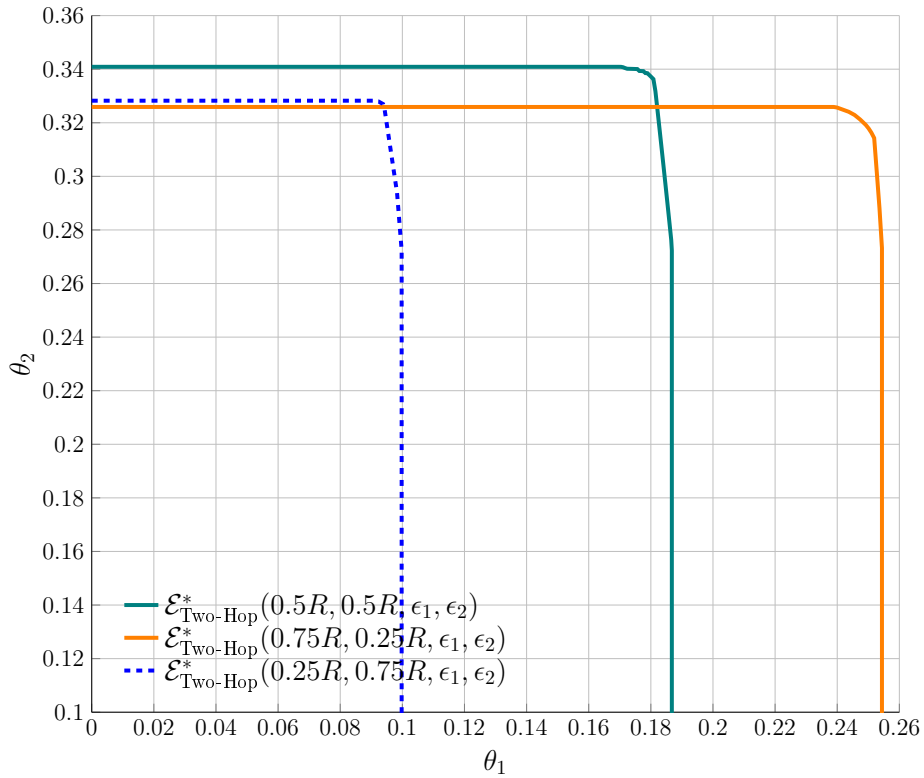


Figure 4.7: Exponents regions under symmetric and asymmetric expected-rate constraints for Example 3 when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R = 1$.

For the other sub-cases corresponding to the asymmetric rates, we plot their optimal exponents region $\mathcal{E}_{\text{Two-Hop}}^*(0.75R, 0.25R, \epsilon_1, \epsilon_2)$ and $\mathcal{E}_{\text{Two-Hop}}^*(0.25R, 0.75R, \epsilon_1, \epsilon_2)$ in Figure 4.7 where we compare them to the optimal exponents region for symmetric rates $\mathcal{E}_{\text{Two-Hop}}^*(0.5R, 0.5R, \epsilon_1, \epsilon_2)$. It is clear that having larger portion of the total rate R dedicated for R_1 leads to larger θ_1 values. However, having larger portion of the total rate R dedicated for R_2 does not necessarily lead to larger θ_2 values.

We mainly observe that in the second sub-case θ_1 values are boosted compared to both other sub-cases and θ_2 values are very close to those of the third sub-case, but are degraded compared to the ones of symmetric rates. In addition, the whole exponents region corresponding to the third sub-case, $\mathcal{E}_{\text{Two-Hop}}^*(0.25R, 0.75R, \epsilon_1, \epsilon_2)$ is totally included in the exponents region of the symmetric rates $\mathcal{E}_{\text{Two-Hop}}^*(0.5R, 0.5R, \epsilon_1, \epsilon_2)$. These observations are in fact justified by the assumed parameters in this example, where one can verify that $\eta_2(\cdot)$ is only slightly larger than $\eta_1(\cdot)$. Thus, allocating the larger portion of the total rate ($0.75R$) to the second link will only lead to a slight increase in the θ_2 values compared to the second sub-case. In fact, the maximum achievable exponents given by the each communication link are given by $\eta_1^*(R_1) = I(Y_0; Y_1) = 0.26766$ and $\eta_2^*(R_2) = I(Y_1; Y_2) = 0.27433$, see Figure 4.8. Recall also that the θ_2 error exponent is an accumulation of the error exponents given by both functions $\eta_1(\cdot) + \eta_2(\cdot)$. The very close values of the two functions for the same rate r , $\eta_1(r)$ and $\eta_2(r)$ ($r \in [0, 1]$), together with the concavity and monotonicity of these functions, induce that to obtain the largest θ_2 values in this example, the total rate needs to be distributed almost equally between both links. In addition, θ_1 error exponent is only dependent on rate R_1 and thus the higher R_1 we have, the higher the values of θ_1 can be. All of the above explains the superiority of the error exponent region obtained when $R_1 = R_2 = 0.5$ over the one obtained when $R_1 = 0.25, R_2 = 0.75$, and the tradeoff between the exponents regions of the sub-cases $R_1 = R_2 = 0.5$ and $R_1 = 0.75, R_2 = 0.25$.

Remark 3. *Note that when $\eta_1(r) \geq \eta_2(r)$, one would expect that both exponents will be boosted by having higher available rate at the first link R_1 , but that might not be true due to the concavity of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$. In this case, the error exponent at the relay θ_1 will be boosted because it only depends on R_1 . However for the error exponent at the receiver θ_2 , distributing the total rate between R_1 and R_2 is better than allocating all the rate to R_1 in some cases. See Figures 4.9 and 4.10 for an example when $\eta_1(r) > \eta_2(r)$ but distributing the total rate into $R_1 = R_2 = 0.5R$ or $R_1 = 0.7R, R_2 = 0.3R$ yields larger error exponents θ_2 than allocating all of the rate to R_1 .*

As we can see in the above remark, an interesting question arises when we have freedom to arbitrarily

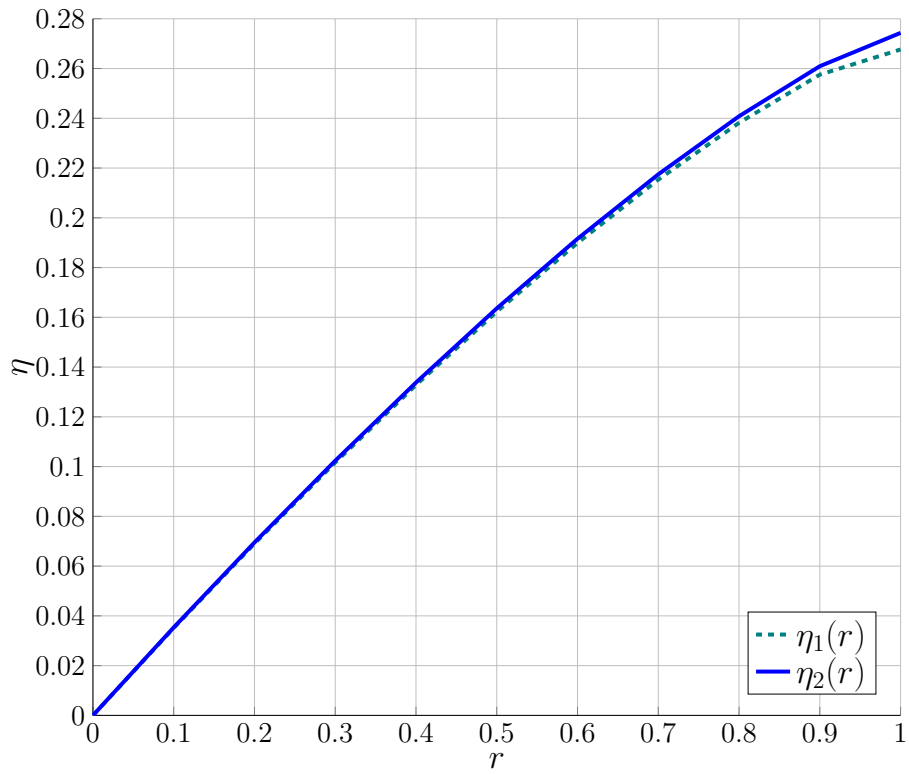


Figure 4.8: Functions $\eta_1(r)$ and $\eta_2(r)$ for Example 3.

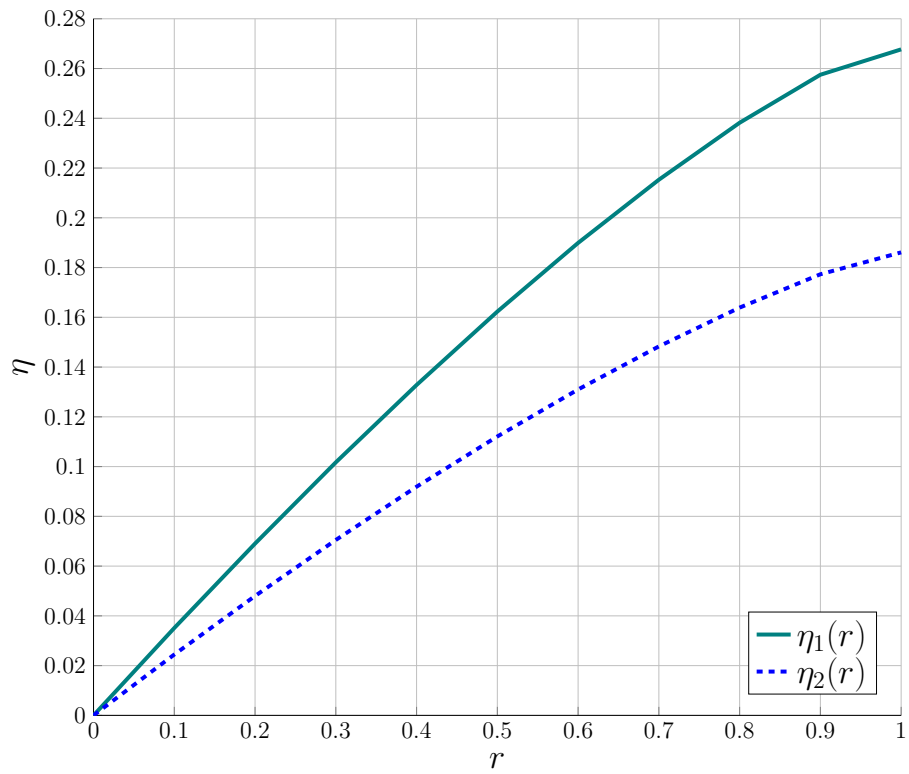


Figure 4.9: Functions $\eta_1(r)$ and $\eta_2(r)$ for Example 3 when $p_{Y_0} = 0.6$, $p_S = 0.75$, and $p_T = 0.8$.

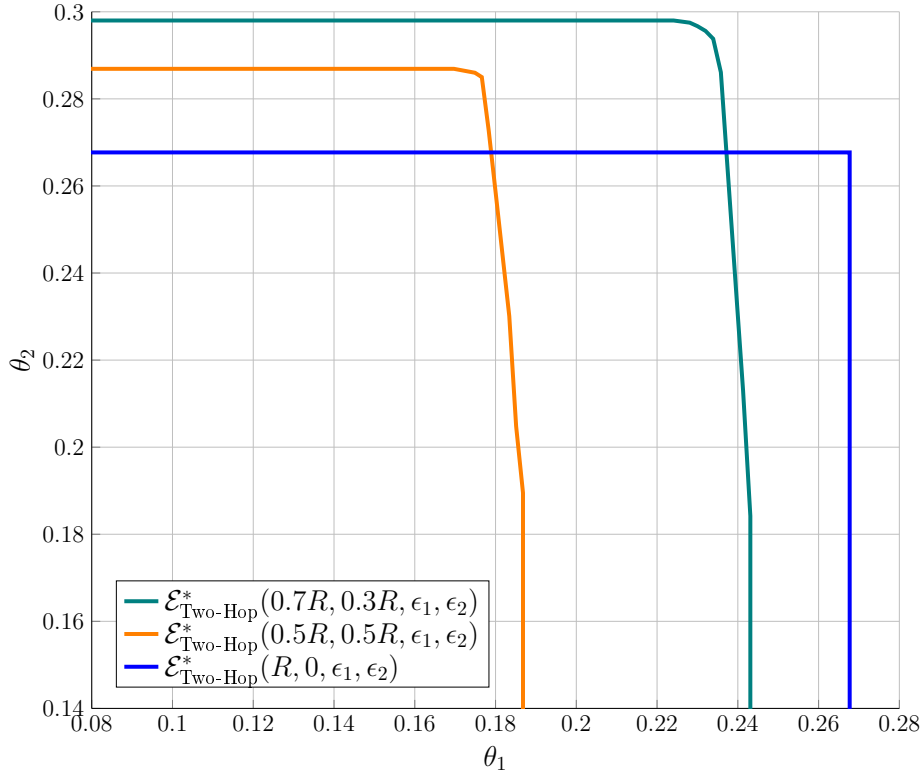


Figure 4.10: Exponents regions under expected-rate constraints subject to different distributions of the total rate $R = R_1 + R_2 = 1$ for Example 3 when $p_{Y_0} = 0.6, p_S = 0.75, p_T = 0.8$ and $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$.

distribute the total rate to the two links: “What is the best rate distribution to maximize the error exponents θ_1 and θ_2 ?” While the answer for maximizing θ_1 is obvious in allocating all the rate to R_1 , the answer is not obvious for θ_2 and requires solving the following optimization problem when $\epsilon_1 > \epsilon_2$

$$\begin{aligned} \max_{\substack{R_1, R_2, R_{\{2\},1}, R_{\{2\},2}, R_{\{1,2\},1}, R_{\{1,2\},2} \geq 0: \\ R_1 + R_2 = R \\ R_1 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},1} + (1 - \epsilon_1)R_{\{1,2\},1} \\ R_2 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},2} + (1 - \epsilon_1)R_{\{1,2\},2}}} \min \{ \eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2}) \}. \end{aligned} \quad (4.46)$$

However when $\epsilon_1 \leq \epsilon_2$, the optimization problem simplifies into:

$$\max_{\substack{R_1, R_2 \geq 0: \\ R_1 + R_2 = R}} \eta_1(R_1/(1 - \epsilon_2)) + \eta_2(R_2/(1 - \epsilon_2)). \quad (4.47)$$

Remark 4. In the special case when $\eta_1(r) = \eta_2(r)$ for all $r \in [0, 1]$, allocating the total rate R to R_1 allows to maximize θ_1 , while distributing the total rate R equally between R_1 and R_2 , i.e. $R_1 = R_2 = 0.5R$ allows to maximize θ_2 . Varying the rate distribution between these two choices of total rate distribution yields the entire Pareto-optimal boundary of the fundamental exponents region when optimizing over sum-rate constraints. These observations follow by concavity and monotonicity arguments of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$.

4.3.5 Converse Proof to Theorem 9

4.3.5.1 Outline of the Converse Proof

The main idea of this proof is to divide the set of strongly jointly typical sequences $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1})$ into four subsets by conditioning on whether the probability of correctly guessing the null hypothesis $\mathcal{H} = 0$ at each decision center is greater or smaller than a small positive number η , and then to apply parallel change of measure arguments under each subset to obtain different lower and upper bounds on the rates and error exponents, see Figure 4.11. To obtain the lower bounds on the rates, we use single-letterization steps under each subset followed by probability laws on expectation to combine the different terms. For the upper bounds on the error exponents, under each subset, we use the blowing-up lemma for each DC with probability of correctly guessing $\mathcal{H} = 0$ larger or equal to η followed by standard single-letterization steps. Then the general upper bound at each DC is given by the minimum between the resulting expressions at each DC. During the process, we choose auxiliary random variables to map the mutual information quantities with the $\eta(\cdot)$ functions, where the auxiliary random variables need to satisfy some Markov chains. However, the change of measure arguments lead to the loss of the initial i.i.d.ness properties that are usually used to prove the required Markov chains. In fact, one Markov chain still holds by construction while we prove the other one holds asymptotically in the limit as $n \rightarrow \infty$.

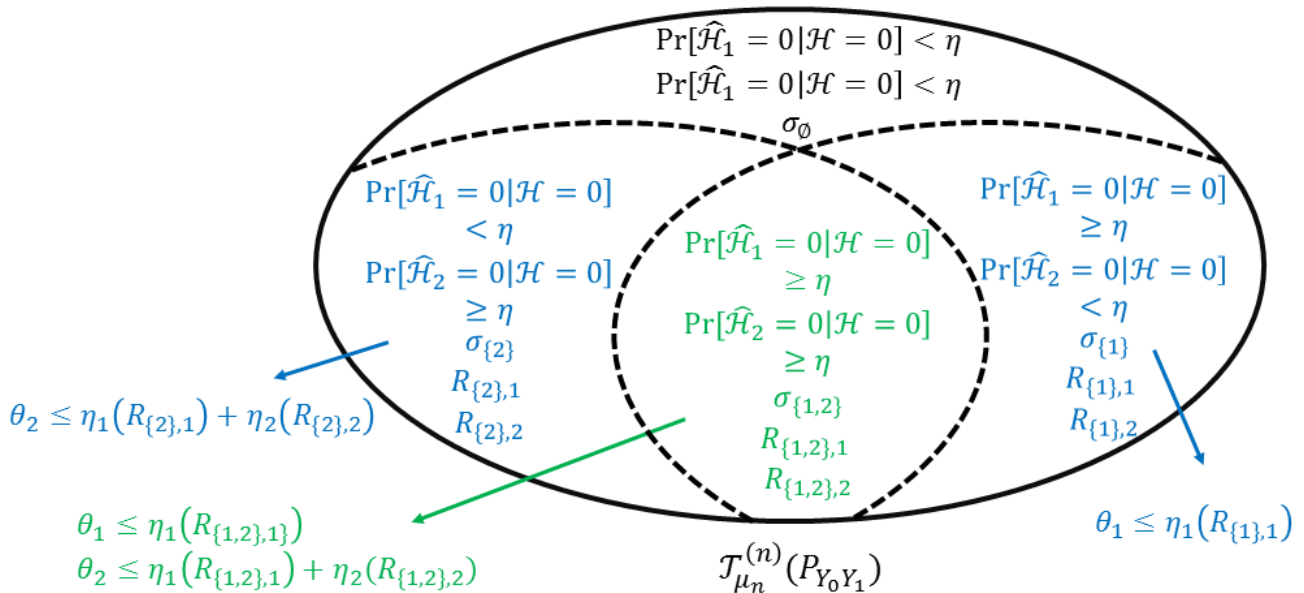


Figure 4.11: Sketch of partitioning $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1})$ and applying parallel change of measure arguments in the converse proof to Theorem 9.

To simplify the presentation of the proof, we first propose, in Subsection 4.3.5.2, the auxiliary Lemma 4

(proved in Appendix B.5) that uses the change of measure arguments and the blowing-up lemma to get the bounds on the rates and exponents. Then we apply this lemma under the different subsets of $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})$ to derive a general outer bound on the exponents region (this lemma does not apply to the set in which both DCs have probability of guessing correctly $\mathcal{H} = 0$ below η). This general bound is valid for all values of ϵ_1, ϵ_2 and is characterized in Proposition 1. In Subsection 4.3.5.3, we simplify the general outer bound depending on the three cases $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$ by showing the optimal values of the probabilities of the different subsets. Indeed, the subset where the DC with larger ϵ -margin is the only one that guesses correctly $\mathcal{H} = 0$ with non-vanishing probability (as $\eta \rightarrow 0$), vanishes (see Figure 4.12 for an illustration in the special case when $\epsilon_1 > \epsilon_2$).

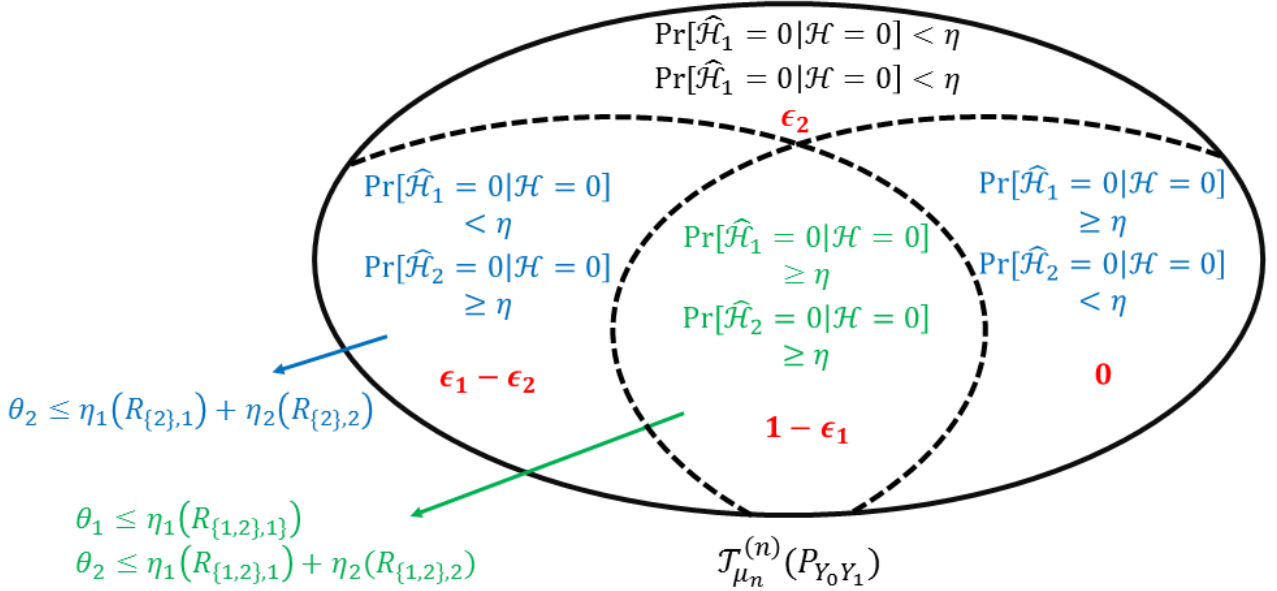


Figure 4.12: Sketch of the simplification in the converse proof to Theorem 9 when $\epsilon_1 > \epsilon_2$.

4.3.5.2 An Auxiliary Lemma and a General Outer Bound

Consider a sequence (in n) of encoding and decision functions $\{(\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ satisfying the constraints on the expected rates and error probabilities in (4.7).

Lemma 4. Fix a small number $\eta > 0$, a blocklength n , and a set $\mathcal{D} \subseteq \mathcal{Y}_0^n \times \mathcal{Y}_1^n$ of probability exceeding η .

Let the tuple $(\tilde{M}_1, \tilde{M}_2, \tilde{Y}_0^n, \tilde{Y}_1^n, \tilde{Y}_2^n)$ follow the pmf

$$P_{\tilde{M}_1 \tilde{M}_2 \tilde{Y}_0^n \tilde{Y}_1^n \tilde{Y}_2^n}(\mathbf{m}_1, \mathbf{m}_2, y_0^n, y_1^n, y_2^n) \triangleq P_{Y_0^n Y_1^n Y_2^n}(y_0^n, y_1^n, y_2^n) \cdot \frac{\mathbb{1}\{(y_0^n, y_1^n) \in \mathcal{D}\}}{P_{Y_0^n Y_1^n}(\mathcal{D})} \cdot \mathbb{1}\{\phi_1^{(n)}(y_0^n) = \mathbf{m}_1\} \cdot \mathbb{1}\{\phi_2^{(n)}(y_1^n, \phi_1^{(n)}(y_0^n)) = \mathbf{m}_2\}. \quad (4.48)$$

Further, define

$$U_1 \triangleq (\tilde{M}_1, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, T) \quad (4.49)$$

$$U_2 \triangleq (\tilde{M}_2, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, T) \quad (4.50)$$

$$\tilde{Y}_i \triangleq \tilde{Y}_{i,T}, \quad i \in \{0, 1, 2\}, \quad (4.51)$$

where T is uniform over $\{1, \dots, n\}$ and independent of all previously defined random variables. Notice the Markov chain $U_2 \rightarrow \tilde{Y}_1 \rightarrow \tilde{Y}_2$. The following (in)equalities hold:

$$H(\tilde{M}_1) \geq nI(U_1; \tilde{Y}_0) + \log P_{Y_0^n Y_1^n}(\mathcal{D}), \quad (4.52)$$

$$H(\tilde{M}_2) \geq nI(U_2; \tilde{Y}_1) + \log P_{Y_0^n Y_1^n}(\mathcal{D}), \quad (4.53)$$

$$I(U_1; \tilde{Y}_1 | \tilde{Y}_0) = \phi_1(n), \quad (4.54)$$

where $\phi_1(n)$ is a function that tends to 0 as $n \rightarrow \infty$.

If

$$\Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, Y_1^n = y_1^n] \geq \eta, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (4.55)$$

then

$$-\frac{1}{n} \log \beta_{2,n} \leq I(U_1; \tilde{Y}_1) + I(U_2; \tilde{Y}_2) + \phi_2(n), \quad (4.56)$$

and if

$$\Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, Y_1^n = y_1^n] \geq \eta, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (4.57)$$

then

$$-\frac{1}{n} \log \beta_{1,n} \leq I(U_1; \tilde{Y}_1) + \phi_3(n), \quad (4.58)$$

where $\phi_2(n), \phi_3(n)$ are functions that tend to 0 as $n \rightarrow \infty$.

Proof: See Appendix B.5. ■

With this lemma, we can prove the desired general outer bound on the exponents region.

Proposition 1. *Given $\epsilon_1, \epsilon_2, R_1, R_2 \geq 0$. The fundamental exponents region $\mathcal{E}_{\text{Two-Hop}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is included in the set of all (θ_1, θ_2) pairs satisfying*

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\}, \quad (4.59a)$$

$$\theta_2 \leq \min\{\eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2})\}, \quad (4.59b)$$

for rates $R_{\{1\},1}, R_{\{1,2\},1}, R_{\{1,2\},2}, R_{\{2\},1}, R_{\{2\},2} \geq 0$ and numbers $\sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}} \geq 0$ so that $\sigma_{\{1\}} + \sigma_{\{2\}} +$

$\sigma_{\{1,2\}} \leq 1$ and

$$\sigma_{\{1\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_1 \quad (4.59c)$$

$$\sigma_{\{2\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_2 \quad (4.59d)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}, \quad (4.59e)$$

and so that the following rate constraints are satisfied:

$$R_1 \geq \sigma_{\{1\}} R_{\{1\},1} + \sigma_{\{1,2\}} R_{\{1,2\},1} + \sigma_{\{2\}} R_{\{2\},1}, \quad (4.59f)$$

$$R_2 \geq \sigma_{\{1,2\}} R_{\{1,2\},2} + \sigma_{\{2\}} R_{\{2\},2}. \quad (4.59g)$$

It can be shown that the outer bound on the fundamental exponents region given in this proposition is tight. We however only need and prove the converse result here.

Proof: Fix a positive $\eta > 0$. Set $\mu_n = n^{-1/3}$, and define the sets

$$\mathcal{B}_1(\eta) \triangleq \{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) : \Pr[\hat{\mathcal{H}}_1 = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \geq \eta\}, \quad (4.60)$$

$$\mathcal{B}_2(\eta) \triangleq \{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) : \Pr[\hat{\mathcal{H}}_2 = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \geq \eta\}, \quad (4.61)$$

$$\mathcal{D}_{\{1,2\}}(\eta) \triangleq \mathcal{B}_1(\eta) \cap \mathcal{B}_2(\eta), \quad (4.62)$$

$$\mathcal{D}_{\{1\}}(\eta) \triangleq \mathcal{B}_1(\eta) \setminus \mathcal{D}_{\{1,2\}}(\eta), \quad (4.63)$$

$$\mathcal{D}_{\{2\}}(\eta) \triangleq \mathcal{B}_2(\eta) \setminus \mathcal{D}_{\{1,2\}}(\eta). \quad (4.64)$$

Further define for each n the probabilities

$$\Delta_{\mathcal{I}} \triangleq P_{Y_0^n Y_1^n}(\mathcal{D}_{\mathcal{I}}(\eta)), \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.65)$$

and notice that by the laws of probability

$$\Delta_{\{1,2\}} + \Delta_{\{1\}} = P_{Y_0^n Y_1^n}(\mathcal{B}_1(\eta)) \quad (4.66)$$

$$\Delta_{\{1,2\}} + \Delta_{\{2\}} = P_{Y_0^n Y_1^n}(\mathcal{B}_2(\eta)) \quad (4.67)$$

$$\Delta_{\{1,2\}} \geq P_{Y_0^n Y_1^n}(\mathcal{B}_1(\eta)) + P_{Y_0^n Y_1^n}(\mathcal{B}_2(\eta)) - 1. \quad (4.68)$$

Now by the type-I error probability constraints (4.7a), we have for $j \in \{1, 2\}$:

$$1 - \epsilon_j \leq \sum_{y_0^n, y_1^n} \Pr[\hat{\mathcal{H}}_j = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (4.69)$$

$$\begin{aligned} &= \sum_{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}} \Pr[\hat{\mathcal{H}}_j = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \\ &+ \sum_{(y_0^n, y_1^n) \notin \mathcal{T}_{\mu_n}^{(n)}} \Pr[\hat{\mathcal{H}}_j = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \end{aligned} \quad (4.70)$$

$$\begin{aligned} &\leq \sum_{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)} \cap \bar{\mathcal{B}}_j(\eta)} \Pr[\hat{\mathcal{H}}_j = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \\ &\quad + \sum_{(y_0^n, y_1^n) \in \mathcal{B}_j(\eta)} P_{Y_0^n Y_1^n}(y_0^n, y_1^n) + \left(1 - P_{Y_0^n Y_1^n}(\mathcal{T}_{\mu_n}^{(n)})\right) \end{aligned} \quad (4.71)$$

$$\leq \eta(1 - P_{Y_0^n Y_1^n}(\mathcal{B}_j(\eta))) + P_{Y_0^n Y_1^n}(\mathcal{B}_j(\eta)) + P_{Y_0^n Y_1^n}(\bar{\mathcal{T}}_{\mu_n}^{(n)}). \quad (4.72)$$

Moreover, by [44, Remark to Lemma 2.12], the probability that the pair (Y_0^n, Y_1^n) lies in the jointly strong typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})$ satisfies

$$P_{Y_0 Y_1}^n \left(\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) \right) \geq 1 - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{4\mu_n^2 n}, \quad (4.73)$$

Thus, by (4.72) and (4.73):

$$P_{Y_0^n Y_1^n}(\mathcal{B}_j(\eta)) \geq \frac{1 - \epsilon_j - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{(1 - \eta)4\mu_n^2 n}, \quad j \in \{1, 2\}, \quad (4.74)$$

and we thus conclude that in the limit $n \rightarrow \infty$ and $\eta \downarrow 0$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} (\Delta_{\{1,2\}} + \Delta_{\{1\}}) \geq 1 - \epsilon_1 \quad (4.75a)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} (\Delta_{\{1,2\}} + \Delta_{\{2\}}) \geq 1 - \epsilon_2 \quad (4.75b)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\} \quad (4.75c)$$

$$\overline{\lim}_{\eta \downarrow 0} \overline{\lim}_{n \rightarrow \infty} \sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} \leq 1. \quad (4.75d)$$

We proceed by applying Lemma 4 to subset $\mathcal{D}_{\mathcal{I}}$ for all $\mathcal{I} \in \mathcal{P}(2)$ with $\Delta_{\mathcal{I}} \geq \eta$. This allows to conclude that for any $\mathcal{I} \in \mathcal{P}(2)$ with $\Delta_{\mathcal{I}} \geq \eta$ there exists a pair $(U_{\mathcal{I},1}, U_{\mathcal{I},2})$ satisfying the Markov chain $U_{\mathcal{I},2} \rightarrow \tilde{Y}_{\mathcal{I},1} \rightarrow \tilde{Y}_{\mathcal{I},2}$ and the (in)equalities

$$H(\tilde{M}_{\mathcal{I},1}) \geq nI(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},0}) + \log P_{Y_0^n Y_1^n}(\mathcal{D}_{\mathcal{I}}), \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.76)$$

$$H(\tilde{M}_{\mathcal{I},2}) \geq nI(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},1}) + \log P_{Y_0^n Y_1^n}(\mathcal{D}_{\mathcal{I}}), \quad \mathcal{I} \in \{\{1, 2\}, \{2\}\}, \quad (4.77)$$

$$\phi_{\mathcal{I},1}(n) = I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1} | \tilde{Y}_{\mathcal{I},0}), \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.78)$$

and

$$-\frac{1}{n} \log \beta_{1,n} \leq I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1}) + \phi_{\mathcal{I},2}(n), \quad \mathcal{I} \in \{\{1\}, \{1, 2\}\}, \quad (4.79)$$

$$-\frac{1}{n} \log \beta_{2,n} \leq I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1}) + I(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},2}) + \phi_{\mathcal{I},3}(n), \quad \mathcal{I} \in \{\{2\}, \{1, 2\}\}, \quad (4.80)$$

where for each \mathcal{I} the functions $\phi_{\mathcal{I},1}(n), \phi_{\mathcal{I},2}(n), \phi_{\mathcal{I},3}(n) \rightarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_{\mathcal{I},0}, \tilde{Y}_{\mathcal{I},1}, \tilde{Y}_{\mathcal{I},2}, \tilde{M}_{\mathcal{I},1}, \tilde{M}_{\mathcal{I},2}$ are defined as in the lemma, when applied to the subset $\mathcal{D}_{\mathcal{I}}$.

To simplify exposition, we assume η very small and $\Delta_{\mathcal{I}} \geq \eta$ for all sets $\mathcal{I} \in \mathcal{P}(2)$. Otherwise the proof

is similar but omitted here.

To summarize:

$$-\frac{1}{n} \log \beta_{1,n} \leq \min \left\{ I(U_{\{1\},1}; \tilde{Y}_{\{1\},1}); I(U_{\{1,2\},1}; \tilde{Y}_{\{1,2\},1}) \right\} + \vartheta_2(n), \quad (4.81)$$

$$-\frac{1}{n} \log \beta_{2,n} \leq \min \left\{ I(U_{\{1,2\},1}; \tilde{Y}_{\{1,2\},1}) + I(U_{\{1,2\},2}; \tilde{Y}_{\{1,2\},2}); \right. \\ \left. I(U_{\{2\},1}; \tilde{Y}_{\{2\},1}) + I(U_{\{2\},2}; \tilde{Y}_{\{2\},2}) \right\} + \vartheta_3(n), \quad (4.82)$$

where $\vartheta_2(n)$ and $\vartheta_3(n)$ are functions tending to 0 as $n \rightarrow \infty$.

Further, define the following random variables

$$\tilde{L}_{\mathcal{I},j} \triangleq \text{len}(\tilde{\mathbf{M}}_{\mathcal{I},j}), \quad j \in \{1, 2\}, \quad \mathcal{I} \in \mathcal{P}(2). \quad (4.83)$$

By the rate constraints (4.7d), and the definition of the random variables $\tilde{L}_{\mathcal{I},j}$, we obtain by the total law of expectations:

$$nR_1 \geq \mathbb{E}[L_1] \geq \sum_{\mathcal{I} \in \mathcal{P}(2)} \mathbb{E}[\tilde{L}_{\mathcal{I},1}] \Delta_{\mathcal{I}}. \quad (4.84)$$

Moreover,

$$H(\tilde{\mathbf{M}}_{\mathcal{I},1}) = H(\tilde{\mathbf{M}}_{\mathcal{I},1}, \tilde{L}_{\mathcal{I},1}) \quad (4.85)$$

$$= \sum_{l_{\mathcal{I}}} \Pr[\tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}] H(\tilde{\mathbf{M}}_{\mathcal{I},1} | \tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}) + H(\tilde{L}_{\mathcal{I},1}) \quad (4.86)$$

$$\leq \sum_{l_{\mathcal{I}}} \Pr[\tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}] l_{\mathcal{I}} + H(\tilde{L}_{\mathcal{I},1}) \quad (4.87)$$

$$= \mathbb{E}[\tilde{L}_{\mathcal{I},1}] + H(\tilde{L}_{\mathcal{I},1}), \quad (4.88)$$

which combined with (4.84) establishes

$$\sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} H(\tilde{\mathbf{M}}_{\mathcal{I},1}) \leq \sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} \mathbb{E}[\tilde{L}_{\mathcal{I},1}] + \Delta_{\mathcal{I}} H(\tilde{L}_{\mathcal{I},1}) \quad (4.89)$$

$$\leq nR_1 \left(1 + \sum_{\mathcal{I} \in \mathcal{P}(2)} h_b \left(\frac{\Delta_{\mathcal{I}}}{nR_1} \right) \right), \quad (4.90)$$

where (4.90) holds by (4.84) and because the entropy of the discrete and positive random variable $\tilde{L}_{\mathcal{I},1}$ of mean $\mathbb{E}[\tilde{L}_{\mathcal{I},1}] \leq \frac{nR_1}{\Delta_{\mathcal{I}}}$ is bounded by $\frac{nR_1}{\Delta_{\mathcal{I}}} \cdot h_b \left(\frac{\Delta_{\mathcal{I}}}{nR_1} \right)$, see [23, Theorem 12.1.1].

In a similar way, we obtain

$$\sum_{\mathcal{I} \in \{\{1,2\}, \{2\}\}} \Delta_{\mathcal{I}} H(\tilde{\mathbf{M}}_{\mathcal{I},2}) \leq nR_2 \left(1 + \sum_{\mathcal{I} \in \{\{1,2\}, \{2\}\}} h_b \left(\frac{\Delta_{\mathcal{I}}}{nR_2} \right) \right). \quad (4.91)$$

Then by combining (4.90) and (4.91) with (4.76) and (4.77), noting (4.66) and (4.74), and considering

also (4.81) and (4.82), we have proved so far that for all $n \geq 1$ there exist joint pmfs $P_{U_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}} = P_{U_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},0}}P_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}}$ (abbreviated as $P_{\mathcal{I},1}^{(n)}$) for $\mathcal{I} \in \mathcal{P}(2)$, and $P_{U_{\mathcal{I},2}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}} = P_{U_{\mathcal{I},2}\tilde{Y}_{\mathcal{I},1}}P_{\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}}$ (abbreviated as $P_{\mathcal{I},2}^{(n)}$) for $\mathcal{I} \in \{\{1,2\}, \{2\}\}$ so that the following conditions hold (where I_P indicates that the mutual information should be calculated according to a pmf P):

$$R_1 \geq \sum_{\mathcal{I} \in \mathcal{P}(2)} (I_{P_{\mathcal{I},1}^{(n)}}(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},0}) + g_{\mathcal{I},1}(n)) \cdot g_{\mathcal{I},2}(n, \eta), \quad (4.92a)$$

$$R_2 \geq \sum_{\mathcal{I} \in \{\{1,2\}, \{2\}\}} (I_{P_{\mathcal{I},2}^{(n)}}(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},1}) + g_{\mathcal{I},1}(n)) \cdot g_{\mathcal{I},2}(n, \eta), \quad (4.92b)$$

$$\theta_1 \leq \min \left\{ I_{P_{\{1\},1}^{(n)}}(U_{\{1\},1}; \tilde{Y}_{\{1\},1}), I_{P_{\{1,2\},2}^{(n)}}(U_{\{1,2\},1}; \tilde{Y}_{\{1,2\},1}) \right\} + g_{1,3}(n), \quad (4.92c)$$

$$\theta_2 \leq \min \left\{ I_{P_{\{1,2\},1}^{(n)}}(U_{\{1,2\},1}; \tilde{Y}_{\{1,2\},1}) + I_{P_{\{1,2\},2}^{(n)}}(U_{\{1,2\},2}; \tilde{Y}_{\{1,2\},2}), \right. \\ \left. I_{P_{\{2\},1}^{(n)}}(U_{\{1,2\},1}; \tilde{Y}_{\{2\},1}) + I_{P_{\{2\},2}^{(n)}}(U_{\{2\},2}; \tilde{Y}_{\{2\},2}) \right\} + g_{2,3}(n), \quad (4.92d)$$

$$g_{\mathcal{I},4}(n) = I_{P_{\mathcal{I},1}^{(n)}}(\tilde{Y}_{\mathcal{I},1}; U_{\mathcal{I},1} | \tilde{Y}_{\mathcal{I},0}), \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.92e)$$

for some nonnegative functions $g_{\mathcal{I},1}(n)$, $g_{\mathcal{I},2}(n, \eta)$, $g_{k,3}(n)$, $g_{\mathcal{I},4}(n)$ with the following asymptotic behaviors:

$$\lim_{n \rightarrow \infty} g_{\mathcal{I},1}(n) = 0, \quad \forall \mathcal{I} \in \mathcal{P}(2), \quad (4.93)$$

$$\lim_{n \rightarrow \infty} g_{k,3}(n) = 0, \quad \forall k \in \{1, 2\}, \quad (4.94)$$

$$\lim_{n \rightarrow \infty} g_{\mathcal{I},4}(n) = 0, \quad \forall \mathcal{I} \in \mathcal{P}(2), \quad (4.95)$$

$$\underline{\lim}_{n \rightarrow \infty} (g_{\{1\},2}(n, \eta) + g_{\{1,2\},2}(n, \eta)) \geq \frac{1 - \epsilon_1 - \eta}{1 - \eta}, \quad (4.96)$$

$$\underline{\lim}_{n \rightarrow \infty} (g_{\{1,2\},2}(n, \eta) + g_{\{2\},2}(n, \eta)) \geq \frac{1 - \epsilon_2 - \eta}{1 - \eta}, \quad (4.97)$$

$$\underline{\lim}_{\eta \downarrow 0} \underline{\lim}_{n \rightarrow \infty} g_{\{1,2\},2}(n, \eta) \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}. \quad (4.98)$$

We next observe that by Carathéodory's theorem [48, Appendix C] for each n there must exist random variables $U_{\{1\},1}, U_{\{1,2\},1}, U_{\{2\},1}, U_{\{1,2\},2}, U_{\{2\},2}$ satisfying (4.92) over alphabets of sizes

$$|\mathcal{U}_{\mathcal{I},1}| \leq |\mathcal{Y}_0| \cdot |\mathcal{Y}_1| + 2, \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.99)$$

$$|\mathcal{U}_{\mathcal{I},2}| \leq |\mathcal{Y}_1| + 1, \quad \mathcal{I} \in \{\{1,2\}, \{2\}\}. \quad (4.100)$$

Then we invoke the Bolzano-Weierstrass theorem and consider an increasing subsequence of positive numbers $\{n_k\}_{k=1}^{\infty}$ such that the following subsequences converge:

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}U_{\mathcal{I},1}}^{(n_k)} = P_{\tilde{Y}_{\mathcal{I},0}Y_{\mathcal{I},1}U_{\mathcal{I},1}}^*, \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.101)$$

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}U_{\mathcal{I},2}}^{(n_k)} = P_{\tilde{Y}_{\mathcal{I},1}Y_{\mathcal{I},2}U_{\mathcal{I},2}}^*, \quad \mathcal{I} \in \{\{1,2\}, \{2\}\}. \quad (4.102)$$

Considering further an appropriate sequence of diminishing η -values, we conclude by (4.92a)–(4.92d) and

(4.75) that:

$$R_1 \geq \sigma_{\{1\}} \cdot I_{P_{\{1,1\},1}^*}(U_{\{1\},1}; Y_{\{1\},0}) + \sigma_{\{1,2\}} \cdot I_{P_{\{1,2\},1}^*}(U_{\{1,2\},1}; Y_{\{1,2\},0}) + \sigma_{\{2\}} \cdot I_{P_{\{2,1\},1}^*}(U_{\{2\},1}; Y_{\{2\},0}), \quad (4.103)$$

$$R_2 \geq \sigma_{\{1,2\}} \cdot I_{P_{\{1,2\},2}^*}(U_{\{1,2\},2}; Y_{\{1,2\},1}) + \sigma_{\{2\}} \cdot I_{P_{\{2,2\},2}^*}(U_{\{2\},2}; Y_{\{2\},1}), \quad (4.104)$$

$$\theta_1 \leq \min \left\{ I_{P_{\{1,1\},1}^*}(U_{\{1\},1}; Y_{\{1\},1}), I_{P_{\{1,2\},1}^*}(U_{\{1,2\},1}; Y_{\{1,2\},1}) \right\}, \quad (4.105)$$

$$\theta_2 \leq \min \left\{ I_{P_{\{1,2\},1}^*}(U_{\{1,2\},1}; Y_{\{1,2\},1}) + I_{P_{\{1,2\},2}^*}(U_{\{1,2\},2}; Y_{\{1,2\},2}), \right. \\ \left. I_{P_{\{2,1\},1}^*}(U_{\{2\},1}; Y_{\{2\},1}) + I_{P_{\{2,2\},2}^*}(U_{\{2\},2}; Y_{\{2\},2}) \right\} \quad (4.106)$$

for some numbers $\sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}} > 0$ satisfying $\sigma_{\{1\}} + \sigma_{\{2\}} + \sigma_{\{1,2\}} \leq 1$ and

$$\sigma_{\{1\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_1, \quad (4.107a)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}, \quad (4.107b)$$

$$\sigma_{\{2\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_2. \quad (4.107c)$$

Notice further that since for any $\mathcal{I} \in \mathcal{P}(2)$ and any k the pair $(\tilde{Y}_{\mathcal{I},0}^{(n_k)}, \tilde{Y}_{\mathcal{I},1}^{(n_k)})$ lies in the jointly typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0 Y_1})$, we have $|P_{\tilde{Y}_{\mathcal{I},0} \tilde{Y}_{\mathcal{I},1}} - P_{Y_0 Y_1}| \leq \mu_{n_k}$ and thus the limiting pmfs satisfy $P_{\tilde{Y}_{\mathcal{I},0} \tilde{Y}_{\mathcal{I},1}}^* = P_{Y_0 Y_1}$. Moreover, since for each n_k the random variable $\tilde{Y}_{\mathcal{I},2}$ is drawn according to $P_{Y_2|Y_1}$ given $\tilde{Y}_{\mathcal{I},1}$, irrespective of $\tilde{Y}_{\mathcal{I},0}$, the limiting pmfs also satisfy $P_{\tilde{Y}_{\mathcal{I},2}|\tilde{Y}_{\mathcal{I},0} \tilde{Y}_{\mathcal{I},1}}^* = P_{Y_2|Y_1}$. We also notice for all $\mathcal{I} \in \{\{1,2\}, \{2\}\}$ that under $P_{Y_{\mathcal{I},1} Y_{\mathcal{I},2} U_{\mathcal{I},2}}^*$ the Markov chain

$$U_{\mathcal{I},2} \rightarrow Y_{\mathcal{I},1} \rightarrow Y_{\mathcal{I},2}, \quad (4.108)$$

holds because $U_{\mathcal{I},2} \rightarrow \tilde{Y}_{\mathcal{I},1} \rightarrow \tilde{Y}_{\mathcal{I},2}$ forms a Markov chain for any n_k . Finally, by continuity considerations and by (4.92e), the following Markov chain must hold under $P_{\tilde{Y}_{\mathcal{I},0} Y_{\mathcal{I},1} U_{\mathcal{I},1}}^*$ for all $\mathcal{I} \in \mathcal{P}(2)$:

$$U_{\mathcal{I},1} \rightarrow Y_{\mathcal{I},0} \rightarrow Y_{\mathcal{I},1}. \quad (4.109)$$

Using the definitions of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$, we thus proved that for any pair of achievable exponents (θ_1, θ_2) there exist rates $R_{\{1\},1}, R_{\{1,2\},1}, R_{\{2\},1}, R_{\{1,2\},2}, R_{\{2\},2} > 0$ satisfying

$$\theta_1 \leq \min \left\{ \eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1}) \right\}, \quad (4.110a)$$

$$\theta_2 \leq \min \left\{ \eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2}) \right\}, \quad (4.110b)$$

and numbers $\sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}} > 0$ satisfying $\sigma_{\{1\}} + \sigma_{\{2\}} + \sigma_{\{1,2\}} \leq 1$, Inequalities (4.107), and the following two rate constraints:

$$R_1 \geq \sigma_{\{1\}} \cdot R_{\{1\},1} + \sigma_{\{1,2\}} \cdot R_{\{1,2\},1} + \sigma_{\{2\}} \cdot R_{\{2\},1}, \quad (4.110c)$$

$$R_2 \geq \sigma_{\{1,2\}} \cdot R_{\{1,2\},2} + \sigma_{\{2\}} \cdot R_{\{2\},2}. \quad (4.110d)$$

■

4.3.5.3 Simplification of the Outer Bound in Proposition 1

We proceed to simplify the outer bound in Proposition 1 depending on the cases $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$. To this end, fix an exponent pair (θ_1, θ_2) in $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$, rates $R_{\{1\},1}, R_{\{1,2\},1}, R_{\{1,2\},2}, R_{\{2\},1}, R_{\{2\},2} \geq 0$ and numbers $\sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}} \geq 0$ summing to less than or equal to 1 and satisfying constraints (4.59).

4.3.5.3.1 The case $\epsilon_1 = \epsilon_2 = \epsilon$

By (4.59):

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\} \quad (4.111)$$

$$\stackrel{(a)}{\leq} \frac{\sigma_{\{1\}}\eta_1(R_{\{1\},1}) + \sigma_{\{1,2\}}\eta_1(R_{\{1,2\},1})}{\sigma_{\{1\}} + \sigma_{\{1,2\}}} \quad (4.112)$$

$$\stackrel{(b)}{\leq} \eta_1\left(\frac{\sigma_{\{1\}}R_{\{1\},1} + \sigma_{\{1,2\}}R_{\{1,2\},1}}{\sigma_{\{1\}} + \sigma_{\{1,2\}}}\right) \quad (4.113)$$

$$\stackrel{(c)}{\leq} \eta_1(R_1/(1-\epsilon)), \quad (4.114)$$

where (a) holds because the minimum is never larger than any linear combination; (b) holds by the concavity of the function $\eta_1(\cdot)$; and (c) holds by the monotonicity of the function $\eta_1(\cdot)$ and because by (4.59) we have $\sigma_{\{1\}}R_{\{1\},1} + \sigma_{\{1,2\}}R_{\{1,2\},1} \leq R_1$ and $\sigma_{\{1\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon$.

Following similar steps, one can prove that

$$\theta_2 \leq \min\left\{\eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2})\right\} \quad (4.115)$$

$$\stackrel{(d)}{\leq} \frac{\sigma_{\{2\}}\eta_1(R_{\{2\},1}) + \sigma_{\{2\}}\eta_2(R_{\{2\},2})}{\sigma_{\{2\}} + \sigma_{\{1,2\}}} + \frac{\sigma_{\{1,2\}}\eta_1(R_{\{1,2\},1}) + \sigma_{\{1,2\}}\eta_2(R_{\{1,2\},2})}{\sigma_{\{2\}} + \sigma_{\{1,2\}}} \quad (4.116)$$

$$\stackrel{(e)}{\leq} \eta_1\left(\frac{\sigma_{\{2\}}R_{\{2\},1} + \sigma_{\{1,2\}}R_{\{1,2\},1}}{\sigma_{\{2\}} + \sigma_{\{1,2\}}}\right) + \eta_2\left(\frac{\sigma_{\{2\}}R_{\{2\},2} + \sigma_{\{1,2\}}R_{\{1,2\},2}}{\sigma_{\{2\}} + \sigma_{\{1,2\}}}\right) \quad (4.117)$$

$$\stackrel{(f)}{\leq} \eta_1(R_1/(1-\epsilon)) + \eta_2(R_2/(1-\epsilon)), \quad (4.118)$$

where (d) holds again because the minimum is never larger than any linear combination; (e) holds by the concavity of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$; and (f) holds because by (4.59) we have $\sigma_{\{2\}}R_{\{2\},i} + \sigma_{\{1,2\}}R_{\{1,2\},i} \leq R_i$, for $i \in \{1, 2\}$, and $\sigma_{\{2\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon$.

This concludes the converse proof to (4.23).

4.3.5.3.2 The case $\epsilon_1 < \epsilon_2$

Choose nonnegative numbers $a_1, a_{1,2}, b_1, b_{1,2}, c_{1,2}$ satisfying

$$a_1 + a_{1,2} \leq \sigma_{\{1\}} \quad (4.119a)$$

$$b_1 + b_{1,2} \leq \sigma_{\{1,2\}} \quad (4.119b)$$

$$c_{1,2} \leq \sigma_{\{2\}} \quad (4.119c)$$

$$a_{1,2} + b_{1,2} = b_{1,2} + c_{1,2} = 1 - \epsilon_2 \quad (4.119d)$$

$$a_1 + b_1 = \epsilon_2 - \epsilon_1. \quad (4.119e)$$

Notice that this set of (in)equalities is equivalent to the two equalities $a_{1,2} = c_{1,2} = 1 - \epsilon_2 - b_{1,2}$ and $a_1 = \epsilon_2 - \epsilon_1 - b_1$ and the three inequalities:

$$1 - \epsilon_1 - b_{1,2} - b_1 \leq \sigma_{\{1\}} \quad (4.120a)$$

$$b_1 + b_{1,2} \leq \sigma_{\{1,2\}} \quad (4.120b)$$

$$1 - \epsilon_2 - b_{1,2} \leq \sigma_{\{2\}}. \quad (4.120c)$$

Through the Fourier-Motzkin Elimination (FME) Algorithm it can be verified that above three inequalities (4.120) have a nonnegative solution pair $(b_1, b_{1,2})$ with corresponding nonnegative values for $a_{1,2}, c_{1,2}, a_1$, whenever

$$0 \leq \sigma_{\mathcal{I}}, \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.121a)$$

$$1 - \epsilon_i \leq \sigma_{\{i\}} + \sigma_{\{1,2\}}, \quad i \in \{1, 2\}, \quad (4.121b)$$

$$0 \leq \epsilon_2 - \epsilon_1, \quad (4.121c)$$

which hold by assumption, see (4.59). The existence of the desired nonnegative numbers $a_1, a_{1,2}, b_1, b_{1,2}, c_{1,2}$ satisfying (4.119) is thus established.

With the chosen numbers, we form

$$\tilde{R}_{\{1,2\},1} := \max \left\{ \frac{a_{1,2}R_{\{1\},1} + b_{1,2}R_{\{1,2\},1}}{1 - \epsilon_2}, \frac{b_{1,2}R_{\{1,2\},1} + c_{1,2}R_{\{2\},1}}{1 - \epsilon_2} \right\}, \quad (4.122a)$$

$$\tilde{R}_{\{1,2\},2} := \frac{b_{1,2}R_{\{1,2\},2} + c_{1,2}R_{\{2\},2}}{1 - \epsilon_2}, \quad (4.122b)$$

$$\tilde{R}_{\{1\},1} := \frac{a_1R_{\{1\},1} + b_1R_{\{1,2\},1}}{\epsilon_2 - \epsilon_1}. \quad (4.122c)$$

We show that exponents (θ_1, θ_2) and rates $\tilde{R}_{\{1\},1}, \tilde{R}_{\{1,2\},1}$ and $\tilde{R}_{\{1,2\},2}$ satisfy constraints (4.24). To this

end, notice that

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\} \quad (4.123)$$

$$\stackrel{(a)}{\leq} \frac{a_1\eta_1(R_{\{1\},1}) + b_1\eta_1(R_{\{1,2\},1})}{\epsilon_2 - \epsilon_1} \quad (4.124)$$

$$\stackrel{(b)}{\leq} \eta_1\left(\frac{a_1R_{\{1\},1} + b_1R_{\{1,2\},1}}{\epsilon_2 - \epsilon_1}\right) \quad (4.125)$$

$$\stackrel{(c)}{\leq} \eta_1\left(\tilde{R}_{\{1\},1}\right), \quad (4.126)$$

where (a) holds because the minimum is smaller than any linear combination and because $a_1 + b_1 = \epsilon_2 - \epsilon_1$; (b) holds by the concavity of the function $\eta_1(\cdot)$; and (c) holds by the definition of rate $R_{\{1\},1}$. In a similar way we have:

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\} \quad (4.127)$$

$$\leq \frac{a_{1,2}\eta_1(R_{\{1\},1}) + b_{1,2}\eta_1(R_{\{1,2\},1})}{1 - \epsilon_2} \quad (4.128)$$

$$\leq \eta_1\left(\frac{a_{1,2}R_{\{1\},1} + b_{1,2}R_{\{1,2\},1}}{1 - \epsilon_2}\right) \quad (4.129)$$

$$\leq \eta_1\left(\tilde{R}_{\{1,2\},1}\right), \quad (4.130)$$

where the last step holds by the monotonicity of the function $\eta_1(\cdot)$ and because by definition $\tilde{R}_{\{1,2\},1} \geq \frac{a_{1,2}R_{\{1\},1} + b_{1,2}R_{\{1,2\},1}}{1 - \epsilon_2}$. Thus, by (4.126) and (4.130):

$$\theta_1 \leq \min\left\{\eta_1\left(\tilde{R}_{\{1\},1}\right), \eta_1\left(\tilde{R}_{\{1,2\},1}\right)\right\}. \quad (4.131)$$

We continue to notice

$$\theta_2 \leq \min\left\{\eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2})\right\} \quad (4.132)$$

$$\stackrel{(d)}{\leq} \frac{b_{1,2}\eta_1(R_{\{1,2\},1}) + b_{1,2}\eta_2(R_{\{1,2\},2})}{1 - \epsilon_2} + \frac{c_{1,2}\eta_1(R_{\{2\},1}) + c_{1,2}\eta_2(R_{\{2\},2})}{1 - \epsilon_2} \quad (4.133)$$

$$\stackrel{(e)}{\leq} \eta_1\left(\frac{b_{1,2}R_{\{1,2\},1} + c_{1,2}R_{\{2\},1}}{1 - \epsilon_2}\right) + \eta_2\left(\frac{b_{1,2}R_{\{1,2\},2} + c_{1,2}R_{\{2\},2}}{1 - \epsilon_2}\right) \quad (4.134)$$

$$\stackrel{(f)}{\leq} \eta_1\left(\tilde{R}_{\{1,2\},1}\right) + \eta_2\left(\tilde{R}_{\{1,2\},2}\right), \quad (4.135)$$

where (d) holds because the minimum is smaller than any linear combination and because $b_{1,2} + c_{1,2} = 1 - \epsilon_2$; (e) holds concavity of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$; and (f) holds by the definitions of rates $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ and by the monotonicity of the function $\eta_1(\cdot)$.

From the rate constraints in (4.59), we further obtain

$$R_1 \geq \sigma_{\{1\}}R_{\{1\},1} + \sigma_{\{2\}}R_{\{2\},1} + \sigma_{\{1,2\}}R_{\{1,2\},1} \quad (4.136)$$

$$\stackrel{(g)}{\geq} (a_1 + a_{1,2})R_{\{1\},1} + c_{1,2}R_{\{2\},1} + (b_1 + b_{1,2})R_{\{1,2\},1} \quad (4.137)$$

$$= (\epsilon_2 - \epsilon_1) \left(\frac{a_1 R_{\{1\},1} + b_1 R_{\{1,2\},1}}{\epsilon_2 - \epsilon_1} \right) + (1 - \epsilon_2) \left(\frac{a_{1,2} R_{\{1\},1} + c_{1,2} R_{\{2\},1} + b_{1,2} R_{\{1,2\},1}}{1 - \epsilon_2} \right) \quad (4.138)$$

$$\stackrel{(h)}{\geq} (\epsilon_2 - \epsilon_1) \tilde{R}_{\{1\},1} + (1 - \epsilon_2) \tilde{R}_{\{1,2\},1} \quad (4.139)$$

and

$$R_2 \geq \sigma_{\{1,2\}} R_{\{1,2\},2} + \sigma_{\{2\}} R_{\{2\},2} \quad (4.140)$$

$$\stackrel{(g)}{\geq} b_{1,2} R_{\{1,2\},2} + c_{1,2} R_{\{2\},2} \quad (4.141)$$

$$= (1 - \epsilon_2) \left(\frac{b_{1,2} R_{\{1,2\},2} + c_{1,2} R_{\{2\},2}}{1 - \epsilon_2} \right) \quad (4.142)$$

$$\stackrel{(h)}{=} (1 - \epsilon_2) \tilde{R}_{\{1,2\},2}, \quad (4.143)$$

where inequalities (g) hold because $a_1 + a_{1,2} \leq \sigma_{\{1\}}$, $c_{1,2} \leq \sigma_{\{2\}}$, and $b_1 + b_{1,2} \leq \sigma_{\{1,2\}}$, see (4.119); and (h) holds by the definitions of rates $\tilde{R}_{\{1\},1}$, $\tilde{R}_{\{1,2\},1}$, and $\tilde{R}_{\{1,2\},2}$ and because

$$a_{1,2} R_{\{1\},1} + c_{1,2} R_{\{2\},1} + b_{1,2} R_{\{1,2\},1} \geq \max\{a_{1,2} R_{\{1\},1} + b_{1,2} R_{\{1,2\},1}, b_{1,2} R_{\{1,2\},1} + c_{1,2} R_{\{2\},1}\}. \quad (4.144)$$

The desired converse result to (4.24) then follows by combining (4.131), (4.135), (4.139), and (4.143), and by noticing that by the monotonicity of the function $\eta_2(\cdot)$ there is no loss in optimality to restrict to rates $\tilde{R}_{\{1,2\},2} = R_2/(1 - \epsilon_2)$.

4.3.5.3.3 The case $\epsilon_1 > \epsilon_2$

The proof is similar to the case $\epsilon_1 < \epsilon_2$. We present it here for completeness.

Choose nonnegative numbers $a_{1,2}, b_2, b_{1,2}, c_2, c_{1,2}$ satisfying

$$a_{1,2} \leq \sigma_{\{1\}} \quad (4.145a)$$

$$b_2 + b_{1,2} \leq \sigma_{\{1,2\}} \quad (4.145b)$$

$$c_2 + c_{1,2} \leq \sigma_{\{2\}} \quad (4.145c)$$

$$a_{1,2} + b_{1,2} = b_{1,2} + c_{1,2} = 1 - \epsilon_1 \quad (4.145d)$$

$$b_2 + c_2 = \epsilon_1 - \epsilon_2, \quad (4.145e)$$

which is equivalent to the three equalities $a_{1,2} = c_{1,2} = 1 - \epsilon_1 - b_{1,2}$ and $c_2 = \epsilon_1 - \epsilon_2 - b_2$ and the three inequalities

$$1 - \epsilon_1 - b_{1,2} \leq \sigma_{\{1\}} \quad (4.146a)$$

$$b_2 + b_{1,2} \leq \sigma_{\{1,2\}} \quad (4.146b)$$

$$1 - \epsilon_2 - b_2 - b_{1,2} \leq \sigma_{\{2\}}. \quad (4.146c)$$

Through FME it can be shown that a nonnegative pair $(b_2, b_{1,2})$ satisfying (4.146) exists and the corresponding values for $a_{1,2}, c_{1,2}, c_2$ are nonnegative whenever

$$0 \leq \sigma_{\mathcal{I}}, \quad \mathcal{I} \in \mathcal{P}(2), \quad (4.147a)$$

$$1 - \epsilon_i \leq \sigma_{\{i\}} + \sigma_{\{1,2\}}, \quad i \in \{1, 2\}, \quad (4.147b)$$

$$0 \leq \epsilon_1 - \epsilon_2, \quad (4.147c)$$

which hold by assumption, see (4.59).

Define the new rates

$$\tilde{R}_{\{1,2\},1} := \max \left\{ \frac{a_{1,2}R_{\{1\},1} + b_{1,2}R_{\{1,2\},1}}{1 - \epsilon_1}, \frac{b_{1,2}R_{\{1,2\},1} + c_{1,2}R_{\{2\},1}}{1 - \epsilon_1} \right\}, \quad (4.148)$$

$$\tilde{R}_{\{1,2\},2} := \frac{b_{1,2}R_{\{1,2\},2} + c_{1,2}R_{\{2\},2}}{1 - \epsilon_1}, \quad (4.149)$$

$$\tilde{R}_{\{2\},i} := \frac{b_2R_{\{1,2\},i} + c_2R_{\{2\},i}}{\epsilon_1 - \epsilon_2}, \quad i \in \{1, 2\} \quad (4.150)$$

We show that the exponents θ_1, θ_2 and the rates $\tilde{R}_{\{2\},1}, \tilde{R}_{\{2\},2}, \tilde{R}_{\{1,2\},1}$ and $\tilde{R}_{\{1,2\},2}$ satisfy constraints (4.25).

To this end, notice that by similar arguments as in the preceding subsections:

$$\theta_1 \leq \min \{ \eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1}) \} \quad (4.151)$$

$$\leq \frac{a_{1,2}\eta_1(R_{\{1\},1}) + b_{1,2}\eta_1(R_{\{1,2\},1})}{1 - \epsilon_1} \quad (4.152)$$

$$\leq \eta_1 \left(\frac{a_{1,2}R_{\{1\},1} + b_{1,2}R_{\{1,2\},1}}{1 - \epsilon_1} \right) \quad (4.153)$$

$$\leq \eta_1 \left(\tilde{R}_{\{1,2\},1} \right). \quad (4.154)$$

Moreover,

$$\theta_2 \leq \min \{ \eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2}) \} \quad (4.155)$$

$$\leq \frac{b_{1,2}\eta_1(R_{\{1,2\},1}) + b_{1,2}\eta_2(R_{\{1,2\},2})}{1 - \epsilon_1} + \frac{c_{1,2}\eta_1(R_{\{2\},1}) + c_{1,2}\eta_2(R_{\{2\},2})}{1 - \epsilon_1} \quad (4.156)$$

$$\leq \eta_1 \left(\frac{b_{1,2}R_{\{1,2\},1} + c_{1,2}R_{\{2\},1}}{1 - \epsilon_1} \right) + \eta_2 \left(\frac{b_{1,2}R_{\{1,2\},2} + c_{1,2}R_{\{2\},2}}{1 - \epsilon_1} \right) \quad (4.157)$$

$$\leq \eta_1 \left(\tilde{R}_{\{1,2\},1} \right) + \eta_2 \left(\tilde{R}_{\{1,2\},2} \right) \quad (4.158)$$

and

$$\theta_2 \leq \frac{b_2\eta_1(R_{\{1,2\},1}) + b_2\eta_2(R_{\{1,2\},2})}{\epsilon_1 - \epsilon_2} + \frac{c_2\eta_1(R_{\{2\},1}) + c_2\eta_2(R_{\{2\},2})}{\epsilon_1 - \epsilon_2} \quad (4.159)$$

$$\leq \eta_1 \left(\frac{b_2R_{\{1,2\},1} + c_2R_{\{2\},1}}{\epsilon_1 - \epsilon_2} \right) + \eta_2 \left(\frac{b_2R_{\{1,2\},2} + c_2R_{\{2\},2}}{\epsilon_1 - \epsilon_2} \right) \quad (4.160)$$

$$\leq \eta_1 \left(\tilde{R}_{\{2\},1} \right) + \eta_2 \left(\tilde{R}_{\{2\},2} \right). \quad (4.161)$$

Combining (4.158) and (4.161) we obtain:

$$\theta_2 \leq \min \left\{ \eta_1 \left(\tilde{R}_{\{1,2\},1} \right) + \eta_1 \left(\tilde{R}_{\{1,2\},2} \right), \eta_1 \left(\tilde{R}_{\{2\},1} \right) + \eta_1 \left(\tilde{R}_{\{2\},2} \right) \right\}. \quad (4.162)$$

From the rate constraints in (4.59), inequalities (4.145), and the definitions of the rates $\tilde{R}_{\{2\},1}, \tilde{R}_{\{2\},2}, \tilde{R}_{\{1,2\},1}, \tilde{R}_{\{1,2\},2}$, we obtain:

$$R_1 \geq \sigma_{\{1\}} R_{\{1\},1} + \sigma_{\{1,2\}} R_{\{1,2\},1} + \sigma_{\{2\}} R_{\{2\},1} \quad (4.163)$$

$$\geq a_{1,2} R_{\{1\},1} + (c_2 + c_{1,2}) R_{\{2\},1} + (b_2 + b_{1,2}) R_{\{1,2\},1} \quad (4.164)$$

$$= (\epsilon_1 - \epsilon_2) \left(\frac{b_2 R_{\{1,2\},1} + c_2 R_{\{2\},1}}{\epsilon_1 - \epsilon_2} \right) + (1 - \epsilon_1) \left(\frac{a_{1,2} R_{\{1\},1} + c_{1,2} R_{\{2\},1} + b_{1,2} R_{\{1,2\},1}}{1 - \epsilon_1} \right) \quad (4.165)$$

$$\geq (\epsilon_1 - \epsilon_2) \tilde{R}_{\{2\},1} + (1 - \epsilon_1) \tilde{R}_{\{1,2\},1} \quad (4.166)$$

and

$$R_2 \geq \sigma_{\{1,2\}} R_{\{1,2\},2} + \sigma_{\{2\}} R_{\{2\},2} \quad (4.167)$$

$$\geq (b_2 + b_{1,2}) R_{\{1,2\},2} + (c_2 + c_{1,2}) R_{\{2\},2} \quad (4.168)$$

$$= (1 - \epsilon_1) \left(\frac{b_{1,2} R_{\{1,2\},2} + c_{1,2} R_{\{2\},2}}{1 - \epsilon_1} \right) + (\epsilon_1 - \epsilon_2) \left(\frac{b_2 R_{\{1,2\},2} + c_2 R_{\{2\},2}}{1 - \epsilon_1} \right) \quad (4.169)$$

$$= (1 - \epsilon_1) \tilde{R}_{\{1,2\},2} + (\epsilon_1 - \epsilon_2) \tilde{R}_{\{2\},2}. \quad (4.170)$$

Combining (4.154), (4.162), (4.166), and (4.170) establishes the desired converse result in (4.25).

4.4 A system with K -hops

We generalize our two-hop setup and results to K hops, i.e., to $K - 1$ relays. We first describe the setup in Section 4.4.1. Then, we review the optimal coding scheme and exponents region under maximum-rate constraints in [34] for vanishing type-I error probabilities and we give our strong converse result in Section 4.4.2. In Section 4.4.3, we present our results on the general optimal coding scheme and exponents region under expected-rate constraints followed by simplification results for the special case $K = 3$ in Section 4.4.4.

4.4.1 The Setup

Consider a system with a transmitter T_0 observing the source sequence Y_0^n , $K - 1$ relays labeled R_1, \dots, R_{K-1} and observing sequences Y_1^n, \dots, Y_{K-1}^n , respectively, and a receiver R_K observing sequence Y_K^n .

The source sequences $(Y_0^n, Y_1^n, \dots, Y_K^n)$ are distributed according to one of two distributions depending on a binary hypothesis $\mathcal{H} \in \{0, 1\}$:

$$\text{under } \mathcal{H} = 0 : (Y_0^n, Y_1^n, \dots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0 Y_1 \dots Y_K}; \quad (4.171a)$$

$$\text{if } \mathcal{H} = 1 : (Y_0^n, Y_1^n, \dots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1} \cdot \dots \cdot P_{Y_K} \quad (4.171b)$$

for a given pmf $P_{Y_0 Y_1 \dots Y_K}$ and where $P_{Y_0}, P_{Y_1}, \dots, P_{Y_K}$ denote the marginals of the joint pmf $P_{Y_0 Y_1 \dots Y_K}$.

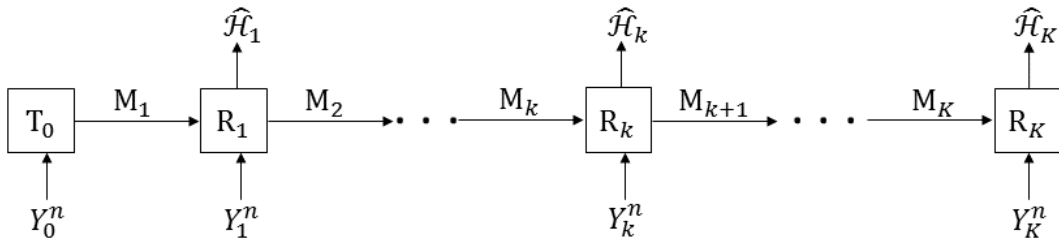


Figure 4.13: Cascaded K -hop setup with K DCs.

Communication takes place over K hops as illustrated in Figure 4.13. The transmitter T_0 sends a message $\mathbf{M}_1 = \phi_0^{(n)}(Y_0^n)$ to the first relay R_1 , which sends a message $\mathbf{M}_2 = \phi_1^{(n)}(Y_1^n, \mathbf{M}_1)$ to the second relay and so on. The communication is thus described by encoding functions

$$\phi_0^{(n)}: \mathcal{Y}_0^n \rightarrow \{0, 1\}^* \quad (4.172)$$

$$\phi_k^{(n)}: \mathcal{Y}_k^n \times \{0, 1\}^* \rightarrow \{0, 1\}^*, \quad k \in \{1, \dots, K - 1\}, \quad (4.173)$$

so that the produced message strings

$$\mathbf{M}_1 = \phi_0^{(n)}(Y_0^n) \quad (4.174)$$

$$\mathbf{M}_{k+1} = \phi_k^{(n)}(Y_k^n, \mathbf{M}_k), \quad k \in \{1, \dots, K - 1\}, \quad (4.175)$$

satisfy either the maximum-rate constraints

$$\text{len}(\mathbf{M}_k) \leq nR_k, \quad k \in \{1, \dots, K\}, \quad (4.176)$$

or the expected-rate constraints

$$\mathbb{E}[\text{len}(\mathbf{M}_k)] \leq nR_k, \quad k \in \{1, \dots, K\}. \quad (4.177)$$

Each relay R_1, \dots, R_{K-1} as well as the receiver R_K , produces a guess of the hypothesis \mathcal{H} . These guesses are described by guessing functions

$$g_k^{(n)}: \mathcal{Y}_k^n \times \{0, 1\}^* \rightarrow \{0, 1\}, \quad k \in \{1, \dots, K\}, \quad (4.178)$$

where we request that the guesses

$$\hat{\mathcal{H}}_{k,n} = g_k^{(n)}(Y_k^n, \mathbf{M}_k), \quad k \in \{1, \dots, K\}, \quad (4.179)$$

have type-I error probabilities

$$\alpha_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0], \quad k \in \{1, \dots, K\}, \quad (4.180)$$

not exceeding given thresholds $\epsilon_1, \epsilon_2, \dots, \epsilon_K > 0$, and type-II error probabilities

$$\beta_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1], \quad k \in \{1, \dots, K\}, \quad (4.181)$$

decaying to 0 exponentially fast with largest possible exponents.

Definition 9. Given maximum type-I error probabilities $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [0, 1]$ and rates $R_1, R_2, \dots, R_K \geq 0$. The exponent tuple $(\theta_1, \theta_2, \dots, \theta_K)$ is called $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$ -achievable if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, g_2^{(n)}, \dots, g_K^{(n)}\}_{n \geq 1}$ satisfying for each $k \in \{1, \dots, K\}$:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{k,n} \leq \epsilon_k, \quad (4.182a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_{k,n}} \geq \theta_k, \quad (4.182b)$$

and

$$\text{len}(\mathbf{M}_k) \leq nR_k, \quad (4.182c)$$

for the setup under maximum-rate constraints, or

$$\mathbb{E}[\text{len}(\mathbf{M}_k)] \leq nR_k, \quad (4.182d)$$

for the setup under expected-rate constraints.

Definition 10. The fundamental exponents region is defined as the closure of the set of all $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$ -achievable exponent pairs $(\theta_1, \theta_2, \dots, \theta_K)$ for given rates $R_1, R_2, \dots, R_K \geq 0$. It is denoted $\mathcal{E}_{K\text{-Hop}, \max}^*(R_1, R_2, \dots, R_K, \epsilon_1, \epsilon_2, \dots, \epsilon_K)$ under maximum-rate constraints, and $\mathcal{E}_{K\text{-Hop}}^*(R_1, R_2, \dots, R_K, \epsilon_1, \epsilon_2, \dots, \epsilon_K)$ under expected-rate constraints.

4.4.2 Maximum-Rate Constraints

In this section, we present the optimal coding scheme suggested by [34] for the K -hop setup. This coding scheme can achieve the optimal error exponent $\theta_{K\text{-Hop,max}}^*(R_1, R_2, \dots, R_K, \epsilon_1, \epsilon_2, \dots, \epsilon_K)$ for all $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [0, 1]$. Our strong converse result is presented in Section 4.4.2.3.

4.4.2.1 Optimal Coding Scheme under Maximum-Rate Constraints

In a similar manner to the two-hop network, the optimal coding scheme for K -hops uses Han's scheme in the following manner.

Han's scheme is first applied between T_0 and R_1 . If R_1 declares $\hat{\mathcal{H}}_1 = 1$, then it informs R_2 of this event by sending the single-bit message $\mathbf{M}_2 = [0]$, and R_2 declares $\hat{\mathcal{H}}_2 = 1$. However if Han's scheme between T_0 and R_1 results in R_1 declaring $\hat{\mathcal{H}}_1 = 0$, then R_1 and R_2 run a separate Han's scheme, and R_2 produces the guess $\hat{\mathcal{H}}_2$ indicated by this second Han's scheme.

The same strategy applies to all next relays R_k ($k \in \{2, \dots, K-1\}$). If the decision taken by R_k is $\hat{\mathcal{H}}_k = 1$, the single-bit message $\mathbf{M}_k = [0]$ is forwarded to R_{k+1} informing it to declare $\hat{\mathcal{H}}_{k+1} = 1$, otherwise a separate Han's scheme is run between R_k and R_{k+1} . Note that for $k = K-1$, the next node is the receiver R_K .

For completeness, we describe the details of the coding scheme as follows. Fix a blocklength n and choose the following parameters:

- a small positive number $\mu > 0$; and
- conditional pmfs $P_{U_k|Y_{k-1}}$ for all $k \in \{1, \dots, K\}$ leading to $P_{Y_{k-1}Y_kU_k} := P_{Y_{k-1}Y_k}P_{U_k|Y_{k-1}}$,

where all mutual information quantities will be evaluated according to the joint pmfs $P_{Y_{k-1}Y_kU_k}$ as defined above.

Randomly generate the codebooks for all $k \in \{1, \dots, K\}$

$$\mathcal{C}_{U_k} \triangleq \left\{ u_k^n(m_k) : m_k \in \left\{ 1, \dots, 2^{n(I(U_k; Y_{k-1}) + \mu)} \right\} \right\} \quad (4.183)$$

by drawing all entries i.i.d. according to the marginal pmfs P_{U_k} .

T_0 : Assume it observes $Y_0^n = y_0^n$. Then, it looks for indices m_1 satisfying

$$(u_1^n(m_1), y_0^n) \in \mathcal{T}_\mu^{(n)}(P_{U_1 Y_0}),$$

randomly picks one of these indices, and sends its corresponding bit-string

$$\mathbf{M}_1 = [\text{bin}(m_1)]. \quad (4.184)$$

If no such index m_1 exists, then T_0 sends

$$\mathbf{M}_1 = [0]. \quad (4.185)$$

R_k ($\forall k \in \{1, \dots, K-1\}$): Assume it observes $Y_k^n = y_k^n$ and receives the bit-string message $\mathbf{M}_k = \mathbf{m}_k$.

If $\mathbf{m}_k = [0]$, then it produces

$$\mathbf{M}_{k+1} = [0], \quad \text{and} \quad \hat{\mathcal{H}}_k = 1. \quad (4.186)$$

Else it sets $m_k = \text{dec}(\mathbf{m}_k)$ and checks if

$$(u_k^n(m_k), y_k^n) \in \mathcal{T}_\mu^{(n)}(P_{U_k Y_k}). \quad (4.187)$$

If the check fails, it follows the degenerate scheme in (4.186). However, if the check is successful, R_k declares $\hat{\mathcal{H}}_k = 0$, looks for indices m_{k+1} satisfying $(u_{k+1}^n(m_{k+1}), y_k^n) \in \mathcal{T}_\mu^{(n)}(P_{U_{k+1} Y_k})$, randomly picks one of them, and sends

$$\mathbf{M}_{k+1} = [\text{bin}(m_{k+1})] \quad (4.188)$$

to R_{k+1} .

If no such index m_{k+1} exists, it directly sends the single-bit message

$$\mathbf{M}_{k+1} = [0]. \quad (4.189)$$

R_K : Assume it observes the sequence $Y_K^n = y_K^n$ and receives message $\mathbf{M}_K = \mathbf{m}_K$.

If $\mathbf{m}_K = [0]$, it declares $\hat{\mathcal{H}}_K = 1$.

Else it sets $m_K = \text{dec}(\mathbf{m}_K)$, and checks if

$$(u_K^n(m_K), y_K^n) \in \mathcal{T}_\mu^{(n)}(P_{U_K Y_K}).$$

It declares $\hat{\mathcal{H}}_K = 0$ if the check succeeds, and $\hat{\mathcal{H}}_K = 1$ otherwise.

4.4.2.2 Optimal Exponents Region

The fundamental exponents region of this setup was only established for vanishing type-I error probabilities, i.e., when all type-I error thresholds $\epsilon_1, \dots, \epsilon_K \downarrow 0$ in [34]. This result is presented in this subsection and we strengthen it by our strong converse result in Section 4.4.2.3.

Definition 11. For any $\ell \in \{1, \dots, K\}$, define the function

$$\eta_\ell: \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+ \quad (4.190)$$

$$R \mapsto \max_{\substack{P_{U|Y_{\ell-1}}: \\ R \geq I(U; Y_{\ell-1})}} I(U; Y_\ell). \quad (4.191)$$

The functions η_1, \dots, η_K are concave and monotonically non-decreasing. The proof is analogous to the proof of Lemma 1 presented in Appendix A.1, and omitted for brevity. Notice further that in the maximization determining $\eta_\ell(R)$ it suffices to consider distributions $P_{U|Y_{\ell-1}}$ on alphabets of sizes $|\mathcal{Y}_{\ell-1}| + 1$, see [1]. In the following, we abbreviate the expression $\lim_{\epsilon_1 \downarrow 0, \epsilon_2 \downarrow 0, \dots, \epsilon_K \downarrow 0} \mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ by $\mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, 0, \dots, 0)$.

Theorem 10 (Proposition 5 in [34]). *The fundamental exponents region under the maximum-rate constraints (4.182c) and vanishing type-I error constraints satisfies*

$$\mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, 0, \dots, 0) = \left\{ (\theta_1, \dots, \theta_K) : \theta_k \leq \sum_{\ell=1}^k \eta_\ell(R_\ell), k \in \{1, \dots, K\} \right\} \quad (4.192)$$

Notice that in this K -hop setup, each decision center accumulates all the error exponents on the various links from the transmitter to the decision center. The fundamental exponents region is thus given by a K -dimensional hyper-rectangle. That means, each decision center can simultaneously achieve its optimal error exponent independently from the other DCs in the system.

4.4.2.3 Strong Converse Result

Theorem 11. *The fundamental exponents region under the maximum-rate constraints (4.182c) is independent of the tuple $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$ for all $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [0, 1]$:*

$$\mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K) = \mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, 0, \dots, 0) \quad (4.193)$$

Proof: Achievability follows by setting $\epsilon_1 = \epsilon_2 = \dots = \epsilon_K = 0$. The strong converse proof is in Appendix B.6. ■

Due to the above theorem, we abbreviate $\mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ by $\mathcal{E}_{\text{K-Hop, max}}^*(R_1, \dots, R_K)$.

4.4.3 Expected-Rate Constraints

4.4.3.1 Optimal Coding Scheme under Expected-Rate Constraints

Similarly to the two-hop scheme, the terminals multiplex different subschemes depending on the sequence Y_0^n observed at the transmitter T_0 . To this end, partition the set \mathcal{Y}_0^n into disjoint subsets \mathcal{D}_\emptyset and $\{\mathcal{D}_I\}_{I \in \mathcal{P}(K)}$

so that the probabilities

$$\sigma_{\mathcal{I}} := \Pr[Y_0^n \in \mathcal{D}_{\mathcal{I}}] \quad (4.194)$$

satisfy

$$1 - \sum_{k \in \mathcal{S}} \epsilon_k \leq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \mathcal{S} \subseteq \mathcal{I}}} \sigma_{\mathcal{I}}, \quad \mathcal{S} \subseteq \{1, \dots, K\}, \quad (4.195a)$$

$$\sum_{\mathcal{I} \in \mathcal{P}(K)} \sigma_{\mathcal{I}} \leq 1. \quad (4.195b)$$

The main structure of the coding scheme is illustrated in Figure 4.14.

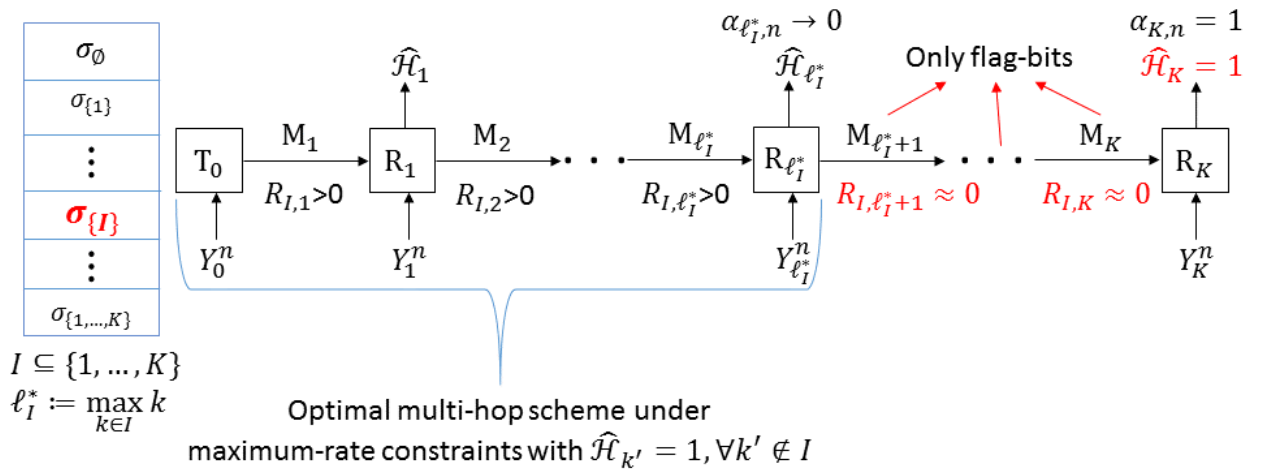


Figure 4.14: Illustration of the general optimal coding scheme for the K -hop setup with K DCs under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.

In our multiplexed schemes, the index \mathcal{I} of $\mathcal{D}_{\mathcal{I}}$ indicates that if T_0 's observation Y_0^n lies in $\mathcal{D}_{\mathcal{I}}$, then all terminals R_k , for $k \in \mathcal{I}$, attempt to correctly guess hypothesis \mathcal{H} , while all terminals R_k , for $k \notin \mathcal{I}$, simply declare $\hat{\mathcal{H}}_k = 1$. If $Y_0^n \in \mathcal{D}_{\emptyset}$, then *all* terminals R_1, \dots, R_K simply declare $\hat{\mathcal{H}} = 1$.

The transmitter T_0 adds K flag-bits to its message M_1 to inform R_1 about the set $\mathcal{D}_{\mathcal{I}}$ containing its observation Y_0^n , and thus about the choice of the employed coding scheme. These flag-bits are forwarded by all relays R_1, \dots, R_{K-1} at the beginning of their messages M_2, \dots, M_K so as to pass the information to all terminals in the network.

We describe the different multiplexed coding schemes in more detail. Let $\ell_{\mathcal{I}}^*$ be the largest index in set \mathcal{I} :

$$\ell_{\mathcal{I}}^* := \max_{k \in \mathcal{I}} k, \quad (4.196)$$

and choose a set of rates

$$\{R_{\mathcal{I},\ell} : \mathcal{I} \in \mathcal{P}(K), \ell \in \{1, \dots, \ell_{\mathcal{I}}^*\}\} \quad (4.197)$$

satisfying

$$R_\ell \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \ell_{\mathcal{I}}^* \geq \ell}} \sigma_{\mathcal{I}} \cdot R_{\mathcal{I},\ell}, \quad \ell \in \{1, \dots, K\}. \quad (4.198)$$

We will see that the choice of the various rates determines the tradeoff between the different exponents $\theta_1, \dots, \theta_K$. Rates $\{R_{\mathcal{I},\ell}: \ell \in \{1, \dots, \ell_{\mathcal{I}}^*\}\}$ are used in the subscheme employed when $Y_0^n \in \mathcal{D}_{\mathcal{I}}$, where under this event only the messages on the first $\ell_{\mathcal{I}}^*$ links have positive rates, while messages on the last $K - \ell_{\mathcal{I}}^*$ links are of zero rate. The reason is that terminals $R_{\ell_{\mathcal{I}}^*+1}, \dots, R_K$ simply declare $\hat{\mathcal{H}} = 1$ and thus messages $M_{\ell_{\mathcal{I}}^*+1}, \dots, M_K$ only have to convey the zero-rate information that $Y_0^n \in \mathcal{D}_{\mathcal{I}}$.

Subscheme for $Y_0^n \in \mathcal{D}_\emptyset$: All terminals T_0 and R_1, \dots, R_{K-1} send the length- K all-zero bit string over the respective communication links:

$$M_1 = \dots = M_K = [0, 0, \dots, 0]. \quad (4.199)$$

Upon receiving this all-zero flag, relays R_1, \dots, R_{K-1} and receiver R_K all declare

$$\hat{\mathcal{H}}_1 = \dots = \hat{\mathcal{H}}_K = 1. \quad (4.200)$$

Communication is thus only used to inform the relays and the receiver about the scheme to employ, or equivalently the event $Y_0^n \in \mathcal{D}_\emptyset$, without providing any further information about the correct hypothesis.

Subscheme for $Y_0^n \in \mathcal{D}_{\mathcal{I}}$, for $\mathcal{I} \in \mathcal{P}(K)$: In this case, only terminals R_k , for $k \in \mathcal{I}$, attempt to correctly guess hypothesis \mathcal{H} ; all other terminals R_k , for $k \notin \mathcal{I}$, directly declare $\hat{\mathcal{H}}_k = 1$.

Terminals $T_0, R_1, \dots, R_{\ell_{\mathcal{I}}^*}$ apply a given $\ell_{\mathcal{I}}^*$ -hop hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\mathcal{I},1}, \dots, R_{\mathcal{I},\ell_{\mathcal{I}}^*}$ on the first $\ell_{\mathcal{I}}^*$ links. To inform all relays and the receiver about the scheme to use, terminals T_0, R_1, \dots, R_{K-1} append a K -length flag sequence describing set \mathcal{I} at the beginning of their messages. We propose that this flag sequence shows bit 1 at all positions $k \in \mathcal{I}$ and bit 0 at all positions $k \notin \mathcal{I}$. Notice that Messages $M_{\ell_{\mathcal{I}}^*+1}, \dots, M_K$ consist of only the flag sequence.

All terminals R_k with $k \in \mathcal{I}$ declare the hypothesis indicated by the employed multi-hop hypothesis testing scheme. The remaining terminals R_k with $k \notin \mathcal{I}$ simply declare

$$\hat{\mathcal{H}}_k = 1, \quad k \notin \mathcal{I}. \quad (4.201)$$

Analysis: By (4.194) and (4.198), and because transmission of K bits hardly changes the rate for sufficiently large blocklengths, the proposed overall scheme respects the expected-rate constraints R_1, \dots, R_K on

the K links for large values of n . Appendix B.7 proves that when the optimal multi-hop hypothesis testing schemes with vanishing type-I error probability [34] are used as the various subschemes, then the overall scheme satisfies the type-I error constraints $\epsilon_1, \dots, \epsilon_K$ and achieves the error exponents in the following Theorem 12.

4.4.3.2 Optimal General Exponents Region

Theorem 12. *The fundamental exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ is equal to the set of all non-negative tuples $(\theta_1, \dots, \theta_K)$ satisfying*

$$\theta_k \leq \min_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \sum_{\ell=1}^k \eta_\ell(R_{\mathcal{I}, \ell}), \quad (4.202a)$$

for some nonnegative rates $\{R_{\mathcal{I}, 1}, \dots, R_{\mathcal{I}, \ell_{\mathcal{I}}^*}\}_{\mathcal{I} \in \mathcal{P}(K)}$ and nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(K)}$ satisfying

$$R_k \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \leq \ell_{\mathcal{I}}^*}} \sigma_{\mathcal{I}} \cdot R_{\mathcal{I}, k}, \quad k \in \{1, \dots, K\}, \quad (4.202b)$$

$$\max \left\{ 0, 1 - \sum_{k \in \mathcal{S}} \epsilon_k \right\} \leq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \mathcal{S} \subseteq \mathcal{I}}} \sigma_{\mathcal{I}}, \quad \mathcal{S} \subseteq \{1, \dots, K\}, \quad (4.202c)$$

$$\sum_{\mathcal{I} \in \mathcal{P}(K)} \sigma_{\mathcal{I}} \leq 1. \quad (4.202d)$$

Proof: Achievability is based on the coding scheme presented in the previous subsection and analyzed in Appendix B.7. The converse is proved in Section 4.4.3.4. ■

4.4.3.3 Discussion of the Benefits of Expected-Rate Constraints

Similar observations apply to the general Theorem 12 as for $K = 2$. In particular, irrespective of the ordering of the permissible type-I error probabilities, the largest exponent achievable at a decision center k is given by

$$\theta_{k, \max} := \sum_{\ell=1}^k \eta_\ell \left(\frac{R_\ell}{1 - \epsilon_k} \right), \quad (4.203)$$

which intuitively can be obtained when we have to satisfy only the type-I error constraint at the decision center k and utilize the available resources to maximize its exponent as if it is the only DC in the system. It coincides with the optimal exponent under maximum-rate constraint and vanishing type-I error probabilities, see Theorem 10, but where the rates are boosted by the factor $(1 - \epsilon_k)^{-1}$. In fact, $\theta_k = \theta_{k, \max}$ is

achieved by choosing the first k rates as:¹

$$R_{\mathcal{I},\ell} = \frac{R_\ell}{1 - \epsilon_k}, \quad \mathcal{I} \ni k, \ell \in \{1, \dots, k\}. \quad (4.204)$$

This choice imposes that $\sigma_{\mathcal{I}} R_{\mathcal{I},\ell} = 0$ for all \mathcal{I} not containing k and all $\ell \in \{1, \dots, k\}$. As a consequence, the optimal performance for a decision center $R_{k'}$, for $k' < k$, is

$$\theta_{k'} = \sum_{\ell=1}^{k'} \eta_\ell \left(\frac{R_\ell}{1 - \epsilon_k} \right), \quad \text{if } \epsilon_{k'} > \epsilon_k \quad (4.205)$$

$$\theta_{k'} = 0, \quad \text{if } \epsilon_{k'} < \epsilon_k, \quad (4.206)$$

where the performance in (4.205) is obtained by setting $\sigma_{\mathcal{I}} = 0$ for all \mathcal{I} containing an index $k' < k$ with $\epsilon'_{k'} > \epsilon_k$ and by setting the corresponding rates to infinity. Notice that $\sigma_{\mathcal{I}}$ cannot be chosen equal to 0 for all sets \mathcal{I} containing index $k' < k$ when $\epsilon_{k'} < \epsilon_k$ because Constraint (4.202c) implies that at least one of these σ -values is positive, which by $\sigma_{\mathcal{I}} R_{\mathcal{I},\ell} = 0$ implies that the corresponding rates $R_{\mathcal{I},\ell} = 0$, for all $\ell = 1, \dots, k$, causing $\theta_{k'}$ to degrade to 0. We conclude that under (4.203), for any $k' < k$, when $\epsilon'_{k'} \geq \epsilon_k$ then exponent $\theta_{k'}$ is degraded from its maximum value because all rates are only boosted by the factor $(1 - \epsilon_k)^{-1}$ and not by the larger factor $(1 - \epsilon_{k'})^{-1}$, and when $\epsilon'_{k'} < \epsilon_k$ the exponent $\theta_{k'}$ completely degrades to 0.

With appropriate choices for the rates on the last $(K - k)$ links, different tradeoffs between the exponents $\theta_{k+1}, \dots, \theta_K$ can be achieved. In particular, it is possible that an exponent $\theta_{k'}$, for $k' > k$, experiences its maximum rate-boost $(1 - \epsilon_{k'})^{-1}$ on some of these links. On the first k links, any exponent $\theta_{k+1}, \dots, \theta_K$ experiences a rate-boost of $(1 - \epsilon_k)^{-1}$ if the corresponding $\epsilon_{k'} > \epsilon_k$, whereas the contributions of the first k links completely degrade to 0 if $\epsilon_{k'} < \epsilon_k$.

Further notice the following property of the region in Theorem 12.

Lemma 5. Consider a set of nonnegative numbers $\{R_{\mathcal{I},1}, \dots, R_{\mathcal{I},\ell^*}\}_{\mathcal{I} \in \mathcal{P}(K)}$ and $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(K)}$ satisfying (4.202) for exponents $(\theta_1, \dots, \theta_K)$. Let $\mathcal{I}', \mathcal{I}'' \in \mathcal{P}(K)$ and $\Gamma \in [0, \sigma_{\mathcal{I}''}]$ be so that

$$\mathcal{I}' \subseteq \mathcal{I}'' \quad (4.207)$$

and

$$\max \left\{ 0, 1 - \sum_{k \in \mathcal{S}} \epsilon_k \right\} + \Gamma \leq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \mathcal{S} \subseteq \mathcal{I}}} \sigma_{\mathcal{I}}, \quad \mathcal{S} \subseteq \mathcal{I}'', \mathcal{S} \not\subseteq \mathcal{I}'. \quad (4.208)$$

Then, the new nonnegative numbers

$$\tilde{\sigma}_{\mathcal{I}'} = \sigma_{\mathcal{I}'} + \Gamma \quad (4.209)$$

¹This choice assumes that the ordering (4.216) is strict, i.e., no two ϵ -values coincide. Moreover, when some of the available rates R_1, \dots, R_k are sufficiently large so as to saturate the functions $\eta_\ell(R_\ell)$, then other choices are possible.

$$\tilde{\sigma}_{\mathcal{I}''} = \sigma_{\mathcal{I}''} - \Gamma \quad (4.210)$$

$$\tilde{\sigma}_{\mathcal{I}} = \sigma_{\mathcal{I}}, \quad \mathcal{I} \in \mathcal{P}(K) \setminus \{\mathcal{I}', \mathcal{I}''\}, \quad (4.211)$$

and rates,

$$\tilde{R}_{\mathcal{I}', \ell} = \frac{\sigma_{\mathcal{I}'} \cdot R_{\mathcal{I}', \ell} + \Gamma \cdot R_{\mathcal{I}'', \ell}}{\tilde{\sigma}_{\mathcal{I}'}} , \quad \ell \in \{1, \dots, \ell_{\mathcal{I}'}^*\}, \quad (4.212)$$

$$\tilde{R}_{\mathcal{I}, \ell} = R_{\mathcal{I}, \ell}, \quad \mathcal{I} \in \mathcal{P}(K) \setminus \{\mathcal{I}'\}, \ell \in \{1, \dots, \ell_{\mathcal{I}}^*\}. \quad (4.213)$$

also satisfy (4.202) for exponents $(\theta_1, \dots, \theta_K)$.

Proof: Above rate-definitions essentially only shift the term $\Gamma \cdot R_{\mathcal{I}'', \ell}$ from $\sigma_{\mathcal{I}''} R_{\mathcal{I}'', \ell}$ to $\tilde{\sigma}_{\mathcal{I}'} \tilde{R}_{\mathcal{I}', \ell}$, and therefore the rate-constraints (4.202b) remain valid also for the new numbers. Similarly, constraint (4.202d) remains valid since the sum of all σ -values is preserved. Notice further that the σ -values included in Constraint (4.202c) for $\mathcal{S} \not\subseteq \mathcal{I}'$ remain unchanged by (4.211) and for $\mathcal{S} \subseteq \mathcal{I}'$ their sum is preserved by (4.209) and (4.210). For $\mathcal{S} \not\subseteq \mathcal{I}'$ but $\mathcal{S} \subseteq \mathcal{I}''$, Constraint (4.202c) is satisfied by Assumption (4.208). It remains to check the validity of (4.202a) for the new rate-values. By (4.213) the constraint remains unchanged for all $k \notin \mathcal{I}'$. For $k \in \mathcal{I}'$, we notice that by (4.207) the minimum in (4.202a) includes both sets \mathcal{I}' and \mathcal{I}'' and this minimum cannot be smaller for the new rates because:

$$\min \left\{ \sum_{\ell=1}^k \eta_{\ell}(R_{\mathcal{I}', \ell}), \sum_{\ell=1}^k \eta_{\ell}(R_{\mathcal{I}'', \ell}) \right\} \leq \min \left\{ \sum_{\ell=1}^k \left(\frac{\sigma_{\mathcal{I}'}}{\tilde{\sigma}_{\mathcal{I}'}} \eta_{\ell}(R_{\mathcal{I}', \ell}) + \frac{\Gamma}{\tilde{\sigma}_{\mathcal{I}'}} \eta_{\ell}(R_{\mathcal{I}'', \ell}) \right), \sum_{\ell=1}^k \eta_{\ell}(R_{\mathcal{I}'', \ell}) \right\} \quad (4.214)$$

$$\leq \min \left\{ \sum_{\ell=1}^k \eta_{\ell}(\tilde{R}_{\mathcal{I}', \ell}), \sum_{\ell=1}^k \eta_{\ell}(\tilde{R}_{\mathcal{I}'', \ell}) \right\}, \quad (4.215)$$

where the first inequality holds because the minimum of two numbers cannot exceed any convex combination of the numbers, and the second inequality holds by the concavity and monotonicity of the functions $\{\eta_{\ell}(\cdot)\}_{\ell}$. \blacksquare

Above lemma indicates that when evaluating the fundamental exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ in Theorem 12 one can restrict to sets of parameters $\{\sigma_{\mathcal{I}}\}$ that satisfy some of the constraints (4.202c) with equality and set certain σ -values to 0. In fact, we conjecture that the simplified expression for the exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ in Conjecture 13 ahead holds, where we define a permutation $\pi: \{1, \dots, K\} \rightarrow \{1, \dots, K\}$ that orders the ϵ -values in decreasing order:

$$\epsilon_{\pi(1)} \geq \epsilon_{\pi(2)} \geq \dots \geq \epsilon_{\pi(K)}, \quad (4.216)$$

and sets $\epsilon_{\pi(0)} := 1$. We observe that the expression in Conjecture 13 is obtained from Theorem 12 by setting

$$\sigma_{\{\pi(i), \dots, \pi(K)\}} = \epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}, \quad i \in \{1, \dots, K\}, \quad (4.217)$$

and all other σ -values to 0, and by renaming rates $R_{\{\pi(i), \dots, \pi(K)\}, \ell}$ to $R_{i, \ell}$ and $\ell_{\{\pi(i), \dots, \pi(K)\}}^*$ to ℓ_i^* . The region in Conjecture 13 is thus an achievable exponents region, but we conjecture it is optimal.

Conjecture 13. *The fundamental exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ is the set of all exponent tuples $(\theta_1, \dots, \theta_K)$ that satisfy*

$$\theta_k \leq \min_{i \in \{1, \dots, \pi(k)\}} \left[\sum_{\ell=1}^k \eta_\ell (R_{i, \ell}) \right], \quad k \in \{1, \dots, K\}, \quad (4.218a)$$

for some nonnegative rates $\{R_{i, \ell}\}$ satisfying

$$R_\ell \geq \sum_{\substack{i \in \{1, \dots, K\}: \\ \ell_i^* \geq \ell}} (\epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}) R_{i, \ell}, \quad \ell \in \{1, \dots, K\}, \quad (4.218b)$$

where

$$\ell_i^* := \max_{\ell} \{\ell : \ell \in \{\pi(i), \dots, \pi(K)\}\}. \quad (4.218c)$$

Linking this conjecture to the coding scheme in the previous Subsection 4.4.3.1, we observe that if it holds, then the optimal coding scheme only multiplexes $K + 1$ coding schemes (instead of 2^K schemes as implied by Theorem 12), where the i -th scheme is applied with probability $\epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}$ and is intended only for the DCs with $(K - i + 1)$ -th smallest type-I error constraints. The structure of the simplified coding scheme is presented in Figure 4.15.

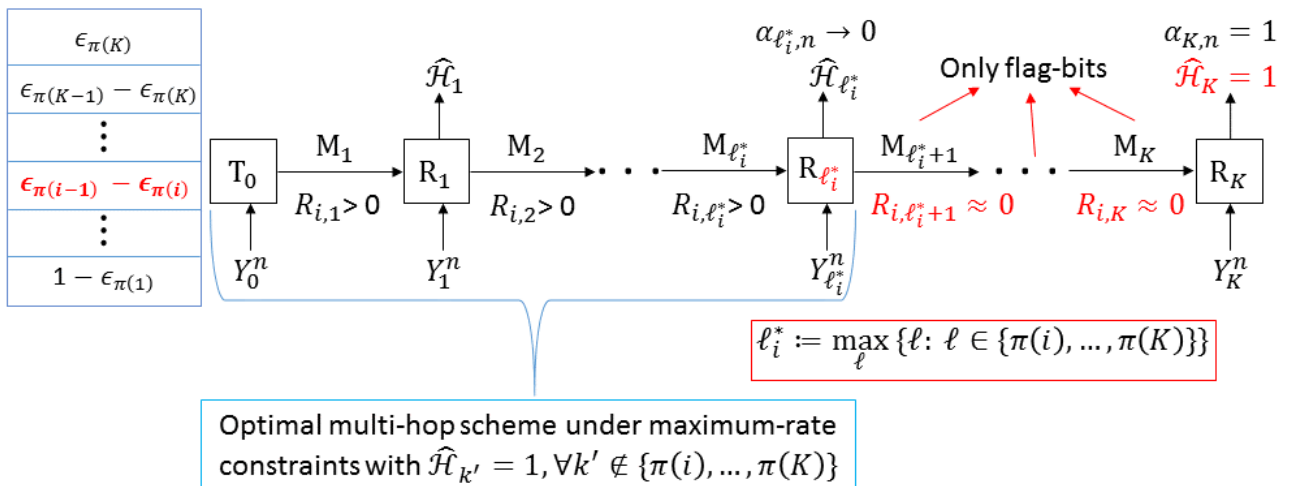


Figure 4.15: Illustration of the conjectured simplified optimal coding scheme for the K -hop setup with K DCs under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.

4.4.3.4 Converse Proof to Theorem 12

Here we propose a new proof method that has some differences compared to the proof method that we presented in Section 4.3.5. The main difference is that this new proof method does not require the blowing-up lemma as we now partition the set of strongly jointly typical sequences among *all* observations $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1 \dots Y_K})$ into 2^K subsets by *considering all possible combinations of the decisions taken at the different DCs*. This indicates that the probabilities are now either 1 or 0 and thus there is no need for the blowing-up lemma, see Figure 4.16 for an illustration when $K = 2$. Another difference in this method is that now by conditioning on all sequences we lose *all* the i.i.d.ness properties and thus we need to prove that all the Markov chains hold asymptotically in the limit as $n \rightarrow \infty$. We next present the technical details of the proof.

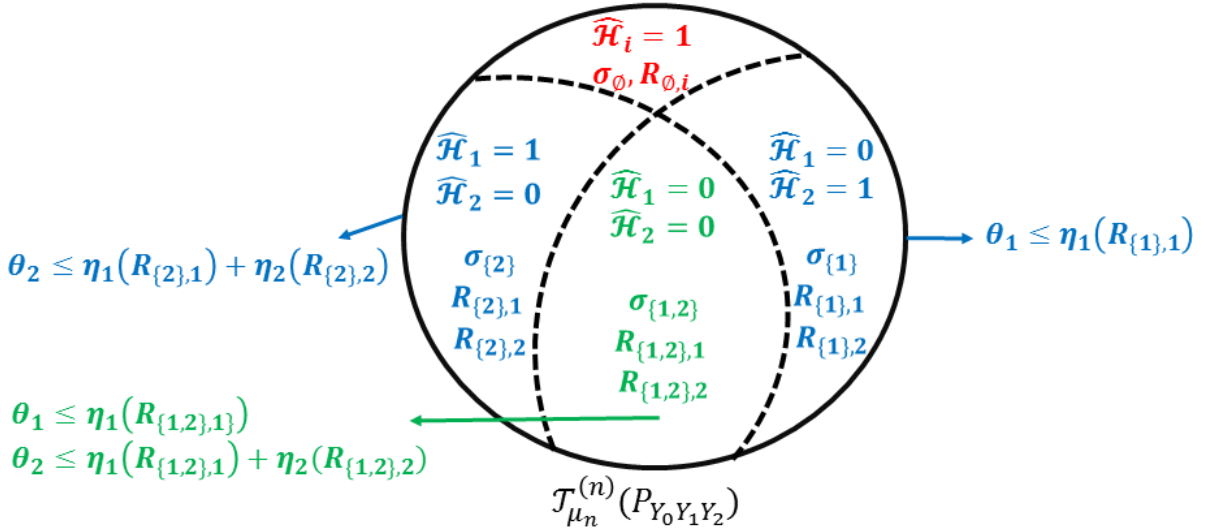


Figure 4.16: Sketch of partitioning $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1 Y_2})$ and applying parallel change of measure arguments in the converse proof to Theorem 12 when $K = 2$.

Fix an exponent-tuple $(\theta_1, \dots, \theta_K)$ in the exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$, and a sequence (in n) of encoding and decision functions $\{(\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_K^{(n)}, g_1^{(n)}, \dots, g_K^{(n)})\}_{n \geq 1}$ achieving this tuple, i.e., satisfying constraints (4.182).

Our proof relies on the following lemma (which generalizes Lemma 4 to $K \geq 2$).

Lemma 6. *Given a small number $\eta > 0$. Fix a blocklength n and a set $\mathcal{D} \subseteq \mathcal{Y}_0^n \times \mathcal{Y}_1^n \times \dots \times \mathcal{Y}_K^n$ of probability exceeding δ , and let the tuple $(\tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_K, \tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_K^n)$ follow the pmf*

$$P_{\tilde{M}_1 \tilde{M}_2 \dots \tilde{M}_K \tilde{Y}_0^n \tilde{Y}_1^n \dots \tilde{Y}_K^n}(\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_K, y_0^n, y_1^n, \dots, y_K^n) \triangleq P_{Y_0^n Y_1^n \dots Y_K^n}(y_0^n, y_1^n, \dots, y_K^n) \cdot \frac{\mathbb{1}\{(y_0^n, y_1^n, \dots, y_K^n) \in \mathcal{D}\}}{P_{Y_0^n Y_1^n \dots Y_K^n}(\mathcal{D})} \cdot \mathbb{1}\{\phi_1(y_0^n) = \mathbf{m}_1\}$$

$$\cdot \mathbb{1}\{\phi_2(y_1^n, \phi_1(y_0^n)) = \mathbf{m}_2\} \cdot \cdots \cdot \mathbb{1}\{\phi_K(y_{K-1}^n, \phi_{K-1}(y_{K-2}^n, \phi_{K-2}(\cdots, \phi_1(y_0^n)))) = \mathbf{m}_K\}. \quad (4.219)$$

Further, define the auxiliary random variables

$$U_k \triangleq (\tilde{\mathbf{M}}_k, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, \dots, \tilde{Y}_K^{T-1}, T), \quad k \in \{1, \dots, K\}, \quad (4.220)$$

$$\tilde{Y}_k \triangleq \tilde{Y}_{k,T}, \quad k \in \{0, 1, \dots, K\}, \quad (4.221)$$

where T is uniform over $\{1, \dots, n\}$ and independent of the tuple $(\tilde{\mathbf{M}}_1, \tilde{\mathbf{M}}_2, \dots, \tilde{\mathbf{M}}_K, \tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_K^n)$.

For any $k \in \{1, \dots, K\}$, the following (in)equalities hold:

$$H(\tilde{\mathbf{M}}_k) \geq nI(U_k; \tilde{Y}_{k-1}) + \log P_{Y_0^n Y_1^n \dots Y_K^n}(\mathcal{D}), \quad (4.222)$$

$$I(U_k; \tilde{Y}_k | \tilde{Y}_{k-1}) = \phi_{1,k}(n), \quad (4.223)$$

where $\phi_{1,k}(n)$ is a function that tends to 0 as $n \rightarrow \infty$.

If

$$\Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, \dots, Y_K^n = y_K^n] \geq \eta, \quad \forall (y_0^n, \dots, y_K^n) \in \mathcal{D}, \quad (4.224)$$

then

$$-\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1, (Y_0^n, \dots, Y_K^n) \in \mathcal{D}] \leq \sum_{\ell=1}^k I(U_\ell; \tilde{Y}_\ell) + \phi_{2,k}(n), \quad (4.225)$$

where $\phi_{2,k}(n)$ are functions that tend to 0 as $n \rightarrow \infty$.

Proof: See Appendix B.8. ■

We continue to prove Theorem 12. Fix a small positive number $\eta > 0$ and set $\mu_n = n^{-1/3}$. Define for each index $k \in \{1, \dots, K\}$ the set

$$\mathcal{B}_k(\eta) \triangleq \{(y_0^n, \dots, y_K^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \dots Y_K}) : \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, \dots, Y_K^n = y_K^n] \geq \eta\}, \quad (4.226)$$

and for each subset $\mathcal{I} \in \mathcal{P}(K)$ the set

$$\begin{aligned} \mathcal{D}_{\mathcal{I}}(\eta) \triangleq & \left\{ (y_0^n, \dots, y_K^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \dots Y_K}) : \right. \\ & \forall k \in \mathcal{I}, \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, \dots, Y_K^n = y_K^n] \geq \eta, \\ & \left. \text{and } \forall k \notin \mathcal{I}, \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 0, Y_0^n = y_0^n, \dots, Y_K^n = y_K^n] < \eta \right\}. \end{aligned} \quad (4.227)$$

Notice that the sets $\{\mathcal{D}_{\mathcal{I}}(\eta)\}_{\mathcal{I}}$ are disjoint and

$$\bigcup_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \mathcal{D}_{\mathcal{I}}(\eta) = \mathcal{B}_k(\eta). \quad (4.228)$$

We continue to notice in a similar manner to steps (4.69)–(4.74) that by [44, Remark to Lemma 2.12]

and the type-I error probability constraints in (4.182a), for any $k \in \{1, \dots, K\}$:

$$P_{Y_0^n Y_1^n \dots Y_{K-1}^n}(\mathcal{B}_k(\eta)) \geq \frac{1 - \epsilon_k - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0| \cdots |\mathcal{Y}_K|}{(1 - \eta)(4\mu_n^2 n)}. \quad (4.229)$$

Defining

$$\Delta_{\mathcal{I}} := P_{Y_0^n Y_1^n \dots Y_K^n}(\mathcal{D}_{\mathcal{I}}(\eta)), \quad (4.230)$$

we conclude by (4.228), by standard laws of probability, and the disjointness of the sets $\{\mathcal{D}_{\mathcal{I}}(\eta)\}_{\mathcal{I}}$, that in the limit as $n \rightarrow \infty$ and $\eta \downarrow 0$, for any subset $\mathcal{S} \subseteq \{1, \dots, K\}$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \mathcal{S} \subseteq \mathcal{I}}} \Delta_{\mathcal{I}} \geq \max \left\{ 1 - \sum_{k \in \mathcal{S}} \epsilon_k, 0 \right\}. \quad (4.231)$$

We now apply Lemma 6 to every subset $\mathcal{D}_{\mathcal{I}}$, for $\mathcal{I} \in \mathcal{P}(K)$ with $\Delta_{\mathcal{I}} \geq \eta$. This allows to conclude that for any such \mathcal{I} there exist random variables $\{U_{\mathcal{I},1}, \dots, U_{\mathcal{I},\ell_{\mathcal{I}}^*}\}$ so that the random variables $(\tilde{M}_{\mathcal{I},1}, \tilde{M}_{\mathcal{I},2}, \dots, \tilde{M}_{\mathcal{I},\ell_{\mathcal{I}}^*}, \tilde{Y}_{\mathcal{I},0}^n, \tilde{Y}_{\mathcal{I},1}^n, \dots, \tilde{Y}_{\mathcal{I},K}^n)$ defined in the lemma satisfy for any $k \in \{1, \dots, K\}$ the (in)equalities

$$H(\tilde{M}_{\mathcal{I},k}) \geq nI(U_{\mathcal{I},k}; \tilde{Y}_{k-1}^n) + \log \Delta_{\mathcal{I}}, \quad (4.232a)$$

$$I(U_{\mathcal{I},k}; \tilde{Y}_{\mathcal{I},k}^n | \tilde{Y}_{\mathcal{I},k-1}^n) = \phi_{1,\mathcal{I},k}(n), \quad (4.232b)$$

and for indices $k \in \mathcal{I}$ moreover:

$$-\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1, (Y_0^n, \dots, Y_K^n) \in \mathcal{D}_{\mathcal{I}}] \leq \sum_{\ell=1}^k I(U_{\mathcal{I},\ell}; \tilde{Y}_{\mathcal{I},\ell}^n) + \phi_{2,\mathcal{I},k}(n), \quad (4.232c)$$

where for each (\mathcal{I}, k) the functions $\phi_{1,\mathcal{I},k}(n)$ and $\phi_{2,\mathcal{I},k}(n) \rightarrow 0$ as $n \rightarrow \infty$.

In the sequel we assume that η is very small and $\Delta_{\mathcal{I}} \geq \eta$ for any \mathcal{I} . Otherwise the proof is similar but omitted for brevity.

We continue with the total law of probability to obtain:

$$-\frac{1}{n} \log \beta_{k,n} \leq \min_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \sum_{\ell=1}^k I(U_{\mathcal{I},\ell}; \tilde{Y}_{\mathcal{I},\ell}^n) + \phi_3(n), \quad (4.233)$$

where $\phi_{3,k}(n)$ is a function that tends to 0 as $n \rightarrow \infty$. We further define the following random variables for $\mathcal{I} \in \mathcal{P}(K)$ and $k \in \{1, \dots, \ell_{\mathcal{I}}^*\}$:

$$\tilde{L}_{\mathcal{I},k} \triangleq \text{len}(\tilde{M}_{\mathcal{I},k}). \quad (4.234)$$

By the rate constraints (4.177) and the total law of expectations:

$$nR_k \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \ell_{\mathcal{I}}^* \geq k}} \mathbb{E}[\tilde{L}_{\mathcal{I},k}] \Delta_{\mathcal{I}}, \quad (4.235)$$

and, similarly to (4.91), we obtain

$$\sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \ell_{\mathcal{I}}^* \geq k}} \Delta_{\mathcal{I}} H(\tilde{M}_{\mathcal{I},k}) \leq nR_k \left(1 + \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \ell_{\mathcal{I}}^* \geq k}} h_b \left(\frac{\Delta_{\mathcal{I}}}{nR_k} \right) \right). \quad (4.236)$$

Then combining (4.236) with (4.232a) and (4.233), and taking $n \rightarrow \infty$ and $\eta \downarrow 0$, we obtain that

$$R_k \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \ell_{\mathcal{I}}^* \geq k}} \Delta_{\mathcal{I}}^* \cdot I(U_{\mathcal{I},k}^*; \tilde{Y}_{k-1}^*), \quad (4.237a)$$

$$\theta_k \leq \min_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \sum_{\ell=1}^k I(U_{\mathcal{I},\ell}^*; \tilde{Y}_{\mathcal{I},\ell}^*) \quad (4.237b)$$

for some random variables $(\tilde{Y}_0^*, \dots, \tilde{Y}_K^*) \sim P_{Y_0 Y_1 \dots Y_K}$ and $\{U_{\mathcal{I},1}^*, \dots, U_{\mathcal{I},\ell_{\mathcal{I}}^*}^*\}_{\mathcal{I}}$ that by (4.232b) satisfy the Markov chains

$$U_{\mathcal{I},k}^* \rightarrow Y_{k-1}^* \rightarrow Y_k^* \quad (4.238)$$

and nonnegative numbers $\{\Delta_{\mathcal{I}}^*\}_{\mathcal{I}}$ that by (4.231) satisfy for any subset $\mathcal{S} \subseteq \{1, \dots, K\}$:

$$\sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ \mathcal{S} \subseteq \mathcal{I}}} \Delta_{\mathcal{I}}^* \geq \max \left\{ 1 - \sum_{k \in \mathcal{S}} \epsilon_k, 0 \right\}. \quad (4.239)$$

4.4.4 The Special Case $K = 3$

4.4.4.1 Simplified Optimal Coding Scheme and Exponents Region under Expected-Rate Constraints

In this simplified coding scheme, the terminals multiplex *four* different subschemes depending on the sequence y_0^n observed at the transmitter T_0 instead of *eight* subschemes as in the general coding scheme in Section 4.4.3.1. To this end, partition the set \mathcal{Y}_0^n into four disjoint subsets $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ so that the subsets $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ are of largest possible probabilities but not exceeding

$$\Pr [Y_0^n \in \mathcal{D}_i] \leq \epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}, \quad i \in \{1, 2, 3\}. \quad (4.240a)$$

Since $\mathcal{D}_0, \dots, \mathcal{D}_3$ form a partition:

$$\Pr [Y_0^n \in \mathcal{D}_0] \geq \epsilon_{\pi(3)}, \quad (4.240b)$$

where all inequalities (4.240) turn into equalities as $n \rightarrow \infty$.

We describe the coding scheme in detail, and in particular the employed subschemes for the various sets

$\mathcal{D}_0, \dots, \mathcal{D}_3$. To this end, choose a set of rates

$$\{R_{i,\ell}: i \in \{1, 2, 3\}, \ell \in \{1, \dots, \ell_i^*\}\} \quad (4.241)$$

satisfying

$$R_\ell > \sum_{\substack{i \in \{1, 2, 3\}: \\ \ell_i^* \geq \ell}} (\epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}) R_{i,\ell}, \quad \ell \in \{1, 2, 3\}, \quad (4.242)$$

where recall that ℓ_i^* is the largest index in the set $\{\pi(i), \dots, \pi(3)\}$.

Subscheme for $Y_0^n \in \mathcal{D}_0$: All terminals T_0 , R_1 , and R_2 send the two-bit zeros-string over the respective communication links:

$$\mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}_3 = [0, 0]. \quad (4.243)$$

Upon receiving this all-zero flag, terminals R_1 , R_2 , and R_3 all declare

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = \hat{\mathcal{H}}_3 = 1. \quad (4.244)$$

Subscheme for $Y_0^n \in \mathcal{D}_i$, for $i \in \{1, 2, 3\}$: In this case, only terminals $R_{\pi(i)}, \dots, R_{\pi(3)}$ attempt to correctly guess hypothesis \mathcal{H} ; all other terminals directly declare $\hat{\mathcal{H}} = 1$.

Terminals T_0 , R_1 , \dots , $R_{\ell_i^*}$ apply the optimal ℓ_i^* -hop hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{i,1}, \dots, R_{i,\ell_i^*}$ on the first ℓ_i^* links. To inform all relays and the receiver about the scheme to use, terminals T_0 , R_1, R_2 append the two-length binary description of the index i to their messages. In particular, Messages $\mathbf{M}_{\ell_i^*+1}, \dots, \mathbf{M}_3$ contain only these flag-bits.

Terminals with indices $\{\pi(i), \dots, \pi(3)\}$ declare the hypothesis indicated by the employed multi-hop hypothesis testing scheme. All other DCs simply declare

$$\hat{\mathcal{H}}_{\pi(1)} = \dots = \hat{\mathcal{H}}_{\pi(i-1)} = 1. \quad (4.245)$$

The overall scheme, by (4.240), (4.242) and because transmission of two bits hardly changes the rate for sufficiently large blocklengths, satisfies the expected-rate constraints, the type-I error constraints $\epsilon_1, \epsilon_2, \epsilon_3$, and achieves the error exponents in the following proposition.

Proposition 2. *Conjecture 13 holds for $K = 3$.*

Proof: As already mentioned in Section 4.4.3.3, achievability of the region in (4.218) for any value of K follows by specializing the region in Theorem 12 to the parameter choice in (4.217) and setting all other σ -values to 0, and by renaming rates $R_{\{\pi(i), \dots, \pi(K)\}, \ell}$ as $R_{i,\ell}$. The converse for $K = 3$ is proved in

Section 4.4.4.2. ■

4.4.4.2 Converse Proof to Proposition 2

We start with two auxiliary lemmas.

Lemma 7. *Let $K = 3$. In Theorem 12 it suffices to consider values $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$ so that*

$$\sigma_{\{1,2,3\}} + \sigma_{\{\pi(1),\pi(2)\}} + \sigma_{\{\pi(1),\pi(3)\}} + \sigma_{\{\pi(1)\}} = 1 - \epsilon_{\pi(1)} \quad (4.246)$$

$$\sigma_{\{\pi(1),\pi(2)\}} + \sigma_{\{\pi(2)\}} \geq \sigma_{\{\pi(1),\pi(3)\}} \quad (4.247)$$

Proof: See Appendix B.9. ■

We thus continue with nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$, and $\{R_{\mathcal{I},1}, \dots, R_{\mathcal{I},\ell_{\mathcal{I}}^*}\}_{\mathcal{I} \in \mathcal{P}(3)}$ satisfying (4.202) for $K = 3$ as well as (4.247). The proof of the desired proposition follows by the next lemma (which holds for any positive integer K) and by an appropriate choice of parameters $\{c_{\mathcal{J}}\}$, see (4.268) ahead.

Lemma 8. *Let*

$$\{c_{\mathcal{J}}: \mathcal{J} \in \mathcal{P}(K)\}, \quad (4.248)$$

$$\{\delta_{\mathcal{I},\mathcal{J}}: \mathcal{I}, \mathcal{J} \in \mathcal{P}(K) \text{ and } \mathcal{I} \cap \mathcal{J} \neq \emptyset\} \quad (4.249)$$

be sets of nonnegative integers satisfying

$$\sum_{\substack{\mathcal{J} \in \mathcal{P}(K): \\ \mathcal{I} \cap \mathcal{J} \neq \emptyset}} \delta_{\mathcal{I},\mathcal{J}} \leq \sigma_{\mathcal{I}}, \quad \mathcal{I} \in \mathcal{P}(K), \quad (4.250)$$

and

$$\sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}} \geq c_{\mathcal{J}}, \quad \forall k \in \mathcal{J}, \mathcal{J} \in \mathcal{P}(K). \quad (4.251)$$

Then, the rates

$$\tilde{R}_{\mathcal{J},k} := \max_{\substack{j \in \mathcal{J}: \\ j \geq k}} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ j \in \mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{c_{\mathcal{J}}} R_{\mathcal{I},k}, \quad k \leq \ell_{\mathcal{J}}^*, \mathcal{J} \in \mathcal{P}(K), \quad (4.252)$$

satisfy the following inequalities:

$$\theta_k \leq \min_{\substack{\mathcal{J} \in \mathcal{P}(K): \\ k \in \mathcal{J}}} \sum_{\ell=1}^k \eta_{\ell} \left(\tilde{R}_{\mathcal{J},\ell} \right), \quad k \in \{1, \dots, K\}, \quad (4.253)$$

and

$$R_k \geq \sum_{\substack{\mathcal{J} \in \mathcal{P}(K): \\ k \leq \ell_{\mathcal{J}}^*}} c_{\mathcal{J}} \cdot \tilde{R}_{\mathcal{J},k}, \quad k \in \{1, \dots, K\}. \quad (4.254)$$

Proof: We start by proving (4.253). By (4.202a), for any $k \in \{1, \dots, K\}$ and any set $\mathcal{J} \subseteq \mathcal{P}(K)$ containing index k :

$$\theta_k \leq \min_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \sum_{\ell=1}^k \eta_\ell(R_{\mathcal{I},\ell}) \quad (4.255)$$

$$\stackrel{(a)}{\leq} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{\sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}}} \cdot \sum_{\ell=1}^k \eta_\ell(R_{\mathcal{I},\ell}) \quad (4.256)$$

$$\stackrel{(b)}{\leq} \sum_{\ell=1}^k \eta_\ell \left(\sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{\sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}}} \cdot R_{\mathcal{I},\ell} \right) \quad (4.257)$$

$$\stackrel{(c)}{\leq} \sum_{\ell=1}^k \eta_\ell \left(\sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \in \mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{c_{\mathcal{J}}} \cdot R_{\mathcal{I},\ell} \right) \quad (4.258)$$

$$\stackrel{(d)}{\leq} \sum_{\ell=1}^k \eta_\ell(\tilde{R}_{\mathcal{J},\ell}), \quad (4.259)$$

where (a) holds because the minimum of a set of numbers is never larger than any convex combination of these numbers; (b) holds by the concavity of the functions $\eta_1(\cdot), \dots, \eta_k(\cdot)$; (c) holds by assumption (4.251) and by the monotonicity of the functions $\eta_1(\cdot), \dots, \eta_k(\cdot)$; and (d) holds by the definition of $\tilde{R}_{\mathcal{J},k}$ in (4.252) because $k \geq \ell$ and $k \in \mathcal{J}$ thus $\ell \leq \ell_{\mathcal{J}}^*$.

To prove (4.254), fix $k \in \{1, \dots, K\}$ and for each subset $\mathcal{J} \subseteq \mathcal{P}(K)$ with $\ell_{\mathcal{J}}^* \geq k$ pick an index $j_{\mathcal{J}} \in \mathcal{J}$ so that $j_{\mathcal{J}} \geq k$. Then, by (4.202b):

$$R_k \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{I}}^*}} \sigma_{\mathcal{I}} \cdot R_{\mathcal{I},k}, \quad (4.260)$$

$$\stackrel{(e)}{\geq} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{I}}^*}} \sum_{\substack{\mathcal{J} \in \mathcal{P}(K) \\ \mathcal{I} \cap \mathcal{J} \neq \emptyset}} \delta_{\mathcal{I},\mathcal{J}} \cdot R_{\mathcal{I},k} \quad (4.261)$$

$$= \sum_{\mathcal{J} \in \mathcal{P}(K)} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{I}}^* \\ \mathcal{I} \cap \mathcal{J} \neq \emptyset}} \delta_{\mathcal{I},\mathcal{J}} \cdot R_{\mathcal{I},k} \quad (4.262)$$

$$\stackrel{(f)}{\geq} \sum_{\substack{\mathcal{J} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{J}}^*}} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{I}}^* \\ \mathcal{I} \cap \mathcal{J} \neq \emptyset}} \delta_{\mathcal{I},\mathcal{J}} \cdot R_{\mathcal{I},k} \quad (4.263)$$

$$\stackrel{(f)}{\geq} \sum_{\substack{\mathcal{J} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{J}}^*}} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K) \\ k \leq \ell_{\mathcal{I}}^* \\ j_{\mathcal{J}} \in \mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}} R_{\mathcal{I},k} \quad (4.264)$$

$$\stackrel{(g)}{=} \sum_{\substack{\mathcal{J} \in \mathcal{P}(K): \\ k \leq \ell_{\mathcal{J}}^*}} \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ j_{\mathcal{J}} \in \mathcal{I}}} \delta_{\mathcal{I}, \mathcal{J}} R_{\mathcal{I}, k} \quad (4.265)$$

$$= \sum_{\substack{\mathcal{J} \in \mathcal{P}(K): \\ k \leq \ell_{\mathcal{J}}^*}} c_{\mathcal{J}} \cdot \frac{\sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ j_{\mathcal{J}} \in \mathcal{I}}} \delta_{\mathcal{I}, \mathcal{J}} R_{\mathcal{I}, k}}{c_{\mathcal{J}}}, \quad (4.266)$$

where (e) holds by Assumption (4.250); inequalities (f) hold because we consider less summands and each summand is nonnegative (recall that $j_{\mathcal{J}} \in \mathcal{J}$); and finally (g) holds because the two conditions $j_{\mathcal{J}} \geq k$ and $j_{\mathcal{J}} \in \mathcal{I}$ imply that $\ell_{\mathcal{J}}^* \geq k$.

The proof of the lemma is concluded by recalling the definition of rate $\tilde{R}_{\mathcal{J}, k}$ in (4.252) and noting that Inequality (4.259) holds for any set \mathcal{J} containing k whereas Inequality (4.266) holds for any index $j_{\mathcal{J}} \in \mathcal{J}$ larger than k . \blacksquare

To obtain the desired simplification in Proposition 2 from Theorem 12, define the subsets

$$\mathcal{J}_k := \{\pi(k), \dots, \pi(K)\}, \quad k \in \{1, \dots, K\}, \quad (4.267)$$

and the values $\pi(0) := 0$ and $\epsilon_0 := 1$. Applying above Lemma 8 to the choice

$$c_{\mathcal{J}} := \begin{cases} \epsilon_{\pi(k-1)} - \epsilon_{\pi(k)}, & \mathcal{J} = \mathcal{J}_k, \\ 0, & \text{otherwise,} \end{cases} \quad (4.268)$$

establishes the converse to Conjecture 13 for general values of K , if one renames rates $\tilde{R}_{\mathcal{J}_k, \ell}$ as $\tilde{R}_{k, \ell}$. The proof is concluded by showing that above parameter choice is permissible, i.e., that there exist nonnegative numbers $\{\delta_{\mathcal{I}, \mathcal{J}}\}$ satisfying conditions (4.250) and (4.251) for $\{c_{\mathcal{J}}\}$ in (4.268). For general values of K this seems cumbersome.

For $K = 3$, this can be achieved by means of the Fourier-Motzkin Elimination algorithm [52], which shows the existence of nonnegative numbers $\{\delta_{\mathcal{I}, \mathcal{J}}\}$ satisfying conditions (4.250) and (4.251) for $\{c_{\mathcal{J}}\}$ in (4.268), whenever (redundant conditions are omitted)

$$\sigma_{\{1,2,3\}} + \sigma_{\{\pi(1), \pi(2)\}} + \sigma_{\{\pi(1), \pi(3)\}} + \sigma_{\{\pi(1)\}} \geq 1 - \epsilon_{\pi(1)} \quad (4.269a)$$

$$\sigma_{\{1,2,3\}} + \sigma_{\{\pi(1), \pi(2)\}} + \sigma_{\{\pi(2), \pi(3)\}} + \sigma_{\{\pi(2)\}} \geq 1 - \epsilon_{\pi(2)} \quad (4.269b)$$

$$\sigma_{\{1,2,3\}} + \sigma_{\{\pi(1), \pi(3)\}} + \sigma_{\{\pi(2), \pi(3)\}} + \sigma_{\{\pi(3)\}} \geq 1 - \epsilon_{\pi(3)} \quad (4.269c)$$

and

$$2\sigma_{\{1,2,3\}} + 2\sigma_{\{\pi(1), \pi(2)\}} + \sigma_{\{\pi(1), \pi(3)\}} + \sigma_{\{\pi(2), \pi(3)\}} + \sigma_{\{\pi(1)\}} + \sigma_{\{\pi(2)\}} + \sigma_{\{\pi(3)\}} \geq 1 - \epsilon_{\pi(1)} + 1 - \epsilon_{\pi(3)}. \quad (4.269d)$$

Since Conditions (4.269a)–(4.269c) are satisfied by Assumption (4.202c) and Condition (4.269d) is implied by (4.269a), (4.269b), and (4.247), this concludes the proof for $K = 3$ and thus establishes Proposition 2.

4.5 Summary and Discussion

Introducing multiple decision centers to the multi-hop network in this chapter led to exploring another dimension of the distributed hypothesis testing problem: the tradeoff between the maximum simultaneously achievable error exponents at different centers. While no tradeoff appears neither in the setup under maximum-rate constraints nor under expected-rate constraints with identical type-I error constraints, we show in this chapter that an interesting tradeoff arises in the setup under expected-rate constraints between the decision centers that have different type-I error constraints. The expected-rate constraints give an additional dimension with different margins to exploit for positive acceptable type-I error thresholds at the different decision centers.

In order to obtain the best tradeoff between all decision centers, we proposed a multiplexing and rate-sharing strategy that can achieve the fundamental exponents region under expected-rate constraints. The idea is to multiplex different versions of the best available coding schemes for the same setup under maximum-rate constraints, while varying the distribution of the available rate for the communication links between these different versions depending on the favorable decision centers and on the allowed type-I error probabilities. Since the optimal coding schemes for any number of hops under maximum-rate constraints with vanishing type-I error probabilities already exist, we showed that this proposed strategy achieves the fundamental exponents region. However this strategy can still be applied with generic schemes and it can still achieve remarkable gains compared to the achievable exponents region by the given generic schemes.

Moreover, to prove that the achievable region by our proposed strategy is the fundamental exponents region for the two-hop network, a novel converse proof method is proposed that builds up on the techniques used in the previous chapter such as change of measure arguments, asymptotic Markov chains, blowing-up lemma, and laws of probability on expectation, but also applies other new features. The distinguishing feature here is dividing the set of all observations at all terminals (except the receiver) into disjoint sets where each set corresponds to the observations that yield a positive probability of avoiding type-I error at a certain subset of the decision centers. Change of measure arguments are then applied to restrict to random variables in the given set with positive probability, on which the blowing-up lemma is then applied along with standard inequalities that yield the minimization expressions in the upper-bounds on the error

exponents of the related DCs. For the lower bounds on the rates of any given link, the laws of probability on expectation for all the related sets (i.e. the sets that require communication over the given link) are applied. Finally for the desired Markov chains, the first one is proved asymptotically in the limit as $n \rightarrow \infty$, while the second one holds by construction. These techniques can also be used under maximum-rate constraints for a simpler strong converse proof than the existing one in [35].

For the K -hop setup, we can generalize the same proof method for the desired converse and strong converse results under expected-rate and maximum-rate constraints, respectively. However, we further provide an alternative and simpler proof without using the blowing-up lemma. In this case, the set conditioning is done over all observations at all terminals including the receiver. Thus, we do not need the blowing-up lemma but yet we do not directly get the last Markov chain, but we still can prove this Markov chain asymptotically. Notice that since we already need the same proof steps to prove the $K - 2$ previous Markov chains which are different from the proof of the first Markov chain, then for the K hops it is definitely simpler to consider this alternative proof. As we show in our very recent work in [43], this method can be promising for the strong converse proofs of other setups even in the source coding problems.

Chapter 5

Single Sensor to Multiple DCs

5.1 Introduction

In this chapter, we consider an instance of the distributed binary hypothesis testing problem with multiple DCs. A single sensor communicates to the DCs over a common broadcast link and over individual links to each of them. This problem has been studied under various assumptions in the literature, and we review these previous results in the next section. The work most related to ours is [28] that establishes the fundamental type-II error exponents region under maximum-rate constraints and for testing against independence when the type-I error probabilities vanish. In this chapter, we establish a strong converse result to this problem and also characterize the fundamental exponents region under expected-rate constraints. In the special case of having only a common broadcast link, we provide a simplified representation of the exponents region which also simplifies the required optimal coding scheme to achieve it.

Our main results in this chapter, which are presented ahead in Section 5.5.2, show that the fundamental exponents region is boosted under expected-rate constraints in comparison with the region under maximum-rate constraints, and where a tradeoff is present between the type-II error exponents at the different DCs. The main difference to the multi-hop setup in the previous chapter is that a tradeoff was already present in the setup under maximum-rate constraints. This tradeoff is induced by the common BC channel which has to serve all DCs at the same time. The other type of tradeoff that is present in this chapter is similar to that of the previous chapter, and stems from having different type-I error thresholds resulting in competing margins to be exploited by the different DCs. The findings in this chapter are shown for two DCs for simplicity, however they can be extended to any number of DCs.

5.2 Related Works

Distributed hypothesis testing with a single sensor communicating to multiple DCs has been studied in several works [28–32] under the assumption of maximum-rate constraints. The work in [28] is most related to ours [53], and we review it at the end of this section. We first present an overview of other related works under two instances of the problem: “coherent detection” and “concurrent detection” as termed by the authors in [32]. Under coherent detection, the different DCs need to maximize their error exponents under the same hypothesis; they share the same goal and classification of critical events. However under concurrent detection, different DCs have to maximize their error exponents under different hypotheses.

For coherent detection, the work of [28] focused on having a single sensor communicating with two DCs over three noise-free links: one common broadcasting link to both DCs, and two individual links dedicated for each decision center. In the special case of testing against independence, the authors in [28] exactly characterized the fundamental error exponents region which shows a tradeoff between the type-II error exponents of the two DCs. Their work is extended in [29] to testing against conditional independence when the communication occurs over one common and two individual noise-free links or over a noisy broadcast channel. Under noise-free communication, the authors in [29] derived the fundamental error exponents region for some special cases of the problem where a tradeoff also arises between the type-II error exponents at the DCs. The setup was further studied with a cooperative link from the first decision center to the second decision center in [31, 32]. In addition to the derivation of inner bounds to the exponents region, the authors provided an exact characterization of the exponents region when testing against independence with zero-cooperation rate between the two DCs [31, 32]; a tradeoff arises in this case between the error exponents at the two decision centers. The authors also characterized the fundamental exponents region for the general binary hypothesis testing setup when the communication from the sensor to the two DCs is limited by asymptotically zero-rate constraints; in this case, the exponents region is rectangular and thus both DCs can maximize their type-II error exponents simultaneously.

For the setup with concurrent detection, the authors in [30–32], considered only noise-free communication links in their works with and without a cooperative link. In the absence of the cooperative link under concurrent detection, the setup is extended to K decision centers in [30] where both simple and composite M -ary hypothesis testing are studied. In addition to the derivation of inner bounds to the exponents region for all the mentioned scenarios, their main results include exact characterizations of the exponents region when the communication from the sensor to the decision centers is limited by asymptotically zero-rate con-

straints. In this case, a tradeoff between the error exponents arises whenever the number of communicated bits is not larger than $\log_2 |\mathcal{H}|$, where $|\mathcal{H}|$ denotes the number of possible hypotheses, and otherwise all decision centers can maximize their error exponents simultaneously. In [32], the authors derived another optimality result for a special case under the cooperative concurrent detection with two DCs. In this special case, the marginal distribution of the sensor's observations is different under the two hypotheses and no tradeoff is present between the error exponents of the two DCs.

As mentioned in the beginning of this section, the most related work to ours is the work of Wigger and Timo [28]. The main difference in our work is having the expected-rate instead of the maximum-rate constraints for the setup of distributed binary hypothesis testing against independence. We consider a single sensor communicating over a common BC channel and two private links with two decision centers under coherent detection, i.e. both decision centers wish to maximize their error exponents under the same hypothesis. While a tradeoff already arises in the work of Wigger and Timo [28] under maximum-rate constraints, mainly due to broadcasting common information to both decision centers at the same time, the different type-I error thresholds contribute to an additional tradeoff under expected-rate constraints. We present the system model under both types of rate constraints in the next section followed by the existing optimal coding scheme and exponents region under maximum-rate constraints, see Sections 5.4.1, 5.4.2. In Section 5.4.3, we provide a strong converse result to the main result of [28]. Our work and main results under expected-rate constraints are then presented in Section 5.5.

5.3 The Setup

Consider the distributed hypothesis testing problem in Figure 5.1 in the special case of testing against independence, i.e., depending on the binary hypothesis $\mathcal{H} \in \{0, 1\}$, the tuple (Y_0^n, Y_1^n, Y_2^n) is distributed as:

$$\text{under } \mathcal{H} = 0 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1 Y_2 | Y_0}; \quad (5.1a)$$

$$\text{under } \mathcal{H} = 1 : (Y_0^n, Y_1^n, Y_2^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1 Y_2} \quad (5.1b)$$

for given pmfs P_{Y_0} and $P_{Y_1 Y_2 | Y_0}$ and where $P_{Y_1 Y_2}$ denotes the marginal of the joint pmf $P_{Y_0 Y_1 Y_2} := P_{Y_0} P_{Y_1 Y_2 | Y_0}$.

The system consists of a transmitter T_0 , and two receivers R_1 and R_2 . Transmitter T_0 observes the source sequence Y_0^n and computes three bit-string messages $(M_0, M_1, M_2) = \phi^{(n)}(Y_0^n)$, where the encoding

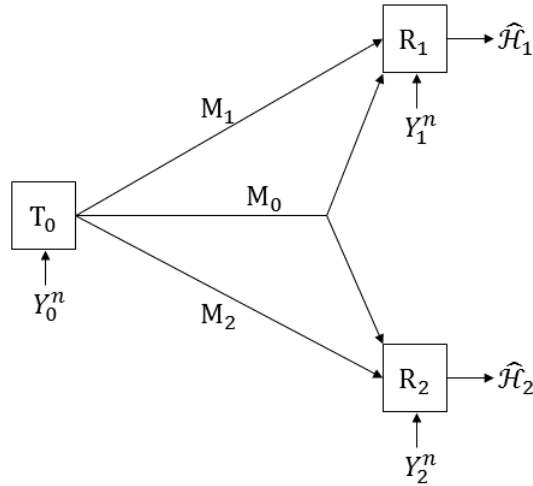


Figure 5.1: Distributed hypothesis testing with a single sensor and two remote decision centers with integrated sensors.

function is of the form $\phi^{(n)} : \mathcal{Y}_0^n \rightarrow \{0, 1\}^* \times \{0, 1\}^* \times \{0, 1\}^*$. Message M_0 is sent to both receivers R_1, R_2 , while message M_1 only to receiver R_1 and message M_2 only to receiver R_2 . The messages have to satisfy the given rate constraints. For maximum-rate constraints, the maximum length of any message should satisfy

$$\text{len}(M_i) \leq nR_i, \quad i \in \{0, 1, 2\}, \quad (5.2)$$

whereas for expected-rate constraints, the expected length of the messages should satisfy

$$\mathbb{E}[\text{len}(M_i)] \leq nR_i, \quad i \in \{0, 1, 2\}. \quad (5.3)$$

Receiver $R_i, i \in \{1, 2\}$, observes the source sequence Y_i^n and with messages M_0, M_i received from T_0 , it produces a guess $\hat{\mathcal{H}}_i$ of the hypothesis \mathcal{H} using a decision function $g_i^{(n)} : \mathcal{Y}_i^n \times \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}$:

$$\hat{\mathcal{H}}_i = g_i^{(n)}(Y_i^n, M_0, M_i) \in \{0, 1\}, \quad i \in \{1, 2\}. \quad (5.4)$$

The goal is to design encoding and decision functions such that their type-I error probabilities

$$\alpha_{i,n} \triangleq \Pr[\hat{\mathcal{H}}_i = 1 | \mathcal{H} = 0], \quad i \in \{1, 2\}, \quad (5.5)$$

stay below given thresholds $\epsilon_i > 0, i \in \{1, 2\}$, and the type-II error probabilities

$$\beta_{i,n} \triangleq \Pr[\hat{\mathcal{H}}_i = 0 | \mathcal{H} = 1], \quad i \in \{1, 2\}, \quad (5.6)$$

decay to 0 with largest possible exponential decay.

Definition 12. Fix maximum type-I error probabilities $\epsilon_1, \epsilon_2 \in [0, 1]$ and rates $R_1, R_2 \geq 0$. The exponent pair (θ_1, θ_2) is called (ϵ_1, ϵ_2) -achievable if there exists a sequence of encoding and decision functions

$\{\phi^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$ satisfying:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{i,n} \leq \epsilon_i, \quad i \in \{1, 2\} \quad (5.7a)$$

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_{i,n}} \geq \theta_i, \quad i \in \{1, 2\}, \quad (5.7b)$$

and

$$\text{len}(\mathbf{M}_i) \leq nR_i, \quad i \in \{0, 1, 2\}, \quad (5.7c)$$

for the setup under maximum-rate constraints, or

$$\mathbb{E}[\text{len}(\mathbf{M}_i)] \leq nR_i, \quad i \in \{0, 1, 2\}, \quad (5.7d)$$

for the setup under expected-rate constraints.

Definition 13. The closure of the set of all (ϵ_1, ϵ_2) -achievable exponent pairs (θ_1, θ_2) is called the fundamental (ϵ_1, ϵ_2) -exponents region and is denoted $\mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ for the setup under maximum-rate constraints and $\mathcal{E}_{\text{BC}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ for the setup under expected-rate constraints.

5.4 Maximum-Rate Constraints

In the following subsection, we present the optimal coding scheme suggested by [28] that can achieve the fundamental exponents region $\mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$. Wigger and Timo [28] have derived a tight upper bound for this region in the limit $\epsilon_1, \epsilon_2 \downarrow 0$. Their result is reported in Subsection 5.4.2. However, as we show later in Theorem 15 in Subsection 5.4.3, we strengthen this result by providing a strong converse that shows that this exponents region is independent of the type-I error probability thresholds ϵ_1, ϵ_2 , and thus it is tight even for non-vanishing type-I error probabilities.

5.4.1 Optimal Coding Scheme under Maximum-Rate Constraints

The optimal coding scheme proposed by [28] for single-sensor two-DCs setup communicating over a common broadcast link and two individual links applies superposition coding on the common message \mathbf{M}_0 sent to both decision centers. Three codebooks are present at T_0 , a common codebook and two superposed codebooks for which, each one intended to a DC. The details of the coding scheme are as follows.

Fix a large blocklength n , small numbers $\mu' < \mu''$, and conditional pmfs $P_{U_0|Y_0}$, $P_{U_1U_2|U_0,Y_0}$ satisfying

the maximum-rate constraints

$$R_0 \geq I(U_0; Y_0) \quad (5.8)$$

$$R_1 \geq I(U_1; Y_0|U_0) \quad (5.9)$$

$$R_2 \geq I(U_2; Y_0|U_0) \quad (5.10)$$

where mutual information quantities are meant with respect to

$$P_{U_0U_1U_2Y_0Y_1Y_2} \triangleq P_{U_0|Y_0} \cdot P_{U_1U_2|U_0Y_0} \cdot P_{Y_0} \cdot P_{Y_1Y_2|Y_0}. \quad (5.11)$$

Randomly generate a codebook

$$C_{U_0} \triangleq \{u_0^n(m_0) : m_0 \in \{1, \dots, \lfloor 2^{nR_0} \rfloor\}\} \quad (5.12)$$

by drawing all entries i.i.d. according to the marginal pmf P_{U_0} . For each codeword $u_0^n(m_0)$, generate the codebooks

$$C_{U_i}(m_0) \triangleq \{u_i^n(m_i|m_0) : m_i \in \{1, \dots, \lfloor 2^{nR_i} \rfloor\}\}, \quad i \in \{1, 2\}, \quad (5.13)$$

by drawing the j -th entry of each codeword according to the conditional marginal pmf $P_{U_i|U_0}$.

All codebooks are revealed to all terminals.

T₀: Assume it observes the sequence $Y_0^n = y_0^n$. Then it looks for a tuple of indices (m_0, m_1, m_2) satisfying

$$(u_0^n(m_0), u_1^n(m_1|m_0), u_2^n(m_2|m_0), y_0^n) \in \mathcal{T}_{\mu'}^n(P_{U_0U_1U_2Y_0}),$$

randomly picks one of these tuples, and sends the corresponding bit-string messages $\mathbf{M}_0, \mathbf{M}_1$ to \mathbf{R}_1 , and $\mathbf{M}_0, \mathbf{M}_2$ to \mathbf{R}_2 . Otherwise, it sends over all links

$$\mathbf{M}_0 = \mathbf{M}_1 = \mathbf{M}_2 = [0]. \quad (5.14)$$

R₁: Assume it observes the sequence $Y_1^n = y_1^n$ and receives the messages $\mathbf{M}_0 = \mathbf{m}_0$ and $\mathbf{M}_1 = \mathbf{m}_1$ from \mathbf{T}_0 .

If $\mathbf{m}_0 = 0$, then \mathbf{R}_1 declares the decision

$$\hat{\mathcal{H}}_1 = 1. \quad (5.15)$$

Else, it sets $m_0 = \text{dec}(\mathbf{m}_0)$, $m_1 = \text{dec}(\mathbf{m}_1)$ and performs the typicality check

$$(u_0^n(m_0), u_1^n(m_1|m_0), y_1^n) \in \mathcal{T}_{\mu'}^n(P_{U_0U_1Y_1}).$$

If the check is successful, then it declares

$$\hat{\mathcal{H}}_1 = 0, \quad (5.16)$$

and otherwise it declares

$$\hat{\mathcal{H}}_1 = 1. \quad (5.17)$$

R₂: Assume it observes the sequence $Y_2^n = y_2^n$ and receives messages $M_0 = m_0$ and $M_2 = m_2$. If $m_0 = [0]$, it declares the decision

$$\hat{\mathcal{H}}_2 = 1. \quad (5.18)$$

Else, it sets $m_0 = \text{dec}(m_0)$, $m_2 = \text{dec}(m_2)$, and performs the typicality check

$$(u_0^n(m_0), u_2^n(m_2|m_0), y_2^n) \in \mathcal{T}_{\mu^n}(P_{U_0 U_2 Y_2}).$$

If the check is successful, then it declares

$$\hat{\mathcal{H}}_2 = 0, \quad (5.19)$$

and otherwise, it declares

$$\hat{\mathcal{H}}_2 = 1. \quad (5.20)$$

In the next subsection, we recall the fundamental exponents region obtained by this coding scheme for the setup under maximum-rate constraints.

5.4.2 The Exponents Region

Under the maximum-rate constraints (5.2), the exponents region $\mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ in the limit as $\epsilon_1, \epsilon_2 \downarrow 0$ was given by the following theorem. In the following, we abbreviate $\lim_{\epsilon_1, \epsilon_2 \downarrow 0} \mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ by $\mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, 0, 0)$

Theorem 14 (Theorem 1 in [28]). *In the limit as $\epsilon_1, \epsilon_2 \downarrow 0$, the fundamental exponents region $\mathcal{E}_{\text{max}}^*(R_0, R_1, R_2, 0, 0)$ is the set of all exponent pairs (θ_1, θ_2) satisfying:*

$$\theta_i \leq I(U_0 U_i; Y_i), \quad i \in \{1, 2\}, \quad (5.21a)$$

for some conditional pmfs $P_{U_0|Y_0}$, $P_{U_i|Y_0}$ satisfying

$$R_0 \geq I(U_0; Y_0), \quad (5.21b)$$

$$R_i \geq I(U_i; Y_0|U_0), \quad i \in \{1, 2\}, \quad (5.21c)$$

and where mutual information quantities are meant with respect to the joint pmf

$$P_{U_0 U_1 U_2 Y_0 Y_1 Y_2} \triangleq P_{U_0|Y_0} \cdot P_{U_1 U_2|U_0 Y_0} \cdot P_{Y_0} \cdot P_{Y_1 Y_2|Y_0}. \quad (5.21d)$$

Proof: See [28]. ■

Notice that the auxiliary random variable U_0 is present in the upper bound expressions for both error exponents. This indicates that a tradeoff might arise between the two error exponents depending on the choice of U_0 in the codebook construction for the communication over the common BC link. The intuition behind this tradeoff stems from the fact that there are different observations present at the two DCs, and thus choosing U_0 in a way that maximizes the mutual information quantity $I(U_0U_i; Y_i)$ for one DC might lead to a compromise in the quantity at the other DC depending on the problem parameters. We provide a numerical example later in Section 5.6.1 where we can visualize this tradeoff.

5.4.3 Strong Converse Result

In the limit $\epsilon_1, \epsilon_2 \downarrow 0$, the exponents region $\mathcal{E}_{\max}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ was determined in [28] and is presented here in this thesis in Section 5.4.2. In our work, we strengthen this result by providing a strong converse result.

Theorem 15. *Under the maximum-rate constraints (5.2), the exponents region $\mathcal{E}_{\max}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ is independent of (ϵ_1, ϵ_2) for all $\epsilon_1 + \epsilon_2 < 1$, and equals the set of (θ_1, θ_2) pairs satisfying:*

$$\theta_i \leq I(U_0U_i; Y_i), \quad i \in \{1, 2\}, \quad (5.22a)$$

for some conditional pmfs $P_{U_0|Y_0}, P_{U_i|Y_0}$ satisfying

$$R_0 \geq I(U_0; Y_0), \quad (5.22b)$$

$$R_i \geq I(U_i; Y_0|U_0), \quad i \in \{1, 2\}. \quad (5.22c)$$

Proof: Achievability is proved in [28]. The converse is proved in Appendix C.3. ■

5.5 Expected-Rate Constraints

We present in the next subsection our optimal coding scheme followed by the fundamental exponents region for the setup under expected-rate constraints. The scheme is based on multiplexing different versions of the optimal coding scheme of the previous section that achieves the fundamental exponents region under maximum-rate constraints, and by the rate-sharing strategy proposed in the previous chapter.

5.5.1 Optimal Coding Scheme under Expected Rate Constraints

In the optimal coding scheme under expected-rate constraints, the three terminals T_0, R_1, R_2 multiplex four different subschemes, and the choice of which subscheme to use depends on the transmitter T_0 's observations Y_0^n . To inform all terminals about the choice of the subscheme, T_0 adds two flag-bits to its messages.

The main distinguishing feature of the different subschemes is the choice of the subset of terminals—either only R_1 or only R_2 , both R_1 and R_2 , or neither of them—which exploits the information in the transmitted messages to produce a guess on the hypothesis \mathcal{H} . The other terminals ignore this communication and simply declare $\hat{\mathcal{H}} = 1$. The different subschemes occupy different communication rates, and as we shall see in Section 4.3.3.2, the allocation of the rates has to be chosen in function of the desired tradeoff between the exponents θ_1 and θ_2 .

To this end, we choose random variables $U_{\{1,2\},0}, U_{\{1,2\},1}, U_{\{1,2\},2}, U_{\{1\},0}, U_{\{1\},1}, U_{\{2\},0}, U_{\{2\},2}$, rates $R_{\{1,2\},0}, R_{\{1,2\},1}, R_{\{1,2\},2}, R_{\{1\},0}, R_{\{1\},1}, R_{\{2\},0}, R_{\{2\},2}$, and probabilities $\sigma_{\{1,2\}}, \sigma_{\{1\}}, \sigma_{\{2\}}$, so that $\sigma_{\{1,2\}} + \sigma_{\{1\}} + \sigma_{\{2\}} \leq 1$, and the following conditions are satisfied for $i \in \{1, 2\}$ and $\mathcal{I} \in \mathcal{P}(2)$:

$$R_0 \geq \sigma_{\{1,2\}} R_{\{1,2\},0} + \sigma_{\{1\}} R_{\{1\},0} + \sigma_{\{2\}} R_{\{2\},0}, \quad (5.23)$$

$$R_i \geq \sigma_{\{1,2\}} R_{\{1,2\},i} + \sigma_{\{i\}} R_{\{i\},i}, \quad (5.24)$$

$$R_{\mathcal{I},0} \geq I(U_{\mathcal{I},0}; Y_0), \quad (5.25)$$

$$R_{\{1,2\},i} \geq I(U_{\{1,2\},i}; Y_0 | U_{\{1,2\},0}), \quad (5.26)$$

$$R_{\{i\},i} \geq I(U_{\{i\},i}; Y_0 | U_{\{i\},0}), \quad (5.27)$$

$$\sigma_{\{1,2\}} + \sigma_{\{i\}} \geq 1 - \epsilon_i, \quad (5.28)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}. \quad (5.29)$$

Then we partition the set \mathcal{Y}_0^n into four disjoint subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{1\}}, \mathcal{D}_{\{2\}}, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under P_{Y_0} the three sets $\mathcal{D}_{\{1\}}, \mathcal{D}_{\{2\}}$ and $\mathcal{D}_{\{1,2\}}$ have largest possible probabilities but limited by

$$\Pr [Y_0^n \in \mathcal{D}_{\{1\}}] \leq \sigma_{\{1\}} \quad (5.30a)$$

$$\Pr [Y_0^n \in \mathcal{D}_{\{2\}}] \leq \sigma_{\{2\}} \quad (5.30b)$$

$$\Pr [Y_0^n \in \mathcal{D}_{\{1,2\}}] \leq \sigma_{\{1,2\}}. \quad (5.30c)$$

As a consequence,

$$\Pr [Y_0^n \in \mathcal{D}_\emptyset] \geq 1 - (\sigma_{\{1\}} + \sigma_{\{2\}} + \sigma_{\{1,2\}}). \quad (5.30d)$$

Notice that as $n \rightarrow \infty$, the four inequalities (5.30) can hold with equality.

Depending on whether Y_0^n lies in \mathcal{D}_\emptyset , $\mathcal{D}_{\{1\}}$, $\mathcal{D}_{\{2\}}$, or $\mathcal{D}_{\{1,2\}}$, the three terminals apply a different subscheme, where the subscript \mathcal{I} of set $\mathcal{D}_{\mathcal{I}}$ indicates the set of receivers that attempt to correctly guess \mathcal{H} in the event $Y_0^n \in \mathcal{D}_{\mathcal{I}}$. To communicate which of the four subschemes is used, T_0 adds a two-bit flag at the beginning of its common message \mathbf{M}_0 to R_1 and R_2 . The main structure of the coding scheme is illustrated in Figure 5.2.

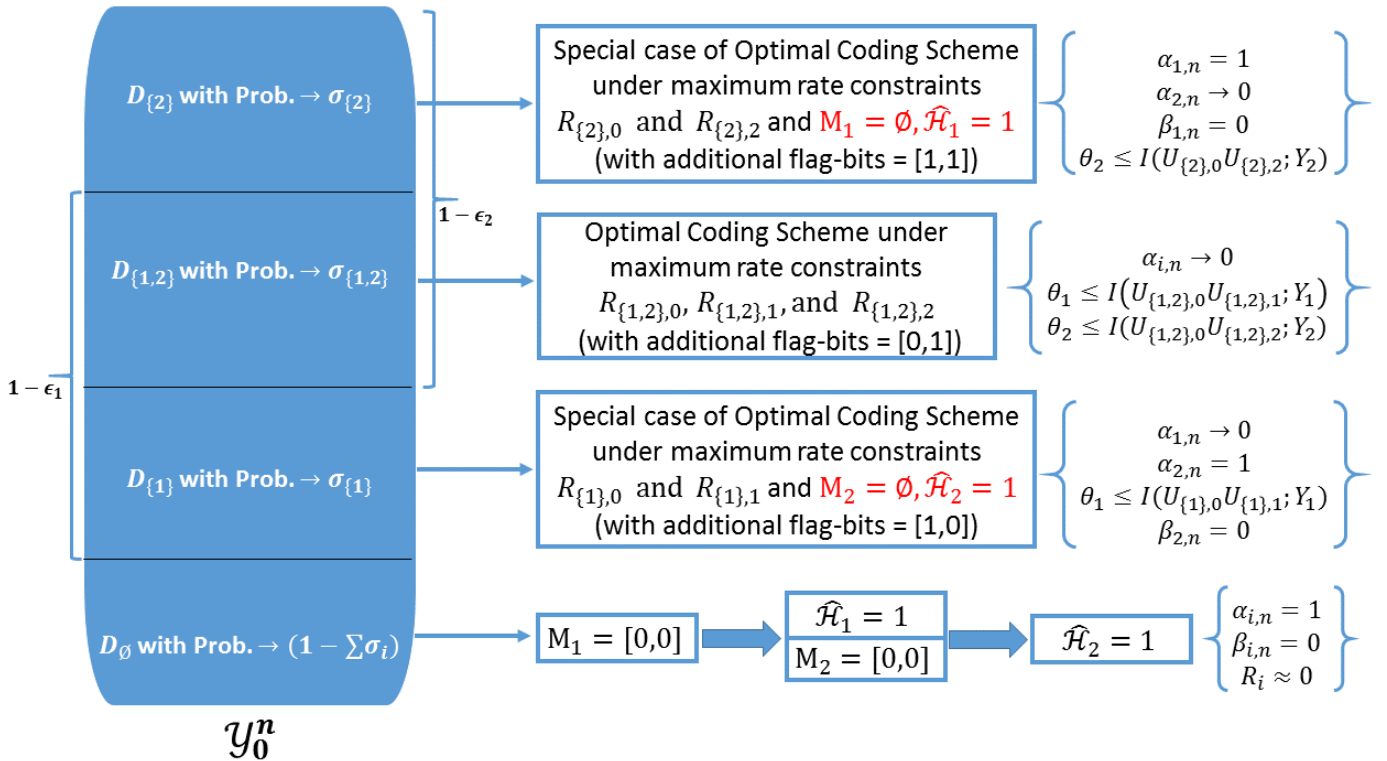


Figure 5.2: Optimal coding scheme for the setup with a single-sensor and two decision centers under expected-rate constraints using the optimal coding scheme under maximum-rate constraints.

If $Y_0^n \in \mathcal{D}_\emptyset$: Transmitter T_0 sends only the flag-bits

$$\mathbf{M}_0 = [0, 0] \quad (5.31)$$

over the common link and nothing over the individual links $\mathbf{M}_1 = \mathbf{M}_2 = \emptyset$. Upon receiving these messages, both decision centers R_1 and R_2 decide on

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1. \quad (5.32)$$

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: T_0 , R_1 , and R_2 all follow the coding scheme with vanishing type-I error probability as described in Section 5.4.1 with the choice of auxiliaries $U_{\{1,2\},0}, U_{\{1,2\},1}, U_{\{1,2\},2}$ and rates $R_{\{1,2\},0}, R_{\{1,2\},1}, R_{\{1,2\},2}$. Additionally, T_0 adds $[0,1]$ -flag bits to the common messages \mathbf{M}_0 to indicate

to R_1 and R_2 that $Y_0^n \in \mathcal{D}_{\{1,2\}}$.

If $Y_0^n \in \mathcal{D}_{\{1\}}$: T_0 and R_1 follow a special case of the coding scheme with vanishing type-I error probability described in Section 5.4.1, where no message is sent over the individual link to R_2 , i.e., $M_2 = \emptyset$. This coding scheme is applied with the choice of auxiliaries $U_{\{1\},0}, U_{\{1\},1}$ and rates $R_{\{1\},0}, R_{\{1\},1}$. Additionally, T_0 adds [1,0]-flag bits to its common messages M_0 to indicate to R_1 and R_2 that $Y_0^n \in \mathcal{D}_{\{1\}}$. We observe that, as indicated by the subscript $\{1\}$ of set $\mathcal{D}_{\{1\}}$, only terminal R_1 attempts to correctly guess \mathcal{H} . Receiver R_2 produces the trivial guess

$$\hat{\mathcal{H}}_2 = 1. \quad (5.33)$$

If $Y_0^n \in \mathcal{D}_{\{2\}}$: Now, T_0 and R_2 follow another special case of the coding scheme with vanishing type-I error probability described in Section 5.4.1, where no message is sent over the individual link to R_1 , i.e., $M_1 = \emptyset$. This coding scheme is applied with the choice of auxiliaries $U_{\{2\},0}, U_{\{2\},2}$ and rates $R_{\{2\},0}, R_{\{2\},2}$. Additionally, T_0 adds [1,1]-flag bits to its common messages M_0 to indicate to R_1 and R_2 that $Y_0^n \in \mathcal{D}_{\{2\}}$. As indicated by the subscript $\{2\}$ of set $\mathcal{D}_{\{2\}}$, only terminal R_2 attempts to correctly guess \mathcal{H} , thus R_1 produces the trivial guess

$$\hat{\mathcal{H}}_1 = 1. \quad (5.34)$$

In summary, this means that for each set $\mathcal{D}_{\mathcal{I}}$ ($\mathcal{I} \in \mathcal{P}(2)$), we apply the optimal coding scheme under maximum-rate constraints in [28] as described in Section 5.4.1, but for each set $\mathcal{D}_{\mathcal{I}}$ we construct different codebooks and use different auxiliaries $U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2}$. In particular, we chose $U_{\{1\},2}$ and $U_{\{2\},1}$ constants, indicating that when $Y_0^n \in \mathcal{D}_{\{i\}}$ then only messages (M_0, M_i) are sent, for $i \in \{1, 2\}$. When $Y_0^n \in \mathcal{D}_{\{1,2\}}$, then all three messages M_0, M_1, M_2 are sent.

5.5.2 The Exponents Region

Our main results are a complete characterization of the fundamental exponents region $\mathcal{E}_{\text{BC}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ under the *expected*-rate constraints in (5.3) which is presented in this section, in addition to a strong converse result under the *maximum*-rate constraints (5.2), which was presented in Section 5.4.3. A main observation to our results, is the new element in the tradeoff between the error exponents of the decision centers θ_1 and θ_2 . While in the previous chapter we highlight the resulting tradeoff from having multiple decision centers with different type-I error constraints, here we have an additional tradeoff due to the common broadcast link from the transmitter to both decision centers. As shown in the previous results presented in Section 5.4.2, in this case even under maximum-rate constraints with vanishing type-I error probabilities, there is a tradeoff

between the decision centers. In the following two subsections, we present our results on the exponents region in the general case with individual and common communication links, and in the special case of having only a common communication link. In the later case, the fundamental exponents region is denoted by $\mathcal{E}_{\text{BC}}^*(R_0, 0, 0, \epsilon_1, \epsilon_2)$, for which we prove that the expression of its characterization simplifies and thus the optimal coding scheme simplifies as well.

5.5.2.1 Individual and Common Communication Links

Theorem 16. *The fundamental (ϵ_1, ϵ_2) -exponents region $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ is the set of all (θ_1, θ_2) pairs satisfying*

$$\theta_i \leq \min \{I(U_{\{1,2\},0}U_{\{1,2\},i}; Y_i), I(U_{\{i\},0}U_{\{i\},i}; Y_i)\}, \quad i \in \{1, 2\} \quad (5.35a)$$

for some nonnegative numbers $\sigma_{\{1,2\}}, \sigma_{\{1\}}, \sigma_{\{2\}}$ with sum ≤ 1 and conditional pmfs $P_{U_{\{1,2\},0}|Y_0}, P_{U_{\{1\},0}|Y_0}, P_{U_{\{2\},0}|Y_0}, P_{U_{\{1,2\},1}|U_{\{1,2\},0}Y_0}, P_{U_{\{1\},1}|U_{\{1\},0}Y_0}, P_{U_{\{1,2\},2}|U_{\{1,2\},0}Y_0}, P_{U_{\{2\},2}|U_{\{2\},0}Y_0}$ satisfying

$$R_0 \geq \sigma_{\{1,2\}}I(U_{\{1,2\},0}; Y_0) + \sigma_{\{1\}}I(U_{\{1\},0}; Y_0) + \sigma_{\{2\}}I(U_{\{2\},0}; Y_0), \quad (5.35b)$$

$$R_i \geq \sigma_{\{1,2\}}I(U_{\{1,2\},i}; Y_0|U_{\{1,2\},0}) + \sigma_{\{i\}}I(U_{\{i\},i}; Y_0|U_{\{i\},0}), \quad i \in \{1, 2\}, \quad (5.35c)$$

$$\text{and } \sigma_{\{1,2\}} + \sigma_{\{i\}} \geq 1 - \epsilon_i, \quad i \in \{1, 2\}, \quad (5.35d)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}, \quad (5.35e)$$

and where the mutual information quantities are calculated according to the joint pmfs

$$P_{Y_0Y_1Y_2U_{\{1,2\},0}U_{\{1,2\},1}U_{\{1,2\},2}} \triangleq P_{Y_0Y_1Y_2}P_{U_{\{1,2\},0}|Y_0}P_{U_{\{1,2\},1}U_{\{1,2\},2}|U_{\{1,2\},0}Y_0} \quad (5.36)$$

$$P_{Y_0Y_1Y_2U_{\{i\},0}U_{\{i\},i}} \triangleq P_{Y_0Y_1Y_2}P_{U_{\{i\},0}|Y_0}P_{U_{\{i\},i}|U_{\{i\},0}Y_0}, \quad i \in \{1, 2\}. \quad (5.37)$$

Proof: The achievability is based on the coding scheme presented in Section 5.5.1 and analyzed in Appendix C.1. The converse is proved in Section 5.7. \blacksquare

5.5.2.2 Only a Common Communication Link

For $R_1 = R_2 = 0$, i.e., without individual communication links, we can simplify the expression for $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$.

Definition 14. Define the two functions

$$\eta_i(R_{\{i\},0}) := \max_{\substack{P_{U_{\{i\},0}|Y_0}: \\ R_{\{i\},0} \geq I(U_{\{i\},0}; Y_0)}} I(U_{\{i\},0}; Y_i), \quad i \in \{1, 2\}, \quad (5.38)$$

where the mutual information quantities are calculated with respect to the joint pmf $P_{U_{\{i\},0}Y_0Y_1Y_2} \triangleq P_{U_{\{i\},0}|Y_0}P_{Y_0Y_1Y_2}$.

Corollary 1. *Let $\pi: \{1, 2\} \rightarrow \{1, 2\}$ be a permutation ordering the ϵ -values in decreasing order:*

$$\epsilon_{\pi(1)} \geq \epsilon_{\pi(2)}. \quad (5.39)$$

Then $\mathcal{E}^*(R_0, 0, 0, \epsilon_1, \epsilon_2)$ is the set of all (θ_1, θ_2) pairs satisfying

$$\theta_{\pi(1)} \leq I(U_0; Y_{\pi(1)}), \quad (5.40a)$$

$$\theta_{\pi(2)} \leq \min \{I(U_0; Y_{\pi(2)}), \eta_{\pi(2)}(R_{\{\pi(2)\},0})\}, \quad (5.40b)$$

for some conditional pmf $P_{U_0|Y_0}$ and rate $R_{\{\pi(2)\},0}$ satisfying

$$R_0 \geq (1 - \epsilon_{\pi(1)})I(U_0; Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})R_{\{\pi(2)\},0}. \quad (5.40c)$$

Proof: See Appendix C.4. ■

5.6 Comparison between Maximum and Expected Rate Constraints

Theorem 16 shows a tradeoff between the two achievable exponents θ_1 and θ_2 . (Figure 5.3 ahead illustrates this tradeoff at hand of a numerical example in the special case $R_1 = R_2 = 0$.) The tradeoff stems technically from the common random variable $U_{\{1,2\},0}$ that is included in the exponent constraint (5.35a) for both $i \in \{1, 2\}$, and from the rate-sharing of the optimal coding scheme under maximum-rate constraints presented in Subsection 5.4.1 for three different choices of $(\sigma_{\mathcal{I}}, U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2})$, for $\mathcal{I} \in \mathcal{P}(2)$.

To see the effect of the expected-rate constraint in (5.3), we compare the fundamental exponents region $\mathcal{E}_{\text{BC}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ with the fundamental exponents region $\mathcal{E}_{\text{BC,max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ under the more stringent maximum-rate constraints (5.2). To this end, notice that (5.22) is obtained from (5.35) by setting $\sigma_{\{1,2\}} = 1$ and thus $\sigma_{\{1\}} = \sigma_{\{2\}} = 0$, which allows us to freely choose the auxiliaries $U_{\{1\},0}, U_{\{2\},0}, U_{\{1\},1}, U_{\{2\},2}$ to maximize the exponent terms corresponding to them by setting $U_{\{i\},0} = U_{\{i\},i} = Y_i$. This is feasible because setting $\sigma_{\{1\}} = \sigma_{\{2\}} = 0$ in (5.35b) and (5.35c) removes any constraints on the choice of the auxiliaries $U_{\{1\},0}, U_{\{2\},0}, U_{\{1\},1}, U_{\{2\},2}$. The notation is then updated by removing the subscript “ $\{1, 2\}$ ” from the remaining terms. Now, due to the strong converse result in Theorem 15, we notice that the fundamental exponents region under maximum-rate constraints for any values of (ϵ_1, ϵ_2) is equal to the fundamental exponents region under expected-rate constraints with zero type-I error thresholds (ϵ_1, ϵ_2) , i.e. $\mathcal{E}_{\text{max}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2) = \mathcal{E}^*(R_0, R_1, R_2, 0, 0)$. Since $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ is

generally increasing in (ϵ_1, ϵ_2) , we conclude that expected-rate constraints allow to boost the exponents region compared to maximum-rate constraints. This conclusion is illustrated in the next subsection at hand of an example for a special case when only the common link between the transmitter and both receivers is present.

5.6.1 Numerical Simulation

The following example illustrates the benefits of expected-rate constraints versus maximum-rate constraints, and the tradeoff between the two exponents in the special case when $R_1 = R_2 = 0$.

Example 4. Consider the following joint pmf $P_{Y_0 Y_1 Y_2}$:

	(Y_1, Y_2)			
Y_0	(0, 0)	(0, 1)	(1, 0)	(1, 1)
0	0.05	0.05	0.15	0.083325
1	0.05	0.15	0.05	0.08335
2	0.15	0.05	0.05	0.083325

For this pmf, Figure 5.3 shows the optimal exponents regions under maximum- and expected-rate constraints when $R_0 = 0.1$ and $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$. The figure illustrates the boost in the exponents region due to the expected-rate constraints. It also emphasizes the benefits of sharing the rate in (5.40c) between two summands, which relate to the fact that depending on the observation Y_0^n we use two variants of the coding scheme in [28] (see Section 5.4.1), one with auxiliary U_0 and the other with an auxiliary $U_{\{\pi(2)\},0}$ that satisfies $I(U_{\{\pi(2)\},0}; Y_0) \leq R_{\{\pi(2)\},0}$ and $I(U_{\{\pi(2)\},0}; Y_1) = \eta_{\pi(2)}(R_{\{\pi(2)\},0})$. Restricting to a single auxiliary U_0 in (5.40) (i.e., no rate-sharing is applied and setting $R_{\{\pi(2)\},0} = I(U_0; Y_0)$) results in an exponents region, denoted $\mathcal{E}_{\text{no-RS}}(R_0, 0, 0, \epsilon_1, \epsilon_2)$ which coincides with $\mathcal{E}_{\text{BC}}^*(R_0, 0, 0, \epsilon_2, \epsilon_2)$ (i.e. when the type-I threshold ϵ_1 is restricted to be equal to ϵ_2 and thus to be smaller) and with $\mathcal{E}_{\text{BC,max}}^*((1 - \epsilon_2)^{-1}R_0, 0, 0, \epsilon_1, \epsilon_2)$ which corresponds to the optimal error exponent under a boosted maximum-rate constraint by the factor $(1 - \epsilon_2)^{-1}$. Note that in this example, $\pi(2) = 2$ since $\epsilon_1 > \epsilon_2$.

5.7 Converse Proof to Theorem 16

Fix an exponent pair in $\mathcal{E}_{\text{BC}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ and a sequence (in n) of encoding and decision functions $\{(\phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ satisfying the constraints on the rate and the error probabilities in (5.7). Our proof relies on the following lemma:

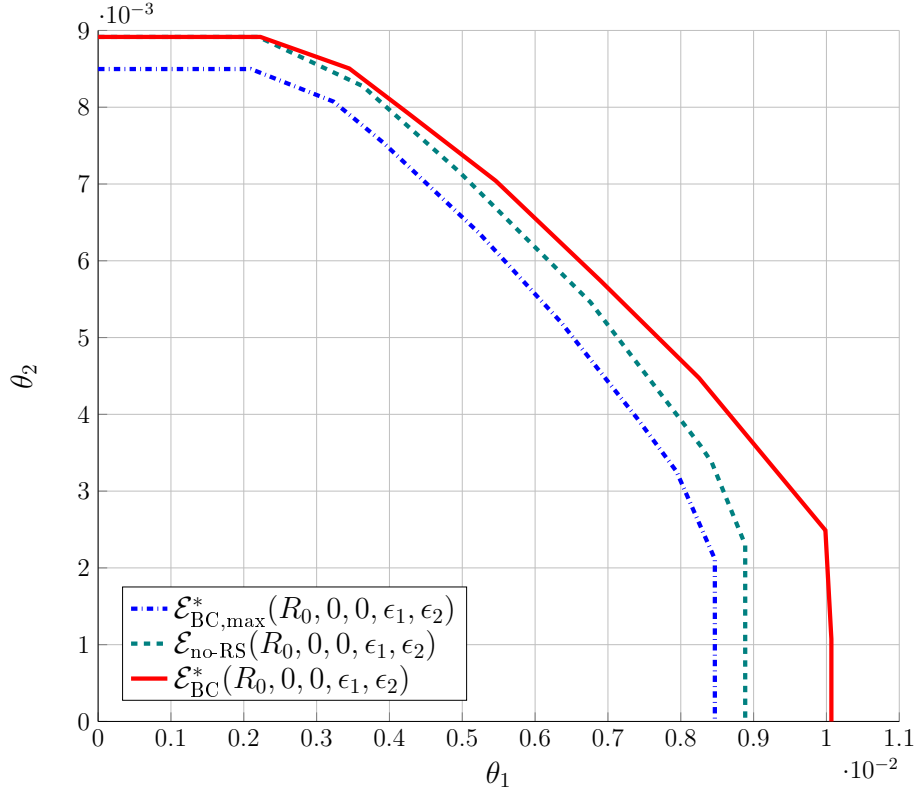


Figure 5.3: Optimal error exponents regions of the BC setup under expected- and maximum-rate constraints for $R_0 = 0.1$, $\epsilon_1 = 0.15$, $\epsilon_2 = 0.05$.

Lemma 9. Fix a blocklength n and a set $\mathcal{D} \subseteq \mathcal{Y}_0^n$ of positive probability, and let the tuple $(\tilde{M}_0, \tilde{M}_1, \tilde{M}_2, \tilde{Y}_0^n, \tilde{Y}_1^n, \tilde{Y}_2^n)$ follow the pmf

$$P_{\tilde{M}_0 \tilde{M}_1 \tilde{M}_2 \tilde{Y}_0^n \tilde{Y}_1^n \tilde{Y}_2^n}(\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2, y_0^n, y_1^n, y_2^n) \triangleq P_{Y_0^n Y_1^n Y_2^n}(y_0^n, y_1^n, y_2^n) \cdot \frac{\mathbb{1}\{y_0^n \in \mathcal{D}\}}{P_{Y_0^n}(\mathcal{D})} \cdot \mathbb{1}\{\phi^{(n)}(y_0^n) = (\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2)\}. \quad (5.41)$$

Further, define $U_0 \triangleq (\tilde{M}_0, \tilde{Y}_0^{T-1}, T)$, $U_1 \triangleq \tilde{M}_1$, $U_2 \triangleq \tilde{M}_2$, $\tilde{Y}_i \triangleq \tilde{Y}_{i,T}$ (for $i \in \{0, 1, 2\}$), where T is uniform over $\{1, \dots, n\}$ and independent of all other random variables. Notice the Markov chain $(U_0, U_1, U_2) \rightarrow \tilde{Y}_0 \rightarrow (\tilde{Y}_1, \tilde{Y}_2)$. Then the following inequalities hold:

$$H(\tilde{M}_0) \geq nI(U_0; \tilde{Y}_0) + \log P_{Y_0^n}(\mathcal{D}), \quad (5.42)$$

$$H(\tilde{M}_i) \geq nI(U_i; \tilde{Y}_0 | U_0), \quad i \in \{1, 2\}. \quad (5.43)$$

Let $\eta > 0$ be arbitrary. For $i \in \{1, 2\}$, if

$$\Pr[\hat{\mathcal{H}}_i = 0 | \mathcal{H} = 0, Y_0^n = y_0^n] \geq \eta, \quad \forall y_0^n \in \mathcal{D}, \quad (5.44)$$

then

$$-\frac{1}{n} \log \beta_{i,n} \leq I(U_0 U_i; \tilde{Y}_i) + \phi_i(n), \quad (5.45)$$

where $\phi_i(n)$ is a function that tends to 0 as $n \rightarrow \infty$.

Proof: See Appendix C.2. ■

We now proceed to prove the converse to Theorem 1. Fix a positive $\eta > 0$. Denote for each blocklength n , the set of strongly typical sequences in \mathcal{Y}_0^n by $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0})$. Set $\mu_n = n^{-2/3}$ and define for $i \in \{1, 2\}$, the sets

$$\mathcal{B}_i(\eta) \triangleq \{y_0^n \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0}) : \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \geq \eta\}, \quad (5.46)$$

$$\mathcal{D}_{\{1,2\}}(\eta) \triangleq \mathcal{B}_1(\eta) \cap \mathcal{B}_2(\eta), \quad (5.47)$$

$$\mathcal{D}_{\{i\}}(\eta) \triangleq \mathcal{B}_i(\eta) \setminus \mathcal{D}_{\{1,2\}}(\eta). \quad (5.48)$$

Further define for each n the probabilities

$$\Delta_{\mathcal{I}} \triangleq P_{Y_0^n}(\mathcal{D}_{\mathcal{I}}(\eta)), \quad \mathcal{I} \in \mathcal{P}(2), \quad (5.49)$$

and notice that by the laws of probability

$$\Delta_{\{1,2\}} + \Delta_{\{i\}} = P_{Y_0^n}(\mathcal{B}_i(\eta)), \quad i \in \{1, 2\}, \quad (5.50)$$

$$\Delta_{\{1,2\}} \geq \max\{P_{Y_0^n}(\mathcal{B}_1(\eta)) + P_{Y_0^n}(\mathcal{B}_2(\eta)) - 1, 0\}. \quad (5.51)$$

By (5.7a), it can be shown that

$$1 - \epsilon_i \leq \eta(1 - P_{Y_0^n}(\mathcal{B}_i(\eta))) + P_{Y_0^n}(\mathcal{B}_i(\eta)) + P_{Y_0^n}(\overline{\mathcal{T}}_{\mu_n}^{(n)}). \quad (5.52)$$

Thus, by (5.52) and [44, Lemma 2.12]:

$$P_{Y_0^n}(\mathcal{B}_i(\eta)) \geq \frac{1 - \epsilon_i - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0|}{(1 - \eta)2\mu_n n}, \quad i \in \{1, 2\}, \quad (5.53)$$

and we conclude that in the limit $n \rightarrow \infty$ and $\eta \downarrow 0$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} (\Delta_{\{1,2\}} + \Delta_{\{i\}}) \geq 1 - \epsilon_i, \quad i \in \{1, 2\} \quad (5.54a)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\} \quad (5.54b)$$

$$\overline{\lim}_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} \leq 1. \quad (5.54c)$$

We proceed by applying Lemma 9 to the set $\mathcal{D}_{\mathcal{I}}$ for any $\mathcal{I} \in \mathcal{P}(2)$ with $\Delta_{\mathcal{I}} \geq 0$, and conclude that there

exists a tuple $(U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2})$ satisfying

$$H(\tilde{\mathbf{M}}_{\mathcal{I},0}) \geq nI(U_{\mathcal{I},0}; \tilde{Y}_{\mathcal{I},0}) + \log P_{Y^n}(\mathcal{D}_{\mathcal{I}}), \quad \mathcal{I} \in \mathcal{P}(2), \quad (5.55)$$

$$H(\tilde{\mathbf{M}}_{\mathcal{I},i}) \geq nI(U_{\mathcal{I},i}; \tilde{Y}_{\mathcal{I},0}|U_{\mathcal{I},0}), \quad i \in \{1, 2\}, \mathcal{I} \in \mathcal{P}(2), \quad (5.56)$$

and for $i \in \{1, 2\}$, $\mathcal{I} \in \{\{1, 2\}, \{i\}\}$:

$$-\frac{1}{n} \log \beta_{i,n} \leq I(U_{\mathcal{I},0}U_{\mathcal{I},i}; \tilde{Y}_{\mathcal{I},i}) + \phi_{\mathcal{I},i}(n), \quad (5.57)$$

where all for all $i \in \{1, 2\}$ and $\mathcal{I} \in \{\{1, 2\}, \{i\}\}$, the function $\phi_{\mathcal{I},i}(n) \downarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_{\mathcal{I},0}, \tilde{Y}_{\mathcal{I},i}, \tilde{\mathbf{M}}_{\mathcal{I},0}, \tilde{\mathbf{M}}_{\mathcal{I},i}$ are defined as in the lemma applied to the subset $\mathcal{D}_{\mathcal{I}}$.

To summarize:

$$-\frac{1}{n} \log \beta_{i,n} \leq \min\{I(U_{\{1,2\},0}U_{\{1,2\},i}; \tilde{Y}_{\{1,2\},i}), I(U_{\{i\},0}U_{\{i\},i}; \tilde{Y}_{\{i\},i})\} + \phi_i(n), \quad (5.58)$$

where $\phi_i(n)$ is a function tending to 0 as $n \rightarrow \infty$.

Define the following random variables for $i \in \{1, 2\}$ and $\mathcal{I} \in \mathcal{P}(2)$

$$\tilde{L}_{\mathcal{I},i} \triangleq \text{len}(\tilde{\mathbf{M}}_{\mathcal{I},i}). \quad (5.59)$$

By the rate constraints (5.3), and the definition of the random variables $\tilde{\mathbf{M}}_{\mathcal{I},i}$, we obtain by the total law of expectations

$$nR_0 \geq \mathbb{E}[L_0] \geq \sum_{\mathcal{I} \in \mathcal{P}(2)} \mathbb{E}[\tilde{L}_{\mathcal{I},0}] \Delta_{\mathcal{I}}. \quad (5.60)$$

Moreover,

$$H(\tilde{\mathbf{M}}_{\mathcal{I},0}) = H(\tilde{\mathbf{M}}_{\mathcal{I},0}, \tilde{L}_{\mathcal{I},0}) \quad (5.61)$$

$$= \sum_{l_{\mathcal{I}}} \Pr[\tilde{L}_{\mathcal{I},0} = l_{\mathcal{I}}] H(\tilde{\mathbf{M}}_{\mathcal{I},0} | \tilde{L}_{\mathcal{I},0} = l_{\mathcal{I}}) + H(\tilde{L}_{\mathcal{I},0}) \quad (5.62)$$

$$\leq \sum_{l_{\mathcal{I}}} \Pr[\tilde{L}_{\mathcal{I},0} = l_{\mathcal{I}}] l_{\mathcal{I}} + H(\tilde{L}_{\mathcal{I},0}) \quad (5.63)$$

$$= \mathbb{E}[\tilde{L}_{\mathcal{I},0}] + H(\tilde{L}_{\mathcal{I},0}), \quad (5.64)$$

which combined with (5.60) establishes

$$\sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} H(\tilde{\mathbf{M}}_{\mathcal{I},0}) \leq \sum_{\mathcal{I} \in \mathcal{P}(2)} \Delta_{\mathcal{I}} \mathbb{E}[\tilde{L}_{\mathcal{I},0}] + \Delta_{\mathcal{I}} H(\tilde{L}_{\mathcal{I},0}) \quad (5.65)$$

$$\leq nR_0 \left(1 + \sum_{\mathcal{I} \in \mathcal{P}(2)} h_b \left(\frac{\Delta_{\mathcal{I}}}{nR_0} \right) \right), \quad (5.66)$$

where (5.66) holds by (5.60) and because the entropy of a discrete and positive random variable $\tilde{L}_{\mathcal{I},0}$ of

mean $\mathbb{E}[\tilde{L}_{\mathcal{I},0}] \leq \frac{nR_0}{\Delta_{\mathcal{I}}}$ is bounded by $\frac{nR_0}{\Delta_{\mathcal{I}}} \cdot h_b\left(\frac{\Delta_{\mathcal{I}}}{nR_0}\right)$, see [23, Theorem 12.1.1].

In a similar way we obtain for $i \in \{1, 2\}$

$$\sum_{\mathcal{I} \in \{\{1,2\}, \{i\}\}} \Delta_{\mathcal{I}} H(\tilde{M}_{\mathcal{I},i}) \leq nR_i \left(1 + \sum_{\mathcal{I} \in \{\{1,2\}, \{i\}\}} h_b\left(\frac{\Delta_{\mathcal{I}}}{nR_i}\right) \right). \quad (5.67)$$

Notice that when $\Delta_{\mathcal{I}} = 0$, the trivial choice $U_{\mathcal{I},i} = \tilde{Y}_{\mathcal{I},i}$ satisfies the inequalities (5.58), (5.66), and (5.67). Therefore, above conclusions hold for $(U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2})$ for any $\mathcal{I} \in \mathcal{P}(2)$.

Combining (5.66) and (5.67) with (5.55) and (5.56), noting (5.50) and (5.53), and considering also (5.58), we have proved so far that for all $n \geq 1$ there exist joint pmfs $P_{U_{\mathcal{I},0}U_{\mathcal{I},1}U_{\mathcal{I},2}\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}} = P_{\tilde{Y}_{\mathcal{I},0}} P_{\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}|\tilde{Y}_{\mathcal{I},0}} P_{U_{\mathcal{I},0}U_{\mathcal{I},1}U_{\mathcal{I},2}|\tilde{Y}_{\mathcal{I},0}}$ (abbreviated as $P_{\mathcal{I}}^{(n)}$) for $\mathcal{I} \in \mathcal{P}(2)$ so that the following conditions hold for $i \in \{1, 2\}$ (where I_P indicates that the mutual information should be calculated according to a pmf P):

$$R_0 \geq \sum_{\mathcal{I} \in \mathcal{P}(2)} (I_{P_{\mathcal{I}}^{(n)}}(U_{\mathcal{I},0}; \tilde{Y}_{\mathcal{I},0}) + g_{\mathcal{I},1}(n)) \cdot g_{\mathcal{I},2}(n, \eta), \quad (5.68a)$$

$$R_i \geq \sum_{\mathcal{I} \in \{\{1,2\}, \{i\}\}} (I_{P_{\mathcal{I}}^{(n)}}(U_{\mathcal{I},i}; \tilde{Y}_{\mathcal{I},0}|U_{\mathcal{I},0})) \cdot g_{\mathcal{I},2}(n, \eta), \quad (5.68b)$$

$$\theta_i \leq \min\{I_{P_{\{1,2\}}^{(n)}}(U_{\{1,2\},0}U_{\{1,2\},i}; \tilde{Y}_{\{1,2\},i}), I_{P_{\{i\}}^{(n)}}(U_{\{i\},0}U_{\{i\},i}; \tilde{Y}_{\{i\},i})\} + g_{\{i\},3}(n), \quad (5.68c)$$

for some nonnegative functions $g_{\mathcal{I},1}(n), g_{\mathcal{I},2}(n, \eta), g_{\{i\},3}(n)$ with the following asymptotic behaviors:

$$\lim_{n \rightarrow \infty} g_{\mathcal{I},1}(n) = 0, \quad \forall \mathcal{I} \in \mathcal{P}(2), \quad (5.69)$$

$$\lim_{n \rightarrow \infty} g_{\{i\},3}(n) = 0, \quad \forall i \in \{1, 2\}, \quad (5.70)$$

$$\lim_{n \rightarrow \infty} (g_{\{1,2\},2}(n, \eta) + g_{\{i\},2}(n, \eta)) \geq \max\left\{\frac{1 - \epsilon_i - \eta}{1 - \eta}, 0\right\}, \quad \forall i \in \{1, 2\}. \quad (5.71)$$

By Carathéodory's theorem [48, Appendix C], there exist for each n , random variables $U_{\{1,2\},0}, U_{\{1\},0}, U_{\{2\},0}, U_{\{1,2\},1}, U_{\{1\},1}, U_{\{1,2\},2}, U_{\{2\},2}$ satisfying (5.68) over alphabets of sizes

$$|\mathcal{U}_{\{1,2\},0}| \leq |\mathcal{Y}_0| + 3, \quad (5.72)$$

$$|\mathcal{U}_{\{i\},0}| \leq |\mathcal{Y}_0| + 2, \quad i \in \{1, 2\}, \quad (5.73)$$

$$|\mathcal{U}_{\mathcal{I},i}| \leq |\mathcal{U}_{\mathcal{I},0}| \cdot |\mathcal{Y}_0| + 1, \quad i \in \{1, 2\}, \mathcal{I} \in \{\{1, 2\}, \{i\}\}. \quad (5.74)$$

Then we invoke the Bolzano-Weierstrass theorem and consider for each $\mathcal{I} \in \mathcal{P}(2)$ a sub-sequence $P_{U_{\mathcal{I},0}U_{\mathcal{I},1}U_{\mathcal{I},2}\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}}^{(n_k)}$ that converges to a limiting pmf $P_{U_{\mathcal{I},0}U_{\mathcal{I},1}U_{\mathcal{I},2}Y_{\mathcal{I},0}Y_{\mathcal{I},1}Y_{\mathcal{I},2}}^*$. For these limiting pmfs, which we abbreviate by $P_{\mathcal{I}}^*$, we conclude by (5.68a)–(5.68c) and (5.54) that for all $i \in \{1, 2\}$:

$$R_0 \geq \sigma_{\{1,2\}} \cdot I_{P_{\{1,2\}}^*}(U_{\{1,2\},0}; Y_{\{1,2\},0}) + \sigma_{\{1\}} \cdot I_{P_{\{1\}}^*}(U_{\{1\},0}; Y_{\{1\},0}) + \sigma_{\{2\}} \cdot I_{P_{\{2\}}^*}(U_{\{2\},0}; Y_{\{2\},0}), \quad (5.75)$$

$$R_i \geq \sigma_{\{1,2\}} \cdot I_{P_{\{1,2\}}^*}(U_{\{1,2\},i}; Y_{\{1,2\},0} | U_{\{1,2\},0}) + \sigma_{\{i\}} \cdot I_{P_{\{i\}}^*}(U_{\{i\},i}; Y_{\{i\},0} | U_{\{i\},0}), \quad (5.76)$$

$$\theta_i \leq \min\{I_{P_{\{1,2\}}^*}(U_{\{1,2\},0} U_{\{1,2\},i}; Y_{\{1,2\},0}), I_{P_{\{i\}}^*}(U_{\{i\},0} U_{\{i\},i}; Y_{\{i\},i})\}, \quad (5.77)$$

where numbers $\sigma_{\{1,2\}}, \sigma_{\{1\}}, \sigma_{\{2\}} > 0$ satisfy $\sigma_{\{1,2\}} + \sigma_{\{1\}} + \sigma_{\{2\}} \leq 1$ and

$$\sigma_{\{1,2\}} + \sigma_{\{i\}} \geq 1 - \epsilon_i, \quad i \in \{1, 2\}, \quad (5.78a)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}. \quad (5.78b)$$

Notice further that since for any $\mathcal{I} \in \mathcal{P}(2)$ and any k , the sequence $\tilde{Y}_{\mathcal{I},0}^{n_k}$ lies in the typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0})$, we have for all $\mathcal{I} \in \mathcal{P}(2)$, $|P_{\tilde{Y}_{\mathcal{I},0}} - P_{Y_0}| \leq \mu_{n_k}$ and thus the limiting pmf satisfies $P_{\tilde{Y}_{\mathcal{I},0}}^* = P_{Y_0}$. Moreover, since for each n_k the pair of random variables $(\tilde{Y}_{\mathcal{I},1}, \tilde{Y}_{\mathcal{I},2})$ is drawn according to $P_{Y_1 Y_2 | Y_0}$ given $\tilde{Y}_{\mathcal{I},0}$, the limiting pmf also satisfies $P_{Y_{\mathcal{I},1} Y_{\mathcal{I},2} | Y_{\mathcal{I},0}}^* = P_{Y_1 Y_2 | Y_0}$. We also notice for all $\mathcal{I} \in \mathcal{P}(2)$ that under $P_{\mathcal{I}}^*$ the Markov chain $(U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2}) \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ holds. This concludes the converse proof.

5.8 Summary and Discussion

In this chapter, another network architecture for multiple DCs was studied where a single sensor communicates with two DCs. Communication is assumed to be over two types of noiseless links: common BC and individual links. Due to the competition between both DCs on the communicated information from the sensor over the common BC, the available resources at the sensor have to be distributed between the two DCs and a tradeoff arises between their error exponents even under maximum-rate constraints or for equal type-I error thresholds. This tradeoff is different from the tradeoff that arises due to the rate-sharing strategy that we proposed in the previous chapter. However, both types of tradeoff are shown to be present for the setup when expected-rate constraints are considered with different type-I error thresholds at the two DCs.

The optimal coding scheme to achieve the fundamental exponents region for the above setup is also based on the rate-sharing strategy as in the previous chapter, where here we multiplex four subschemes using one degenerate scheme and three different versions of the optimal coding scheme under maximum-rate constraints for vanishing type-I error probabilities. The probabilities of these subschemes are subject to some optimization. In the special case of having only the common BC, we showed that the characterization of the fundamental exponents region simplifies and thus the optimal coding scheme simplifies too; it suffices to use only three subschemes with already given probabilities. We conjecture that similar simplifications are possible in the general case. The general results (without simplification) are also extendable to the

setup with K DCs, one common BC, and K individual links.

Chapter 6

Summary and Outlook

6.1 Summary

In this thesis we focused on deriving the fundamental limits for distributed hypothesis testing against independence for networks with multiple sensors and/or multiple decision centers. In addition to establishing new optimality results, we highlight the following conclusions:

- We extended the rate-gain factor of $(1 - \epsilon)^{-1}$ to all communication links in multiple-sensors single-DC networks, where there is a single terminal that can communicate directly or through multiple hops to all other terminals in the network.
- The same conclusion holds for networks with multiple DCs that have identical type-I error thresholds equal to ϵ .
- For networks with multiple DCs that have different type-I error thresholds, we can obtain different rate-boosts on the different communication links. In fact, the different type-I error thresholds lead to different margins for the multiple DCs to exploit in order to increase their type-II error exponents. We thus observe and characterize a tradeoff in the fundamental exponents region between the different DCs. This tradeoff is *the first of its kind*. We propose a new rate-sharing and multiplexing strategy to achieve this tradeoff.
- For some networks with K DCs, we can show that it suffices to multiplex only $K + 1$ schemes, with one scheme being the degenerate scheme of sending single-bit messages and raising alert at all DCs, instead of 2^K schemes as in the general case. We can further identify the multiplexing probabilities

as a function of the given type-I error thresholds. This simplifies the optimal coding schemes and the expressions of the characterizations of the fundamental exponents regions for these networks.

- For BC networks with multiple DCs, a tradeoff was already observed as the different DCs are competing for the shared information over the BC. Our new tradeoff adds up on this tradeoff when the DCs have different type-I error thresholds under expected-rate constraints.
- To prove our converse results, we propose new proof methods combining change of measure arguments with blowing-up lemma and/or our new tool “asymptotic Markov chains”. We observe that in some cases, this new tool suffices and the blowing-up lemma is not required. In addition, our method can be used to prove converse results for other problems that are not necessarily related to hypothesis testing.
- Our converse proof methods also allow one to prove new strong converse results under maximum-rate constraints, or to simplify existing proofs.

6.2 Outlook on Interactive Communication

The single-sensor single-DC setup was studied under interactive communication and maximum-rate constraints in [7,8,10]. In these works, an optimal error exponent is derived when testing against independence under vanishing type-I error probability. As we explained previously, under expected-rate constraints, a $(1 - \epsilon)^{-1}$ rate-gain can be obtained on all links when a type-I error probability of ϵ is tolerated. While this gain is achievable, with our converse proof methods we could not establish the corresponding converse. The problem here is that we could neither use the blowing-up lemma, after applying the change of measure arguments, nor prove the desired asymptotic Markov chains.

6.3 Outlook on Non-Cooperative MAC

The non-cooperative MAC setup remains open even for the special case of testing against independence under maximum-rate constraints. Many works obtained partial results in [2,4–6,9,18,24,26,27]. Our interest is in the maximum gain achievable under expected-rate constraints compared to results under maximum-rate constraints. Using our converse proof methods, we can establish a converse result showing that no better gain than $(1 - \epsilon)^{-1}$ factor could be achieved. However, as we explained before, in order to achieve this rate-gain on all links, we need at least one terminal that communicates with all other terminals in the

network informing them when to apply or not the degenerate scheme, which is not the case in this setup. We thus conjecture for the setup with two sensor-DC links, that the best achievable rate-gain is $(1 - \epsilon_1)^{-1}$ and $(1 - \epsilon_2)^{-1}$ where ϵ_1 and ϵ_2 correspond to the probabilities of two arbitrary subsets of the observations at the first and the second sensor, respectively, such that their joint probability is less than or equal to ϵ . The question would be then, “what are the highest individual probabilities that can be obtained by each set and the impact of the tradeoff between them?”. A new converse proof method is needed to prove this is the best possible gain.

6.4 Extensions to Noisy Communication Channel

The studied setups in this manuscript are of main interest to establish their fundamental limits and understand the gains from expected-rate constraints. However, it is important to build up on them for more practical scenarios where the communication between the different nodes in the system could be over noisy channels. Salehkalaibar and Wigger [20] had proved such an extension for the point-to-point setup over noisy DMC under an expected-rate constraint. The multi-sensor setups studied in this thesis have previously been considered with noisy communication channels but only under maximum-rate constraints [18, 26, 29].

6.5 Extensions to General Hypothesis Testing

The special case of testing against independence in a binary hypothesis testing environment is of significant relevance in distributed control and anomaly detection systems, where observations are correlated in normal situation, but become uncorrelated in case of incident or system failure. However, in other decision systems, used for instance for binary classification, observations are not independent under the alternative hypothesis but follow some other arbitrary distribution. This problem of “general distributed hypothesis testing” seems much more challenging, and as we have mentioned previously is even open in the single-sensor single-DC setup under the standard maximum-rate constraint. Another interesting direction for future research is the study of M -ary hypothesis testing problems, which is widely open except for zero-rate communication. We conjecture that under expected-rate constraints, the proposed multiplexing and rate-sharing strategies can achieve similar gains for general binary hypothesis testing and for M -ary hypothesis testing as we proved for binary hypothesis testing in the different communication scenarios. However it would be very challenging to prove optimality of such strategy under expected-rate constraints without having the optimality results for the analogous setups under maximum-rate constraints.

6.6 Setups with Privacy and Security Constraints

In this thesis, our focus was on maximizing the error exponent at a given DC. In decision systems with sensitive data observed at the sensors, another goal is to protect the privacy of this data from the legitimate DC which is expected to learn only the joint distribution of the data at other terminals but not their private sensitive information. Several works exist on the distributed hypothesis testing problem with privacy constraints and under maximum-rate constraints (for instance, see [54, 55] and references therein). An interesting perspective would be to investigate this problem when expected-rate constraints are considered instead of maximum-rate constraints. One would like to check if the same gains could be achieved as without the privacy constraints or if any new type of tradeoff could arise. Another aspect of the problem under security constraints is studied when an eavesdropper or an adversary attacker try to learn the hypothesis test carried out by a given DC as in [56]. In this case, several security and privacy metrics can be studied and one main concern would be to determine the maximum achievable error exponent at the DC when the error exponent at the attacker should be minimized under expected-rate constraints on the communication links.

6.7 Finite-Blocklength and Machine Learning

Establishing the fundamental limits for the problem of distributed hypothesis testing in the asymptotic regime (as the blocklength $n \rightarrow \infty$), as we do in this thesis for various setups, is essential to understand the impact of different constraints on the optimal performance of the decision centers in a given setup. However when it comes to applications, the sensors in some IoT networks are expected to have limited number of samples, and thus the optimal strategy in the infinite-blocklength regime might not necessarily remain optimal in the finite-blocklength regime. Establishing the fundamental limits for the problem of hypothesis testing in the finite-blocklength regime is very challenging as an exact characterization of the optimal error exponent even in the centralized or simplest point-to-point setup remains an open problem for all kinds of hypothesis testing.

In general, two types of problems have to be solved in distributed hypothesis testing problems: compress the information at the sensors and take a decision at the DCs. To achieve desired compression rates, one could use existing quantization algorithms which might be limited for small rates (in scalar quantization), or computationally infeasible for non-short blocklengths (very long running time in vector quantization). An alternative approach would be to explore deep learning algorithms such as Autoencoders to perform

such compression. Then for the decision rules, typicality checks are shown to be optimal in the infinite-blocklength regime, which we can substitute with correlation tests between the received message(s) at the DC and its local observations in finite-blocklength regime. A simple test would rely on a data-driven correlation factor under null hypothesis, and whenever the testing correlation factor differs by an arbitrary margin from the data-driven one, an alert would be raised. But what if deep neural networks could learn better testing criteria or correlation factor? Intuitively, replacing the whole coding scheme with a neural network responsible for quantization, compression, and decision-rule making, seems to be a promising research topic.

Other interesting research directions that could be analyzed with machine learning tools arise in sequential hypothesis testing problems or in multi-hop multi-DC networks. For example, in sequential hypothesis testing problems, machine learning could be used to determine at which number of observations a sensor should stop sampling and start communicating with the sensor. In multi-hop systems, machine learning could help in identifying in the finite-blocklength regime after how many hops to take the decision so as to save communication bandwidth for other applications. For example, if a certain lower bound on the fundamental exponents region is required and is enough, then “what is the required number of samples to stop sampling and take a final decision?” Or in the case of a multi-hop setup, “at which relay testing should stop and the final decision should be forwarded to the receiver?”, so that the available rate at the next communication links would be saved for other applications sharing the same bandwidth.

Appendix A

Proofs for Chapters 2 and 3

A.1 Proof of Lemma 1: Concavity and Monotonicity of the Function η_1

The function $\eta_1(R)$ is monotonically non-decreasing because larger values of R imply larger optimization domains. Continuity follows simply by the continuity of mutual information.

The concavity of $\eta_1(R)$ follows by the following arguments. Consider rates R and \tilde{R} , and let U^* and \tilde{U}^* be the corresponding solutions to the optimizations in the definition of η_1 . Pick any $\lambda \in [0, 1]$, define $Q \sim \text{Bern}(\lambda)$ independent of $(Y_0, Y_1, U^*, \tilde{U}^*)$, and set

$$U_Q^* = \begin{cases} U^* & \text{if } Q = 0 \\ \tilde{U}^* & \text{if } Q = 1. \end{cases} \quad (\text{A.1})$$

Defining the random variable $V := (U_Q^*, Q)$, we obtain

$$\lambda \cdot \eta_1(R) + (1 - \lambda) \cdot \eta_1(\tilde{R}) = \lambda I(U^*; Y_1) + (1 - \lambda) I(\tilde{U}^*; Y_1) \quad (\text{A.2})$$

$$= I(U_Q^*; Y_1 | Q) \quad (\text{A.3})$$

$$= I(U_Q^*, Q; Y_1) \quad (\text{A.4})$$

$$= I(V; Y_1) \quad (\text{A.5})$$

$$\leq \eta_1(I(V; Y_0)) \quad (\text{A.6})$$

$$\leq \eta_1(\lambda R + (1 - \lambda)\tilde{R}) \quad (\text{A.7})$$

where (A.4) holds because Q is independent of Y_1 , (A.6) holds by the definition of the function η_1 , and (A.7) holds by the monotonicity of the function η_1 and the following set of (in)equalities:

$$I(V; Y_0) = I(U_Q^*, Q; Y_0) = I(U_Q^*; Y_0 | Q) \quad (\text{A.8})$$

$$= \lambda I(U^*; Y_0) + (1 - \lambda) I(\tilde{U}^*; Y_0) \quad (\text{A.9})$$

$$\leq \lambda R + (1 - \lambda) \tilde{R}. \quad (\text{A.10})$$

■

A.2 Converse to Theorem 3

In this proof, we replace the notation of serif font for the messages M_i used in [25] by the sans-serif font \mathbf{M}_i (for $i = 1, 2$) and the notation of X_1, X_2 in [25] by Y_0, Y_1 .

Following the weak converse in [25], we can obtain the upper bound for the error exponent as

$$\lim_{\epsilon \downarrow 0} \theta_{\max}(R_1, R_2, \epsilon) \leq \lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{M}_1 \mathbf{M}_2; Y_2^n) \quad (\text{A.11})$$

Moreover, we can lower bound the rate constraints as

$$nR_1 \geq H(\mathbf{M}_1) \geq I(\mathbf{M}_1; Y_0^n Y_1^n) \quad (\text{A.12})$$

$$= \sum_{i=1}^n I(\mathbf{M}_1 Y_0^{i-1} Y_1^{i-1}; Y_{0,i} Y_{1,i}) \quad (\text{A.13})$$

$$\geq \sum_{i=1}^n I(U_{1,i}; Y_{0,i}), \quad (\text{A.14})$$

where $U_{1,i} = (\mathbf{M}_1, Y_0^{i-1}, Y_1^{i-1})$. We note that the Markov chain $U_{1,i} \rightarrow Y_{0,i} \rightarrow (Y_{1,i}, Y_{2,i})$ holds. Similarly,

$$nR_2 \geq H(\mathbf{M}_2) \geq I(\mathbf{M}_2; Y_0^n Y_1^n | \mathbf{M}_1) \quad (\text{A.15})$$

$$= \sum_{i=1}^n I(\mathbf{M}_2; Y_{0,i} Y_{1,i} | \mathbf{M}_1 Y_0^{i-1} Y_1^{i-1}) \quad (\text{A.16})$$

$$\geq \sum_{i=1}^n I(U_{2,i}; Y_{1,i} | U_{1,i}), \quad (\text{A.17})$$

where $U_{2,i} = \mathbf{M}_2$ and we also note that the Markov chain $U_{2,i} \rightarrow (Y_{1,i}, U_{1,i}) \rightarrow (Y_{0,i}, Y_{2,i})$ holds.

Then we single-letterize the upper bound of the error exponent as follows:

$$I(\mathbf{M}_1 \mathbf{M}_2; Y_2^n) = \sum_{i=1}^n I(\mathbf{M}_1 \mathbf{M}_2 Y_2^{i-1}; Y_{2,i}) \quad (\text{A.18})$$

$$\leq \sum_{i=1}^n I(\mathbf{M}_1 \mathbf{M}_2 Y_0^{i-1} Y_1^{i-1} Y_2^{i-1}; Y_{2,i}) \quad (\text{A.19})$$

$$= \sum_{i=1}^n I(\mathbf{M}_1 \mathbf{M}_2 Y_0^{i-1} Y_1^{i-1}; Y_{2,i}) \quad (\text{A.20})$$

$$= \sum_{i=1}^n I(U_{1,i} U_{2,i}; Y_{2,i}), \quad (\text{A.21})$$

where (A.20) holds by the Markov chain $Y_2^{i-1} \rightarrow (M_1, M_2, Y_0^{i-1}, Y_1^{i-1}) \rightarrow Y_{2,i}$. Our proof is then concluded by defining the time-sharing random variable Q uniform over $\{1, \dots, n\}$ independent of all other random variables, and by defining $U_1 \triangleq (U_{1Q}, Q)$, $U_2 \triangleq (U_{2Q}, Q)$, $Y_0 \triangleq Y_{0Q}$, $Y_1 \triangleq Y_{1Q}$, and $Y_2 \triangleq Y_{2Q}$.

A.3 Strong Converse Proof to Theorem 4

Notice first that the Markov chain $U_2 \rightarrow (U_1, Y_1) \rightarrow (Y_0, Y_2)$ needed for the validation of the distribution of the joint pmf $P_{U_1 U_2 Y_0 Y_1 Y_2} = P_{U_1 | Y_0} P_{U_2 | U_1 Y_1} P_{Y_0 Y_1 Y_2}$, can be replaced by the weaker Markov chain $U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2$ because the right hand-side of (3.23) does not depend on the joint pmf of U_2 and Y_0 .

We proceed to show that

$$\theta \leq \tilde{\eta}(R_1, R_2) \quad (\text{A.22})$$

where $\tilde{\eta}(\cdot, \cdot)$ is defined as follows

$$\tilde{\eta}(R_1, R_2) := \max_{\substack{P_{U_1 | Y_0}, P_{U_2 | U_1 Y_1} : \\ R_1 \geq I(U_1; Y_0) \\ R_2 \geq I(U_2; Y_1 | U_1)}} I(U_1 U_2; Y_2) \quad (\text{A.23})$$

with the auxiliary random variables U_1 and U_2 satisfying the Markov chains $U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ and $U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2$.

Consider a sequence (in n) of encoding and decision functions $\{(\phi_1^{(n)}, \phi_2^{(n)}, g^{(n)})\}$ satisfying the constraints on the rates and error probabilities in (3.10) for maximum-rate constraints setup.

In the following, we use Lemma 2 which is presented in the main text, specifically in the converse proof of Theorem 5 in Section 3.4, in order to prove the desired outer bound on the exponents region.

Fix a positive $\eta > 0$. Set $\mu_n = n^{-1/3}$, and define the sets

$$\mathcal{B}(\eta) \triangleq \{(y_0^n, y_1^n) : \Pr[\hat{\mathcal{H}}_1 = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \geq \eta\}, \quad (\text{A.24})$$

$$\mathcal{D}(\eta) \triangleq \mathcal{B}(\eta) \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}). \quad (\text{A.25})$$

Further define for each n the probability

$$\Delta \triangleq P_{Y_0^n Y_1^n}(\mathcal{D}(\eta)), \quad (\text{A.26})$$

and notice that by the constraint (3.10a) on the type-I error probability:

$$1 - \epsilon \leq \sum_{y_0^n, y_1^n} \Pr[\hat{\mathcal{H}} = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (\text{A.27})$$

$$\leq \sum_{(y_0^n, y_1^n) \in \mathcal{B}(\eta)} P_{Y_0^n Y_1^n}(y_0^n, y_1^n) + \sum_{(y_0^n, y_1^n) \notin \mathcal{B}(\eta)} \Pr[\hat{\mathcal{H}} = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (\text{A.28})$$

$$\leq P_{Y_0^n Y_1^n}(\mathcal{B}(\eta)) + \eta(1 - P_{Y_0^n Y_1^n}(\mathcal{B}(\eta))). \quad (\text{A.29})$$

Thus we have

$$P_{Y_0^n Y_1^n}(\mathcal{B}(\eta)) \geq \frac{1 - \epsilon - \eta}{1 - \eta}. \quad (\text{A.30})$$

Moreover, by [44, Remark to Lemma 2.12], the probability that the pair (Y_0^n, Y_1^n) lies in the strongly jointly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})$ satisfies

$$P_{Y_0 Y_1}^n \left(\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) \right) \geq 1 - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{4\mu_n^2 n}, \quad (\text{A.31})$$

and since for any two events A and B ,

$$\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1, \quad (\text{A.32})$$

then by (A.25), (A.26), (A.30), and (A.31), we obtain

$$\Delta \geq \frac{1 - \epsilon - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{4\mu_n^2 n}. \quad (\text{A.33})$$

We proceed by applying Lemma 2 to the set $\mathcal{D}(\eta)$ with $\Delta \geq \eta$. This allows to conclude that there exists a pair (U_1, U_2) satisfying the Markov chain $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$ and the (in)equalities

$$nR_1 \geq H(\tilde{\mathbf{M}}_1) \geq nI(U_1; \tilde{Y}_0) + \log P_{Y_0^n Y_1^n}(\mathcal{D}), \quad (\text{A.34})$$

$$nR_2 \geq H(\tilde{\mathbf{M}}_2) \geq nI(U_2; \tilde{Y}_1|U_1), \quad (\text{A.35})$$

$$\phi_1(n) = I(U_1; \tilde{Y}_1 \tilde{Y}_2 | \tilde{Y}_0), \quad (\text{A.36})$$

and

$$-\frac{1}{n} \log \beta_n \leq I(U_1 U_2; \tilde{Y}_2) + \phi_2(n), \quad (\text{A.37})$$

where the functions $\phi_1(n), \phi_2(n) \rightarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_0, \tilde{Y}_1, \tilde{Y}_2, \tilde{\mathbf{M}}_1, \tilde{\mathbf{M}}_2$ are defined as in the lemma, when applied to the set $\mathcal{D}(\eta)$.

To simplify exposition, we assume η very small and $\Delta \geq \eta$. Otherwise the proof is similar but omitted here.

For all $n \geq 1$, we abbreviate the joint pmf $P_{U_1 U_2 \tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2} = P_{U_1 | \tilde{Y}_0 \tilde{Y}_1} P_{U_2 | U_1 \tilde{Y}_1} P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2}$ as $P^{(n)}$ and observe that by Carathéodory's theorem [48, Appendix C] for each n there must exist random variables U_1, U_2 satisfying (A.34)–(A.37) over alphabets of sizes

$$|\mathcal{U}_1| \leq |\mathcal{Y}_0| \cdot |\mathcal{Y}_1| + 3, \quad (\text{A.38})$$

$$|\mathcal{U}_2| \leq |\mathcal{U}_1| \cdot |\mathcal{Y}_1| + 1. \quad (\text{A.39})$$

Then we invoke the Bolzano-Weierstrass theorem and consider an increasing subsequence of positive num-

bers $\{n_k\}_{k=1}^{\infty}$ such that the following subsequences converge:

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2 U_1 U_2}^{(n_k)} = P_{Y_0 Y_1 Y_2 U_1 U_2}^*. \quad (\text{A.40})$$

Considering further an appropriate sequence of diminishing η -values, we conclude that:

$$R_1 \geq I_{P^*}(U_1; Y_0), \quad (\text{A.41})$$

$$R_2 \geq I_{P^*}(U_2; Y_1 | U_1), \quad (\text{A.42})$$

$$\theta \leq I_{P^*}(U_1 U_2; Y_2), \quad (\text{A.43})$$

where I_{P^*} indicates that the mutual information should be calculated according to the pmf $P_{Y_0 Y_1 Y_2 U_1 U_2}^*$. Notice further that since for any k the pair $(\tilde{Y}_0^{(n_k)}, \tilde{Y}_1^{(n_k)})$ lies in the jointly typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0 Y_1})$, we have $|P_{\tilde{Y}_0 \tilde{Y}_1} - P_{Y_0 Y_1}| \leq \mu_{n_k}$ and thus the limiting pmf satisfies $P_{\tilde{Y}_0 \tilde{Y}_1}^* = P_{Y_0 Y_1}$. Moreover, since for each n_k the random variable \tilde{Y}_2 is drawn according to $P_{Y_2 | Y_1}$ given \tilde{Y}_1 , irrespective of \tilde{Y}_0 , the limiting pmf also satisfies $P_{\tilde{Y}_2 | Y_0 Y_1}^* = P_{Y_2 | Y_1}$. We also notice that under $P_{Y_0 Y_1 Y_2 U_1 U_2}^*$ the Markov chain

$$U_2 \rightarrow (U_1, Y_1) \rightarrow Y_2, \quad (\text{A.44})$$

holds because $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$ forms a Markov chain for any n_k . Finally, by continuity considerations and by (A.36), the following Markov chain must hold under $P_{Y_0 Y_1 Y_2 U_1 U_2}^*$:

$$U_1 \rightarrow Y_0 \rightarrow (Y_1, Y_2). \quad (\text{A.45})$$

Using the definition of the function $\tilde{\eta}(\cdot, \cdot)$ in (A.23), we thus proved that all achievable error exponents θ are upper-bounded by the error exponent given in (A.22). This concludes our converse proof.

A.4 Analysis of the coding scheme in Subsection 3.3.3.1

Consider the cooperative two-sensor single-DC scheme employed when $Y_0^n \in \mathcal{D}_{\{2\}}$, and let $\hat{\mathcal{H}}_{\{2\}}$ denote the guess produced at R_2 when employing this scheme for any $Y_0^n \in \mathcal{Y}_0^n$. Notice that by assumption the type-I error probability of this scheme tends to 0 as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{2\}} = 1 | \mathcal{H} = 0] = 0. \quad (\text{A.46})$$

Noticing that when $Y_0^n \in \mathcal{D}_\emptyset$, then $\hat{\mathcal{H}} = 1$, and applying the total law of probability, we can have for the type-I error probability of the overall scheme

$$\alpha_n = \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0] \quad (\text{A.47})$$

$$= \Pr[\hat{\mathcal{H}} = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}} = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \quad (\text{A.48})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\}} = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \quad (\text{A.49})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\}} = 1 | \mathcal{H} = 0] \quad (\text{A.50})$$

Combining these inequalities with (A.46), and because in the limit $n \rightarrow \infty$ Inequality (3.25) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \leq \epsilon. \quad (\text{A.51})$$

For the type-II error probability of the overall scheme, we have

$$\beta_n = \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1] \quad (\text{A.52})$$

$$= \Pr[\hat{\mathcal{H}} = 0, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}} = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] \quad (\text{A.53})$$

$$= \Pr[\hat{\mathcal{H}} = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] \quad (\text{A.54})$$

$$= \Pr[\hat{\mathcal{H}}_{\{2\}} = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] \quad (\text{A.55})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{2\}} = 0 | \mathcal{H} = 1]. \quad (\text{A.56})$$

The type-II error exponents of the overall scheme are thus given by the error exponents of the cooperative two-sensor single-DC scheme employed under $Y_0^n \in \mathcal{D}_{\{2\}}$. By [25], for vanishing type-I error probability and under the rate-constraints $R_{\{2\},1}$ and $R_{\{2\},2}$, the exponents in (3.30) are proved achievable.

A.5 Proof of Lemma 2

Note first that by (3.41):

$$D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0 Y_1}^n) \leq \log \Delta_n^{-1}, \quad (\text{A.57})$$

where we define $\Delta_n \triangleq P_{Y_0^n Y_1^n}(\mathcal{D})$.

Further define

$$\tilde{U}_{2,t} \triangleq \tilde{M}_2 \quad (\text{A.58})$$

$$\tilde{U}_{1,t} \triangleq (\tilde{M}_1, \tilde{Y}_0^{t-1}, \tilde{Y}_1^{t-1}), \quad (\text{A.59})$$

and notice that

$$H(\tilde{M}_1) \geq I(\tilde{M}_1; \tilde{Y}_0^n \tilde{Y}_1^n) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0^n Y_1^n}^n) + \log \Delta_n \quad (\text{A.60})$$

$$= H(\tilde{Y}_0^n \tilde{Y}_1^n) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0^n Y_1^n}^n) - H(\tilde{Y}_0^n \tilde{Y}_1^n | \tilde{M}_1) + \log \Delta_n \quad (\text{A.61})$$

$$\geq n[H(\tilde{Y}_{0,T}\tilde{Y}_{1,T}) + D(P_{\tilde{Y}_{0,T}\tilde{Y}_{1,T}}||P_{Y_0Y_1})] - \sum_{t=1}^n H(\tilde{Y}_{0,t}\tilde{Y}_{1,t}|\tilde{U}_{1,t}) + \log \Delta_n \quad (\text{A.62})$$

$$= n[H(\tilde{Y}_{0,T}\tilde{Y}_{1,T}) + D(P_{\tilde{Y}_{0,T}\tilde{Y}_{1,T}}||P_{Y_0Y_1}) - H(\tilde{Y}_{0,T}\tilde{Y}_{1,T}|\tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{A.63})$$

$$\geq n[H(\tilde{Y}_{0,T}\tilde{Y}_{1,T}) - H(\tilde{Y}_{0,T}\tilde{Y}_{1,T}|\tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{A.64})$$

$$= n[I(\tilde{Y}_0\tilde{Y}_1; U_1)] + \log \Delta_n \quad (\text{A.65})$$

$$\geq n \left[I(\tilde{Y}_0; U_1) + \frac{1}{n} \log \Delta_n \right]. \quad (\text{A.66})$$

Here, (A.60) holds by (A.57); (A.62) holds by the super-additivity property in [36, Proposition 1], by the chain rule, and by the definition of $\tilde{U}_{1,t}$ and by defining T uniform over $\{1, \dots, n\}$ independent of all other random variables; (A.65) holds by the non-negativity of the Kullback-Leibler divergence; and (A.65) by the definitions of $U_1, \tilde{Y}_0, \tilde{Y}_1$ in the lemma.

We can lower bound the entropy of \tilde{M}_2 as follows

$$H(\tilde{M}_2) \geq I(\tilde{M}_2; \tilde{Y}_0^n \tilde{Y}_1^n | \tilde{M}_1) \quad (\text{A.67})$$

$$= \sum_{t=1}^n I(\tilde{U}_{2,t}; \tilde{Y}_{0,t} \tilde{Y}_{1,t} | \tilde{U}_{1,t}) \quad (\text{A.68})$$

$$= nI(\tilde{U}_{2,T}T; \tilde{Y}_{0,T} \tilde{Y}_{1,T} | \tilde{U}_{1,T}T) \quad (\text{A.69})$$

$$\geq nI(U_2; \tilde{Y}_1 | U_1). \quad (\text{A.70})$$

Here, (A.67) holds since \tilde{M}_2 is function of \tilde{Y}_1^n and \tilde{M}_1 ; (A.68) holds by the chain rule, the definitions of $\tilde{U}_{1,t}$ and $\tilde{U}_{2,t}$; and (A.70) holds by the definitions of U_2, U_1, \tilde{Y}_1 in the lemma.

We next upper bound the error exponent at the receiver. To this end, fix a sequence of real numbers $\{\ell_n\}_{n=1}^\infty$ satisfying $\lim_{n \rightarrow \infty} \ell_n/n = 0$ and $\lim_{n \rightarrow \infty} \ell_n/\sqrt{n} = \infty$. Define

$$\mathcal{A}_{Y_2}(\mathbf{m}_1, \mathbf{m}_2) \triangleq \{y_2^n : g^{(n)}(\mathbf{m}_1, \mathbf{m}_2, y_2^n) = 0\}, \quad (\text{A.71})$$

and its Hamming neighborhood

$$\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_1, \mathbf{m}_2) \triangleq \{\tilde{y}_2^n : \exists y_2^n \in \mathcal{A}_{Y_2}(\mathbf{m}_1, \mathbf{m}_2) \text{ s.t. } d_H(y_2^n, \tilde{y}_2^n) \leq \ell_n\}. \quad (\text{A.72})$$

Since by Condition (3.48),

$$P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\mathcal{A}_{Y_2}(\mathbf{m}_1, \mathbf{m}_2) | y_0^n, y_1^n) \geq \eta, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (\text{A.73})$$

the blowing-up lemma [40] yields

$$P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_1, \mathbf{m}_2) | y_0^n, y_1^n) \geq 1 - \zeta_n, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (\text{A.74})$$

for real numbers $\zeta_n > 0$ such that $\lim_{n \rightarrow \infty} \zeta_n = 0$.

Define

$$\mathcal{A}_{Y_2} \triangleq \bigcup_{(\mathbf{m}_1, \mathbf{m}_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \{(\mathbf{m}_1, \mathbf{m}_2)\} \times \mathcal{A}_{Y_2}(\mathbf{m}_1, \mathbf{m}_2), \quad (\text{A.75})$$

$$\hat{\mathcal{A}}_{Y_2}^{\ell_n} \triangleq \bigcup_{(\mathbf{m}_1, \mathbf{m}_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \{(\mathbf{m}_1, \mathbf{m}_2)\} \times \hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_1, \mathbf{m}_2), \quad (\text{A.76})$$

and notice that

$$P_{\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2 \tilde{Y}_2^n} \left(\hat{\mathcal{A}}_{Y_2}^{\ell_n} \right) = \sum_{(y_0^n, y_1^n) \in \mathcal{D}} P_{\tilde{Y}_0^n \tilde{Y}_1^n}(y_0^n, y_1^n) \cdot P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\mathcal{A}_{Y_2}(\phi_1(y_0^n), \phi_2(\phi_1(y_0^n), y_1^n))) | y_0^n, y_1^n) \quad (\text{A.77})$$

$$\geq (1 - \zeta_n). \quad (\text{A.78})$$

Then we can write

$$P_{\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2} P_{\tilde{Y}_2^n} \left(\hat{\mathcal{A}}_{Y_2, n}^{\ell_n} \right) \leq P_{\mathcal{M}_1 \mathcal{M}_2} P_{Y_2}^n \left(\hat{\mathcal{A}}_{Y_2, n}^{\ell_n} \right) \Delta_n^{-2} \quad (\text{A.79})$$

$$= \sum_{(\mathbf{m}_1, \mathbf{m}_2) \in \mathcal{M}_1 \times \mathcal{M}_2} P_{\mathcal{M}_1 \mathcal{M}_2}(\mathbf{m}_1, \mathbf{m}_2) P_{Y_2}^n \left(\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_1, \mathbf{m}_2) \right) \Delta_n^{-2} \quad (\text{A.80})$$

$$\leq \sum_{(\mathbf{m}_1, \mathbf{m}_2) \in \mathcal{M}_1 \times \mathcal{M}_2} P_{\mathcal{M}_1 \mathcal{M}_2}(\mathbf{m}_1, \mathbf{m}_2) P_{Y_2}^n(\mathcal{A}_{Y_2}(\mathbf{m}_1, \mathbf{m}_2)) \cdot e^{nh_b(\ell_n/n)} |\mathcal{Y}_2|^{\ell_n} k_n^{\ell_n} \Delta_n^{-2} \quad (\text{A.81})$$

$$= \beta_{2,n} e^{n\delta_n}, \quad (\text{A.82})$$

where $\delta_n \triangleq h_b(\ell_n/n) + \frac{\ell_n}{n} \log(|\mathcal{Y}_2| \cdot k_n) - \frac{2}{n} \log \Delta_n$ and $k_n \triangleq \min_{\substack{y_2, y_2': \\ P_{Y_2}(y_2') > 0}} \frac{P_{Y_2}(y_2)}{P_{Y_2}(y_2')}$. Here, (A.82) holds by [44, Proof of Lemma 5.1].

Combining (A.82) with (A.78) and standard inequalities (see [20, Lemma 1]), we then obtain

$$-\frac{1}{n} \log \beta_n \leq -\frac{1}{n} \log \left(P_{\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2} P_{\tilde{Y}_2^n} \left(\hat{\mathcal{A}}_{Y_2}^{\ell_n} \right) \right) + \delta_n \quad (\text{A.83})$$

$$\leq \frac{1}{n(1 - \zeta_n)} D(P_{\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2 \tilde{Y}_2^n} \| P_{\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2} P_{\tilde{Y}_2^n}) + \delta_n + \frac{1}{n} \quad (\text{A.84})$$

$$= \frac{1}{n(1 - \zeta_n)} I(\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2; \tilde{Y}_2^n) + \delta_n + \frac{1}{n}. \quad (\text{A.85})$$

We further upper-bound the term $I(\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2; \tilde{Y}_2^n)$ as follows

$$I(\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2; \tilde{Y}_2^n) \leq \sum_{t=1}^n I(\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2 \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1} \tilde{Y}_2^{t-1}; \tilde{Y}_{2,t}) \quad (\text{A.86})$$

$$= \sum_{t=1}^n I(\tilde{\mathcal{M}}_1 \tilde{\mathcal{M}}_2 \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}; \tilde{Y}_{2,t}) \quad (\text{A.87})$$

$$= \sum_{t=1}^n I(\tilde{U}_{1,t} \tilde{U}_{2,t}; \tilde{Y}_{2,t}) \quad (\text{A.88})$$

$$= nI(\tilde{U}_{1,T} \tilde{U}_{2,T}; \tilde{Y}_{2,T} | T) \quad (\text{A.89})$$

$$\leq nI(U_1 U_2; \tilde{Y}_2), \quad (\text{A.90})$$

where (A.87) holds by the Markov chain $\tilde{Y}_2^{t-1} \rightarrow (\tilde{M}_1 \tilde{M}_2, \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}) \rightarrow \tilde{Y}_{2,t}$; (A.88) follows by the definitions of $\tilde{U}_{1,t}$ and $\tilde{U}_{2,t}$; and (A.90) follows by the definitions of T , U_1 , U_2 , and \tilde{Y}_2 in the lemma.

We observe that the Markov chain $\tilde{U}_{2,t} \rightarrow (\tilde{U}_{1,t}, \tilde{Y}_{1,t}) \rightarrow \tilde{Y}_{2,t}$ holds for any t , and thus $U_2 \rightarrow (U_1, \tilde{Y}_1) \rightarrow \tilde{Y}_2$ holds as mentioned in the lemma.

Finally, we proceed to prove the Markov chain $U_1 \rightarrow \tilde{Y}_0 \rightarrow (\tilde{Y}_1, \tilde{Y}_2)$ in the limit as $n \rightarrow \infty$. To this end, notice the Markov chain $\tilde{M}_1 \rightarrow \tilde{Y}_0^n \rightarrow (\tilde{Y}_1^n, \tilde{Y}_2^n)$, and thus

$$0 = I(\tilde{M}_1; \tilde{Y}_1^n \tilde{Y}_2^n | \tilde{Y}_0^n) \quad (\text{A.91})$$

$$\geq H(\tilde{Y}_1^n \tilde{Y}_2^n | \tilde{Y}_0^n) - H(\tilde{Y}_1^n \tilde{Y}_2^n | \tilde{Y}_0^n \tilde{M}_1) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n \tilde{Y}_2^n} \| P_{\tilde{Y}_0^n Y_1 Y_2}^n) + \log \Delta_n \quad (\text{A.92})$$

$$\geq n[H(\tilde{Y}_{1,T} \tilde{Y}_{2,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T} \tilde{Y}_{2,T}} \| P_{Y_0 Y_1 Y_2})] + \log \Delta_n - H(\tilde{Y}_1^n \tilde{Y}_2^n | \tilde{Y}_0^n \tilde{M}_1) \quad (\text{A.93})$$

$$\geq n[H(\tilde{Y}_{1,T} \tilde{Y}_{2,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T} \tilde{Y}_{2,T}} \| P_{Y_0 Y_1 Y_2})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{1,t} \tilde{Y}_{2,t} | \tilde{Y}_{0,t} \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1} \tilde{M}_1) \quad (\text{A.94})$$

$$= n[H(\tilde{Y}_{1,T} \tilde{Y}_{2,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T} \tilde{Y}_{2,T}} \| P_{Y_0 Y_1 Y_2})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{1,t} \tilde{Y}_{2,t} | \tilde{Y}_{0,t} \tilde{U}_{1,t}) \quad (\text{A.95})$$

$$\geq n[H(\tilde{Y}_{1,T} \tilde{Y}_{2,T} | \tilde{Y}_{0,T}) - H(\tilde{Y}_{1,T} \tilde{Y}_{2,T} | \tilde{Y}_{0,T}, \tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{A.96})$$

$$\geq nI(\tilde{Y}_1 \tilde{Y}_2; U_1 | \tilde{Y}_0) + \log \Delta_n, \quad (\text{A.97})$$

where (A.93) holds by the super-additivity property in [36, Proposition 1]; (A.94) by the chain rule and since conditioning reduces entropy; (A.95) by the definition of $\tilde{U}_{1,t}$ and by recalling that T is uniform over $\{1, \dots, n\}$ independent of all other random quantities; (A.96) by the non-negativity of the Kullback-Leibler divergence; and finally (A.97) holds by the definitions of U_1 , \tilde{Y}_0 , \tilde{Y}_1 in the lemma.

Appendix B

Proofs for Chapter 4

B.1 Alternative Strong Converse Proof to Theorem 8

Notice that this proof depends on Lemma 4 stated in Section 4.3.5 and proved ahead in Appendix B.5.

Fix an exponent pair in $\mathcal{E}_{\text{Two-Hop,max}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ and a sequence (in n) of encoding and decision functions $\{(\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ satisfying the constraints on the rate and the error probabilities in (4.7a), (4.7b), (4.7c).

Fix a positive $\eta > 0$ and a blocklength n and choose $\mu_n = n^{-1/3}$. Define for $i \in \{1, 2\}$, the sets

$$\mathcal{B}_i(\eta) \triangleq \{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) : \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \geq \eta\}, \quad i \in \{1, 2\}, \quad (\text{B.1})$$

$$\mathcal{D}(\eta) \triangleq \mathcal{B}_1(\eta) \cap \mathcal{B}_2(\eta). \quad (\text{B.2})$$

Further define the probability

$$\Delta \triangleq P_{Y_0^n Y_1^n}(\mathcal{D}(\eta)), \quad (\text{B.3})$$

and notice that by the laws of probability

$$\Delta \geq \max\{P_{Y_0^n Y_1^n}(\mathcal{B}_1(\eta)) + P_{Y_0^n Y_1^n}(\mathcal{B}_2(\eta)) - 1, 0\}. \quad (\text{B.4})$$

By (4.7a), it can further be shown that for $i \in \{1, 2\}$:

$$1 - \epsilon_i \leq \sum_{y_0^n, y_1^n} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \quad (\text{B.5})$$

$$\begin{aligned} &\leq \sum_{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \\ &+ \sum_{(y_0^n, y_1^n) \notin \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1})} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \end{aligned} \quad (\text{B.6})$$

$$\begin{aligned} &\leq \sum_{(y_0^n, y_1^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}) \cap \bar{\mathcal{B}}_i(\eta)} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, Y_1^n = y_1^n, \mathcal{H} = 0] \cdot P_{Y_0^n Y_1^n}(y_0^n, y_1^n) \\ &\quad + \sum_{(y_0^n, y_1^n) \in \mathcal{B}_i(\eta)} P_{Y_0^n Y_1^n}(y_0^n, y_1^n) + (1 - P_{Y_0^n Y_1^n}(\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 Y_1}))) \end{aligned} \quad (\text{B.7})$$

$$\leq \eta(1 - P_{Y_0^n Y_1^n}(\mathcal{B}_i(\eta))) + P_{Y_0^n Y_1^n}(\mathcal{B}_i(\eta)) + P_{Y_0^n Y_1^n}(\bar{\mathcal{T}}_{\mu_n}^{(n)}(P_{Y_0 Y_1})). \quad (\text{B.8})$$

Thus, by (B.8) and [44, Lemma 2.12]:

$$\Delta \geq \max \left\{ \frac{1 - \epsilon_1 - \epsilon_2 - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0| |\mathcal{Y}_1|}{(1 - \eta) 2\mu_n^2 n}, 0 \right\}, \quad (\text{B.9})$$

and we conclude that in the limit as $n \rightarrow \infty$ and $\eta \downarrow 0$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}. \quad (\text{B.10})$$

We proceed by applying Lemma 4 to the set $\mathcal{D}(\eta)$ when it has positive probability $\Delta > 0$, otherwise the proof is similar and omitted for brevity. By this Lemma 4, and using also the maximum-rate constraints (4.7c), alongside the trivial inequality $nR_i \geq H(\tilde{\mathcal{M}}_i)$, for all $i \in \{1, 2\}$, we conclude that there is a pair (U_1, U_2) satisfying the Markov chain $U_2 \rightarrow \tilde{Y}_1 \rightarrow \tilde{Y}_2$ and the (in)equalities

$$nR_1 \geq H(\tilde{\mathcal{M}}_1) \geq nI(U_1; \tilde{Y}_0) + \log P_{Y_0^n Y_1^n}(\mathcal{D}(\eta)), \quad (\text{B.11})$$

$$nR_2 \geq H(\tilde{\mathcal{M}}_2) \geq nI(U_2; \tilde{Y}_1) + \log P_{Y_0^n Y_1^n}(\mathcal{D}(\eta)), \quad (\text{B.12})$$

$$\phi_1(n) = I(U_1; \tilde{Y}_1 | \tilde{Y}_0) \quad (\text{B.13})$$

and

$$-\frac{1}{n} \log \beta_{1,n} \leq I(U_1; \tilde{Y}_1) + \phi_2(n), \quad (\text{B.14})$$

$$-\frac{1}{n} \log \beta_{2,n} \leq I(U_1; \tilde{Y}_1) + I(U_2; \tilde{Y}_2) + \phi_3(n), \quad (\text{B.15})$$

where the functions $\phi_1(n), \phi_2(n) \downarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_0, \tilde{Y}_1, \tilde{Y}_2, \tilde{\mathcal{M}}_1, \tilde{\mathcal{M}}_2$ are defined as in the lemma (Lemma 4) applied to the set $\mathcal{D}(\eta)$.

Thus we have proved so far that for all $n \geq 1$ there exist joint pmfs $P_{U_1 \tilde{Y}_0 \tilde{Y}_1}$ (abbreviated as $P_1^{(n)}$) and $P_{U_2 \tilde{Y}_1 \tilde{Y}_2} = P_{\tilde{Y}_1 \tilde{Y}_2} P_{U_2 | \tilde{Y}_1}$ (abbreviated as $P_2^{(n)}$) so that the following conditions hold

$$R_1 \geq I_{P_1^{(n)}}(U_1; \tilde{Y}_0) + g_1(n), \quad (\text{B.16a})$$

$$R_2 \geq I_{P_2^{(n)}}(U_2; \tilde{Y}_1) + g_1(n), \quad (\text{B.16b})$$

$$g_2(n) = I_{P_1^{(n)}}(U_1; \tilde{Y}_1 | \tilde{Y}_0) \quad (\text{B.16c})$$

$$\theta_1 \leq I_{P_1^{(n)}}(U_1; \tilde{Y}_1) + g_3(n), \quad (\text{B.16d})$$

$$\theta_2 \leq I_{P_1^{(n)}}(U_1; \tilde{Y}_1) + I_{P_2^{(n)}}(U_2; \tilde{Y}_2) + g_4(n), \quad (\text{B.16e})$$

for some nonnegative functions $g_1(n), g_2(n), g_3(n), g_4(n)$ tending to 0 in the limit as $n \rightarrow \infty$.

The rest of the proof follows the same steps as the proof of the converse in Section 4.3.5.

By Carathéodory's theorem [48, Appendix C], there exist for each n random variables U_1, U_2 satisfying (B.16) over alphabets of sizes

$$|\mathcal{U}_1| \leq |\mathcal{Y}_0||\mathcal{Y}_1| + 2, \quad (\text{B.17})$$

$$|\mathcal{U}_2| \leq |\mathcal{Y}_1| + 1. \quad (\text{B.18})$$

Then we invoke the Bolzano-Weierstrass theorem and consider an increasing subsequence of positive numbers $\{n_k\}_{k=1}^{\infty}$ such that the following subsequences converge:

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_0 \tilde{Y}_1 U_1}^{(n_k)} = P_{Y_0 Y_1 U_1}^*, \quad (\text{B.19})$$

$$\lim_{k \rightarrow \infty} P_{\tilde{Y}_1 \tilde{Y}_2 U_2}^{(n_k)} = P_{Y_1 Y_2 U_2}^*. \quad (\text{B.20})$$

Considering further an appropriate sequence of diminishing η -values, and noticing that since the pair $(\tilde{Y}_0^{(n_k)}, \tilde{Y}_1^{(n_k)})$ lies in the jointly typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0 Y_1})$, we have $|P_{\tilde{Y}_0 \tilde{Y}_1} - P_{Y_0 Y_1}| \leq \mu_{n_k}$ and thus the limiting pmfs satisfy $P_{Y_0 Y_1}^* = P_{Y_0 Y_1}$. Moreover, since for each n_k the random variable \tilde{Y}_2 is drawn according to $P_{Y_2|Y_1}$ given \tilde{Y}_1 , irrespective of \tilde{Y}_0 , the limiting pmfs also satisfy $P_{Y_2|Y_0 Y_1}^* = P_{Y_2|Y_1}$. We also notice that under $P_{Y_1 Y_2 U_2}^*$ the Markov chain

$$U_2 \rightarrow Y_1 \rightarrow Y_2, \quad (\text{B.21})$$

holds because $U_2 \rightarrow \tilde{Y}_1 \rightarrow \tilde{Y}_2$ forms a Markov chain for any n_k . Finally, by continuity considerations and by (B.16c), the following Markov chain must hold under $P_{Y_0 Y_1 U_1}^*$:

$$U_1 \rightarrow Y_0 \rightarrow Y_1. \quad (\text{B.22})$$

By the above observations, by (B.16a)–(B.16e), and using the definitions of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$, we thus conclude the strong converse proof.

B.2 Analysis of the coding scheme in Subsection 4.3.3.1.1 for $\epsilon_1 = \epsilon_2 = \epsilon$

Consider the two-hop scheme employed when $Y_0^n \in \mathcal{D}_{\{1,2\}}$, and let $\hat{\mathcal{H}}_{\{1,2\},1}$ and $\hat{\mathcal{H}}_{\{1,2\},2}$ denote the guesses produced at R_1 and R_2 when employing this scheme for any $Y_0^n \in \mathcal{Y}_0^n$. Notice that by assumption the type-I error probabilities of this scheme tend to 0 as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1, 2\}. \quad (\text{B.23})$$

Noticing that when $Y_0^n \in \mathcal{D}_\emptyset$, then $\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1$, and applying the total law of probability, we can write for $k \in \{1, 2\}$:

$$\alpha_{k,n} = \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0] \quad (\text{B.24})$$

$$= \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.25})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.26})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] \quad (\text{B.27})$$

Combining these inequalities with (B.23), and because in the limit $n \rightarrow \infty$ Inequality (4.9) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraints:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{k,n} \leq \epsilon, \quad k \in \{1, 2\}. \quad (\text{B.28})$$

For the type-II error probabilities of the overall scheme we observe for $k \in \{1, 2\}$:

$$\beta_{1,n} = \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1] \quad (\text{B.29})$$

$$= \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.30})$$

$$= \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.31})$$

$$= \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.32})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 0 | \mathcal{H} = 1]. \quad (\text{B.33})$$

The type-II error exponents of the overall scheme are thus given by the error exponents of the two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [33] and because the two-hop scheme has vanishing type-I error probabilities and respect the rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$, the exponents in (4.23) are proved achievable.

B.3 Analysis of the coding scheme in Subsection 4.3.3.1.2 for $\epsilon_2 > \epsilon_1$

Consider the two-hop scheme employed when $Y_0^n \in \mathcal{D}_{\{1,2\}}$, and let $\hat{\mathcal{H}}_{\{1,2\},1}$ and $\hat{\mathcal{H}}_{\{1,2\},2}$ denote the guesses produced at R_1 and R_2 when employing this scheme for any $y_0^n \in \mathcal{Y}_0^n$. Similarly, let $\hat{\mathcal{H}}_{\{1\},1}$ and $\hat{\mathcal{H}}_{\{1\},2}$ denote the guesses produced at R_1 and R_2 when employing the scheme for $Y_0^n \in \mathcal{D}_{\{1\}}$, where we again extend the scheme to the entire set \mathcal{Y}_0^n .

By assumption, the type-I error probabilities of these schemes tend to 0 as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1, 2\} \quad (\text{B.34a})$$

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1, 2\}. \quad (\text{B.34b})$$

Notice that for $Y_0^n \in \mathcal{D}_\emptyset$ both R_1 and R_2 decide on $\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1$. Applying the total law of probability, we can write

$$\alpha_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \quad (\text{B.35})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.36})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.37})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1 | \mathcal{H} = 0] \quad (\text{B.38})$$

Combining this inequality with (B.34), and because in the limit $n \rightarrow \infty$ Inequality (4.13c) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{1,n} \leq \epsilon_1. \quad (\text{B.39})$$

Similarly we have:

$$\alpha_{2,n} = \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0] \quad (\text{B.40})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}}) | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.41})$$

$$= \Pr[Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}}) | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.42})$$

$$\leq \Pr[Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}}) | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1 | \mathcal{H} = 0]. \quad (\text{B.43})$$

Combining this inequality with (B.34), and because in the limit $n \rightarrow \infty$ Inequalities (4.13a) and (4.13c) turn into equalities, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{2,n} \leq \epsilon_2. \quad (\text{B.44})$$

For the relay's type-II error probability in the overall scheme we observe:

$$\beta_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \quad (\text{B.45})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.46})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.47})$$

$$= \Pr[\hat{\mathcal{H}}_{\{1\},1} = 0, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.48})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{1\},1} = 0 | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0 | \mathcal{H} = 1]. \quad (\text{B.49})$$

The relay's type-II error exponent of the overall scheme is thus given by the minimum of the error exponents of the single-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1\}}$ and of two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$.

By [2] and [33] and because these schemes have vanishing type-I error probabilities and respect the rate constraints $R_{\{1\},1}$ and $(R_{\{1,2\},1}, R_{\{1,2\},2})$, respectively, the exponent θ_1 in (4.24) is proved achievable.

It remains to analyze the receiver's type-II error exponent:

$$\beta_{2,n} = \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \quad (\text{B.50})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}}) | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.51})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.52})$$

$$= \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.53})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 0 | \mathcal{H} = 1]. \quad (\text{B.54})$$

The receiver's type-II error exponent of the overall scheme is thus given by the error exponent of the two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [33] and because this scheme has vanishing type-I error probabilities and respects the rate constraints $(R_{\{1,2\},1}, R_{\{1,2\},2})$, the exponent θ_2 in (4.24) is proved achievable.

B.4 Analysis of the coding scheme in Subsection 4.3.3.1.3 for $\epsilon_1 > \epsilon_2$

Now consider the two-hop scheme employed when $Y_0^n \in \mathcal{D}_{\{1,2\}}$, and let $\hat{\mathcal{H}}_{\{1,2\},1}$ and $\hat{\mathcal{H}}_{\{1,2\},2}$ denote the guesses produced at R_1 and R_2 when employing this scheme for any $y_0^n \in \mathcal{Y}_0^n$. Similarly, let $\hat{\mathcal{H}}_{\{2\},2}$ denote the guess produced at R_2 when employing the scheme for $Y_0^n \in \mathcal{D}_{\{2\}}$, where we again extend the scheme to the entire set \mathcal{Y}_0^n . Note that in this case, i.e. when $Y_0^n \in \mathcal{D}_{\{2\}}$, the guess produced by R_1 is $\hat{\mathcal{H}}_{\{2\},1} = 1$.

By assumption, the type-I error probabilities of these schemes have the following limits:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{2\},1} = 1 | \mathcal{H} = 0] = 1, \quad (\text{B.55a})$$

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1 | \mathcal{H} = 0] = 0, \quad (\text{B.55b})$$

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1, 2\}. \quad (\text{B.55c})$$

Notice that for $Y_0^n \in \mathcal{D}_\emptyset$ both R_1 and R_2 decide on $\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1$. Applying the total law of probability, we can write

$$\alpha_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \quad (\text{B.56})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.57})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.58})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1 | \mathcal{H} = 0] \quad (\text{B.59})$$

Combining this inequality with (B.55), and because in the limit $n \rightarrow \infty$ Inequalities (4.19a) and (4.19c) turn into equalities, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{1,n} \leq \epsilon_1. \quad (\text{B.60})$$

Similarly we have:

$$\alpha_{2,n} = \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0] \quad (\text{B.61})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.62})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \quad (\text{B.63})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1 | \mathcal{H} = 0] \quad (\text{B.64})$$

Combining this inequality with (B.55), and because in the limit $n \rightarrow \infty$ Inequality (4.19c) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{2,n} \leq \epsilon_2. \quad (\text{B.65})$$

For the relay's type-II error probability in the overall scheme we observe:

$$\beta_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \quad (\text{B.66})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{2\}}) | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.67})$$

$$= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.68})$$

$$= \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.69})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0 | \mathcal{H} = 1]. \quad (\text{B.70})$$

The relay's type-II error exponent of the overall scheme is thus given by the error exponent of the two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [33] and because this scheme has vanishing type-I error probabilities and respects the rate constraint $R_{\{1,2\},1}$, the exponent θ_1 in (4.25a) is proved achievable.

It remains to analyze the receiver's type-II error exponent:

$$\beta_{2,n} = \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \quad (\text{B.71})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.72})$$

$$= \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.73})$$

$$= \Pr[\hat{\mathcal{H}}_{\{2\},2} = 0, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 1] \quad (\text{B.74})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{2\},2} = 0 | \mathcal{H} = 1] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 0 | \mathcal{H} = 1]. \quad (\text{B.75})$$

The receiver's type-II error exponent of the overall scheme is thus given by the minimum of the error exponents of the two-hop schemes employed under $Y_0^n \in \mathcal{D}_{\{2\}}$ and under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [33] and because these schemes have vanishing type-I error probabilities and respect the rate constraints $(R_{\{2\},1}, R_{\{2\},2})$ and $(R_{\{1,2\},1}, R_{\{1,2\},2})$, the exponent θ_2 in (4.25b) is proved achievable.

B.5 Proof of Lemma 4

Note first that by (4.48):

$$D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0 Y_1}^n) \leq \log \Delta_n^{-1}, \quad (\text{B.76})$$

where we defined $\Delta_n \triangleq P_{Y_0^n Y_1^n}(\mathcal{D})$.

Further define

$$\tilde{U}_{2,t} \triangleq (\tilde{M}_2, \tilde{Y}_0^{t-1}, \tilde{Y}_1^{t-1}) \quad (\text{B.77})$$

$$\tilde{U}_{1,t} \triangleq (\tilde{M}_1, \tilde{Y}_0^{t-1}, \tilde{Y}_1^{t-1}), \quad (\text{B.78})$$

and notice:

$$H(\tilde{M}_1) \geq I(\tilde{M}_1; \tilde{Y}_0^n \tilde{Y}_1^n) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0 Y_1}^n) + \log \Delta_n \quad (\text{B.79})$$

$$= H(\tilde{Y}_0^n \tilde{Y}_1^n) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{Y_0 Y_1}^n) - H(\tilde{Y}_0^n \tilde{Y}_1^n | \tilde{M}_1) + \log \Delta_n \quad (\text{B.80})$$

$$\geq n[H(\tilde{Y}_{0,T} \tilde{Y}_{1,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T}} \| P_{Y_0 Y_1})] - \sum_{t=1}^n H(\tilde{Y}_{0,t} \tilde{Y}_{1,t} | \tilde{U}_{1,t}) + \log \Delta_n \quad (\text{B.81})$$

$$= n[H(\tilde{Y}_{0,T} \tilde{Y}_{1,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T}} \| P_{Y_0 Y_1}) - H(\tilde{Y}_{0,T} \tilde{Y}_{1,T} | \tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{B.82})$$

$$\geq n[H(\tilde{Y}_{0,T} \tilde{Y}_{1,T}) - H(\tilde{Y}_{0,T} \tilde{Y}_{1,T} | \tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{B.83})$$

$$= n[I(\tilde{Y}_0 \tilde{Y}_1; U_1)] + \log \Delta_n \quad (\text{B.84})$$

$$\geq n \left[I(\tilde{Y}_0; U_1) + \frac{1}{n} \log \Delta_n \right]. \quad (\text{B.85})$$

Here, (B.79) holds by (B.76); (B.81) holds by the super-additivity property in [36, Proposition 1], by the chain rule, and by the definition of $\tilde{U}_{1,t}$ and by defining T uniform over $\{1, \dots, n\}$ independent of all other random variables; and (B.84) holds by the definitions of $U_1, \tilde{Y}_0, \tilde{Y}_1$ in the lemma.

We can lower bound the entropy of \tilde{M}_2 in a similar way to obtain:

$$H(\tilde{M}_2) \geq n \left[I(\tilde{Y}_1; U_2) + \frac{1}{n} \log \Delta_n \right]. \quad (\text{B.86})$$

We next upper bound the error exponent at the receiver. To this end, fix a sequence of real numbers

$\{\ell_n\}_{n=1}^\infty$ satisfying $\lim_{n \rightarrow \infty} \ell_n/n = 0$ and $\lim_{n \rightarrow \infty} \ell_n/\sqrt{n} = \infty$. Define

$$\mathcal{A}_{Y_2}(\mathbf{m}_2) \triangleq \{y_2^n : g_2(\mathbf{m}_2, y_2^n) = 0\}, \quad (\text{B.87})$$

and its Hamming neighborhood:

$$\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_2) \triangleq \{\tilde{y}_2^n : \exists y_2^n \in \mathcal{A}_{Y_2}(\mathbf{m}_2) \text{ s.t. } d_H(y_2^n, \tilde{y}_2^n) \leq \ell_n\}. \quad (\text{B.88})$$

Since by Condition (4.55),

$$P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\mathcal{A}_{Y_2}(\mathbf{m}_2) | y_0^n, y_1^n) \geq \eta, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (\text{B.89})$$

the blowing-up lemma [40] yields

$$P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_2) | y_0^n, y_1^n) \geq 1 - \zeta_n, \quad \forall (y_0^n, y_1^n) \in \mathcal{D}, \quad (\text{B.90})$$

for real numbers $\zeta_n > 0$ such that $\lim_{n \rightarrow \infty} \zeta_n = 0$.

Define

$$\mathcal{A}_{Y_2} \triangleq \bigcup_{\mathbf{m}_2 \in \mathcal{M}_2} \{\mathbf{m}_2\} \times \mathcal{A}_{Y_2}(\mathbf{m}_2), \quad (\text{B.91})$$

$$\hat{\mathcal{A}}_{Y_2}^{\ell_n} \triangleq \bigcup_{\mathbf{m}_2 \in \mathcal{M}_2} \{\mathbf{m}_2\} \times \hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_2), \quad (\text{B.92})$$

and notice that

$$P_{\tilde{M}_2 \tilde{Y}_2^n}(\hat{\mathcal{A}}_{Y_2}^{\ell_n}) = \sum_{(y_0^n, y_1^n) \in \mathcal{D}} P_{\tilde{Y}_0^n \tilde{Y}_1^n}(y_0^n, y_1^n) \cdot P_{\tilde{Y}_2^n | \tilde{Y}_0^n \tilde{Y}_1^n}(\mathcal{A}_{Y_2}(\phi_2(\phi_1(y_0^n), y_1^n))) | y_0^n, y_1^n) \quad (\text{B.93})$$

$$\geq (1 - \zeta_n). \quad (\text{B.94})$$

Defining

$$Q_{\tilde{M}_2}(\mathbf{m}_2) \triangleq \sum_{y_1^n, \mathbf{m}_1} P_{\tilde{M}_1}(\mathbf{m}_1) P_{\tilde{Y}_1^n}(y_1^n) \cdot \mathbb{1}\{\phi_2(\mathbf{m}_1, y_1^n) = \mathbf{m}_2\}, \quad (\text{B.95})$$

we can write

$$Q_{\tilde{M}_2} P_{\tilde{Y}_2^n}(\hat{\mathcal{A}}_{Y_2, n}^{\ell_n}) \leq Q_{M_2} P_{Y_2}^n(\hat{\mathcal{A}}_{Y_2, n}^{\ell_n}) \Delta_n^{-3} \quad (\text{B.96})$$

$$= \sum_{\mathbf{m}_2 \in \mathcal{M}_2} Q_{M_2}(\mathbf{m}_2) P_{Y_2}^n(\hat{\mathcal{A}}_{Y_2}^{\ell_n}(\mathbf{m}_2)) \Delta_n^{-3} \quad (\text{B.97})$$

$$\leq \sum_{\mathbf{m}_2 \in \mathcal{M}_2} Q_{M_2}(\mathbf{m}_2) P_{Y_2}^n(\mathcal{A}_{Y_2}(\mathbf{m}_2)) \cdot e^{nh_b(\ell_n/n)} |\mathcal{Y}_2|^{\ell_n} k_n^{\ell_n} \Delta_n^{-3} \quad (\text{B.98})$$

$$= \beta_{2, n} e^{n\delta_n}, \quad (\text{B.99})$$

where $\delta_n \triangleq h_b(\ell_n/n) + \frac{\ell_n}{n} \log(|\mathcal{Y}_2| \cdot k_n) - \frac{3}{n} \log \Delta_n$ and $k_n \triangleq \min_{y_2, y_2': P_{Y_2}(y_2') > 0} \frac{P_{Y_2}(y_2)}{P_{Y_2}(y_2')}$. Here, (B.99) holds by [44, Proof

of Lemma 5.1].

Combining (B.99) with (B.94) and standard inequalities (see [20, Lemma 1]), we then obtain:

$$-\frac{1}{n} \log \beta_{2,n} \leq -\frac{1}{n} \log \left(Q_{\tilde{M}_2} P_{\tilde{Y}_2^n} \left(\hat{A}_{Y_2}^{\ell_n} \right) \right) + \delta_n \quad (\text{B.100})$$

$$\leq \frac{1}{n(1-\zeta_n)} D(P_{\tilde{M}_2 \tilde{Y}_2^n} \| Q_{\tilde{M}_2} P_{\tilde{Y}_2^n}) + \delta_n + \frac{1}{n}. \quad (\text{B.101})$$

We continue to upper bound the divergence term as

$$D(P_{\tilde{M}_2 \tilde{Y}_2^n} \| Q_{\tilde{M}_2} P_{\tilde{Y}_2^n}) = I(\tilde{M}_2; \tilde{Y}_2^n) + D(P_{\tilde{M}_2} \| Q_{\tilde{M}_2}) \quad (\text{B.102})$$

$$\leq I(\tilde{M}_2; \tilde{Y}_2^n) + D(P_{\tilde{Y}_1^n \tilde{M}_1} \| P_{\tilde{Y}_1^n} P_{\tilde{M}_1}) \quad (\text{B.103})$$

$$= I(\tilde{M}_2; \tilde{Y}_2^n) + I(\tilde{M}_1; \tilde{Y}_1^n) \quad (\text{B.104})$$

$$= \sum_{t=1}^n I(\tilde{M}_2; \tilde{Y}_{2,t} | \tilde{Y}_2^{t-1}) + I(\tilde{M}_1; \tilde{Y}_{1,t} | \tilde{Y}_1^{t-1}) \quad (\text{B.105})$$

$$\leq \sum_{t=1}^n I(\tilde{M}_2 \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}; \tilde{Y}_{2,t}) + I(\tilde{M}_1 \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}; \tilde{Y}_{1,t}) \quad (\text{B.106})$$

$$= \sum_{t=1}^n I(\tilde{U}_{2,t}; \tilde{Y}_{2,t}) + I(\tilde{U}_{1,t}; \tilde{Y}_{1,t}) \quad (\text{B.107})$$

$$= n[I(\tilde{U}_{2,T}; \tilde{Y}_{2,T} | T) + I(\tilde{U}_{1,T}; \tilde{Y}_{1,T} | T)] \quad (\text{B.108})$$

$$\leq n[I(\tilde{U}_{2,T} T; \tilde{Y}_{2,T}) + I(\tilde{U}_{1,T} T; \tilde{Y}_{1,T})] \quad (\text{B.109})$$

$$= n[I(U_2; \tilde{Y}_2) + I(U_1; \tilde{Y}_1)]. \quad (\text{B.110})$$

Here (B.103) is obtained by the data processing inequality for KL-divergence; (B.105) by the chain rule; (B.106) by the Markov chain $\tilde{Y}_2^{t-1} \rightarrow (\tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}) \rightarrow \tilde{Y}_{2,t}$; and (B.107)–(B.110) by the definitions of $\tilde{U}_{1,t}, \tilde{U}_{2,t}, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2$.

Following similar steps, we now prove the desired upper bound on the relay's error exponent. Define the acceptance region at R_1 as

$$\mathcal{A}_{Y_1} \triangleq \{(\mathbf{m}_1, y_1^n) : g_1(\mathbf{m}_1, y_1^n) = 0\}. \quad (\text{B.111})$$

Notice that for any given (y_0^n, y_1^n) the pair $(\mathbf{m}_1 = \phi_0(y_0^n), y_1^n)$ lies inside the acceptance region \mathcal{A}_{Y_1} with probability either 0 or 1. Thus, for any $\eta > 0$ Condition (4.57) implies that for all pairs $(y_0^n, y_1^n) \in \mathcal{D}$ the corresponding pairs (\mathbf{m}_1, y_1^n) lie in \mathcal{A}_{Y_1} with probability 1:

$$P_{\tilde{M}_1 \tilde{Y}_1^n}(\mathcal{A}_{Y_1}) = 1. \quad (\text{B.112})$$

Following similar steps as in the analysis of the receiver's error exponent:

$$P_{\tilde{M}_1} P_{\tilde{Y}_1^n}(\mathcal{A}_{Y_1}) \leq P_{M_1} P_{Y_1^n}(\mathcal{A}_{Y_1}) \Delta_n^{-2} \quad (\text{B.113})$$

$$= \beta_{1,n} \Delta_n^{-2}. \quad (\text{B.114})$$

Combining (B.112) and (B.114) with standard inequalities (see [20, Lemma 1]), we further obtain

$$-\frac{1}{n} \log \beta_{1,n} \leq -\frac{1}{n} \log \left(P_{\tilde{M}_1} P_{\tilde{Y}_1^n} (\mathcal{A}_{Y_1,n}) \right) - \frac{2}{n} \log \Delta_n \quad (\text{B.115})$$

$$\leq \frac{1}{n} D(P_{\tilde{M}_1 \tilde{Y}_1^n} \| P_{\tilde{M}_1} P_{\tilde{Y}_1^n}) + \delta'_n \quad (\text{B.116})$$

where $\delta'_n \triangleq -\frac{2}{n} \log \Delta_n + \frac{1}{n}$ and tends to 0 as $n \rightarrow \infty$.

We continue to upper bound the divergence term as

$$D(P_{\tilde{M}_1 \tilde{Y}_1^n} \| P_{\tilde{M}_1} P_{\tilde{Y}_1^n}) = I(\tilde{M}_1; \tilde{Y}_1^n) \quad (\text{B.117})$$

$$= \sum_{t=1}^n I(\tilde{M}_1; \tilde{Y}_{1,t} | \tilde{Y}_1^{t-1}) \quad (\text{B.118})$$

$$\leq \sum_{t=1}^n I(\tilde{M}_1 \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1}; \tilde{Y}_{1,t}) \quad (\text{B.119})$$

$$= \sum_{t=1}^n I(\tilde{U}_{1,t}; \tilde{Y}_{1,t}) \quad (\text{B.120})$$

$$= n[I(\tilde{U}_{1,T}; \tilde{Y}_{1,T} | T)] \quad (\text{B.121})$$

$$\leq n[I(\tilde{U}_{1,T} T; \tilde{Y}_{1,T})] \quad (\text{B.122})$$

$$= n[I(U_1; \tilde{Y}_1)]. \quad (\text{B.123})$$

Here (B.118) holds by the chain rule and (B.120)–(B.123) hold by the definitions of $\tilde{U}_{1,t}, U_1, \tilde{Y}_1$.

Finally, we proceed to prove the Markov chain $U_1 \rightarrow \tilde{Y}_0 \rightarrow \tilde{Y}_1$ in the limit as $n \rightarrow \infty$. To this end, notice the Markov chain $\tilde{M}_1 \rightarrow \tilde{Y}_0^n \rightarrow \tilde{Y}_1^n$, and thus:

$$0 = I(\tilde{M}_1; \tilde{Y}_1^n | \tilde{Y}_0^n) \quad (\text{B.124})$$

$$\geq H(\tilde{Y}_1^n | \tilde{Y}_0^n) - H(\tilde{Y}_1^n | \tilde{Y}_0^n \tilde{M}_1) + D(P_{\tilde{Y}_0^n \tilde{Y}_1^n} \| P_{\tilde{Y}_0^n Y_1}^n) + \log \Delta_n \quad (\text{B.125})$$

$$\geq n[H(\tilde{Y}_{1,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T}} \| P_{Y_0 Y_1})] + \log \Delta_n - H(\tilde{Y}_1^n | \tilde{Y}_0^n \tilde{M}_1) \quad (\text{B.126})$$

$$\geq n[H(\tilde{Y}_{1,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T}} \| P_{Y_0 Y_1})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{1,t} | \tilde{Y}_{0,t} \tilde{Y}_0^{t-1} \tilde{Y}_1^{t-1} \tilde{M}_1) \quad (\text{B.127})$$

$$= n[H(\tilde{Y}_{1,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \tilde{Y}_{1,T}} \| P_{Y_0 Y_1})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{1,t} | \tilde{Y}_{0,t} \tilde{U}_{1,t}) \quad (\text{B.128})$$

$$\geq n[H(\tilde{Y}_{1,T} | \tilde{Y}_{0,T}) - H(\tilde{Y}_{1,T} | \tilde{Y}_{0,T}, \tilde{U}_{1,T}, T)] + \log \Delta_n \quad (\text{B.129})$$

$$\geq nI(\tilde{Y}_1; U_1 | \tilde{Y}_0) + \log \Delta_n, \quad (\text{B.130})$$

where (B.126) holds by the super-additivity property in [36, Proposition 1]; (B.127) by the chain rule and since conditioning reduces entropy; (B.128) by the definition of $\tilde{U}_{1,t}$ and by recalling that T is uniform over

$\{1, \dots, n\}$ independent of all other random quantities; (B.129) by the non-negativity of the Kullback-Leibler divergence; and finally (B.130) holds by the definitions of $U_1, \tilde{Y}_0, \tilde{Y}_1$.

B.6 Strong Converse Proof of Theorem 11

This strong converse proof relies on the method of change of measure arguments and asymptotic Markov chains similarly to the one explained in more detail in Section 4.4.3.4. The main idea here is to apply the change of measure arguments at each DC by restricting to observations in the intersection set between the jointly typical sequences (of all observations until the current DC) and the acceptance region at the DC. This leads to lower bounds on the rates of all links prior to the DC and to an upper bound on the error exponent of the current DC. Since the different bounds have different distributions, the proof is concluded by continuity and convergence arguments in the limit as $n \rightarrow \infty$ where the required Markov chains are also shown to hold asymptotically. The technical proof is presented below.

Fix an exponent-tuple $(\theta_1, \dots, \theta_K)$ in the exponents region $\mathcal{E}^*(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$, and a sequence (in n) of encoding and decision functions $\{(\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, \dots, g_K^{(n)})\}_{n \geq 1}$ achieving this tuple, i.e., satisfying constraints (4.182) for maximum-rate constraints.

Fix an arbitrary $k \in \{1, \dots, K\}$ and set $\mu_n = n^{-1/3}$. Let \mathcal{A}_k denote the acceptance region of R_k , i.e.,

$$\mathcal{A}_k := \{(y_0^n, \dots, y_k^n) : g_k^{(n)}(\mathbf{m}_k, y_k^n) = 0\}, \quad (\text{B.131})$$

where we define recursively $\mathbf{m}_1 := \phi_0^{(n)}(y_0^n)$ and

$$\mathbf{m}_\ell := \phi_{\ell-1}^{(n)}(\mathbf{m}_{\ell-1}, y_{\ell-1}^n), \quad \ell \in \{2, \dots, k\}. \quad (\text{B.132})$$

Define also the intersection of this acceptance region with the typical set:

$$\mathcal{D}_k \triangleq \mathcal{A}_k \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \dots Y_k}). \quad (\text{B.133})$$

By [44, Remark to Lemma 2.12] and the type-I error probability constraints in (4.182a),

$$\Delta_k := P_{Y_0^n Y_1^n \dots Y_k^n}(\mathcal{D}_k) \geq 1 - \epsilon_k - \frac{|\mathcal{Y}_0| \cdots |\mathcal{Y}_k|}{4\mu_n^2 n}, \quad (\text{B.134})$$

and thus $\underline{\lim}_{n \rightarrow \infty} \Delta_k \geq 1 - \epsilon_k > 0$ as $n \rightarrow \infty$.

Let $(\tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_k^n)$ be random variables of joint pmf

$$P_{\tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_k^n}(\tilde{y}_0^n, \tilde{y}_1^n, \dots, \tilde{y}_k^n) = \frac{P_{Y_0^n, Y_1^n, \dots, Y_k^n}(\tilde{y}_0^n, \tilde{y}_1^n, \dots, \tilde{y}_k^n)}{\Delta} \cdot \mathbf{1}\{(\tilde{y}_0^n, \tilde{y}_1^n, \dots, \tilde{y}_k^n) \in \mathcal{D}_k\}. \quad (\text{B.135})$$

Let also $\tilde{\mathbf{M}}_\ell = \phi_{\ell-1}^{(n)}(\tilde{\mathbf{M}}_{\ell-1}, \tilde{Y}_{\ell-1}^n)$ and T be uniform over $\{1, \dots, n\}$ independent of

$(\tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_k^n, \tilde{\mathbf{M}}_1, \dots, \tilde{\mathbf{M}}_k)$, and define $\tilde{Y}_\ell := \tilde{Y}_{\ell,T}$ for $\ell \in \{1, \dots, k\}$.

Lemma 10. *There exist random variables $\{U_1, \dots, U_k\}$ satisfying the (in)equalities*

$$nR_\ell \geq H(\tilde{\mathbf{M}}_\ell) \geq nI(U_\ell; \tilde{Y}_{\ell-1}) + \log \Delta_k, \quad \ell \in \{1, \dots, k\}, \quad (\text{B.136a})$$

$$I(U_\ell; \tilde{Y}_\ell | \tilde{Y}_{\ell-1}) = \phi_{1,\ell}(n), \quad (\text{B.136b})$$

and

$$-\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1, (Y_0^n, \dots, Y_k^n) \in \mathcal{D}_k] \leq \sum_{\ell=1}^k I(U_\ell; \tilde{Y}_\ell) + \phi_2(n), \quad (\text{B.136c})$$

where the functions $\{\phi_{1,\ell}(n)\}_{\ell=1}^k$ and $\phi_2(n)$ all tend to 0 as $n \rightarrow \infty$.

Proof: Similar to the proof of Lemma 6 which is presented later in Appendix B.8, in addition to the trivial inequality $nR_\ell \geq H(\tilde{\mathbf{M}}_\ell)$. ■

The desired bound on θ_k in (4.192) is then obtained from above lemma by taking $n \rightarrow \infty$, as we explain in the following. By Carathéodory's theorem [48, Appendix C], for each n there must exist random variables U_1, \dots, U_k satisfying (B.136) over alphabets of sizes

$$|\mathcal{U}_\ell| \leq |\mathcal{Y}_{\ell-1}| \cdot |\mathcal{Y}_\ell| + 2, \quad \ell \in \{1, \dots, k\}. \quad (\text{B.137})$$

We thus restrict to random variables of above (bounded) supports and invoke the Bolzano-Weierstrass theorem to conclude the existence of a pmf $P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell}^{(\ell)}$ over $\mathcal{Y}_{\ell-1} \times \mathcal{Y}_\ell \times \mathcal{U}_\ell$, also abbreviated as $P^{(\ell)}$, and an increasing subsequence of positive numbers $\{n_i\}_{i=1}^\infty$ satisfying

$$\lim_{i \rightarrow \infty} P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell; n_i} = P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell}^{(\ell)}, \quad \ell \in \{1, \dots, k\}, \quad (\text{B.138})$$

where $P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell; n_i}$ denotes the pmf at blocklength n_i .

By the monotone continuity of mutual information over finite pmfs, we can then deduce that

$$R_\ell \geq I_{P^{(\ell)}}(U_\ell; \tilde{Y}_{\ell-1}), \quad \ell \in \{1, \dots, k\}, \quad (\text{B.139})$$

$$\theta_k \leq \sum_{\ell=1}^k I_{P^{(\ell)}}(U_\ell; \tilde{Y}_\ell), \quad (\text{B.140})$$

where the subscripts indicate that mutual informations should be computed according to the indicated pmfs.

Since for any blocklength n_i the pair $(\tilde{Y}_{\ell-1}^{n_i}, \tilde{Y}_\ell^{n_i})$ lies in the jointly typical set $\mathcal{T}_{\mu_{n_i}}^{(n_i)}(P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell})$, we have $|P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell; n_i} - P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell}| \leq \mu_{n_i}$ and thus the limiting pmfs satisfy $P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell}^{(\ell)} = P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell}$. By similar continuity

considerations and by (B.136b), for all $\ell \in \{1, \dots, k\}$ the Markov chain

$$U_\ell \rightarrow Y_{\ell-1} \rightarrow Y_\ell, \quad (\text{B.141})$$

holds under $P_{Y_{\ell-1}Y_\ell U_\ell}^{(\ell)}$.

By the definitions of the functions $\{\eta_\ell(\cdot)\}$ and by (B.139)–(B.141):

$$\theta_k \leq \sum_{\ell=1}^k \eta_\ell(R_\ell), \quad (\text{B.142})$$

which concludes the proof.

B.7 Analysis of the coding scheme in Section 4.4.3.1

Consider the $\ell_{\mathcal{I}}^*$ -hop hypothesis testing scheme employed when $Y_0^n \in \mathcal{D}_{\mathcal{I}}$, for $\mathcal{I} \in \mathcal{P}(K)$. For any $\mathcal{I} \in \mathcal{P}(K)$, let $\hat{\mathcal{H}}_{\mathcal{I},1}, \dots, \hat{\mathcal{H}}_{\mathcal{I},\ell_{\mathcal{I}}^*}$ denote the guesses produced at terminals $1, \dots, \ell_{\mathcal{I}}^*$ when employing this scheme, where $\ell_{\mathcal{I}}^*$ is defined in (4.196).

By assumption, the type-I error probabilities of these decisions tend to 0 as $n \rightarrow \infty$ for any $\mathcal{I} \in \mathcal{P}(K)$:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\mathcal{I},k} = 1 | \mathcal{H} = 0, Y_0^n \in \mathcal{D}_{\mathcal{I}}] = 0, \quad k \in \mathcal{I}. \quad (\text{B.143})$$

Recalling that decision center k declares $\hat{\mathcal{H}}_k = 1$ whenever $Y_0^n \in \mathcal{D}_\emptyset$ or $Y_0^n \in \mathcal{D}_{\mathcal{I}}$ for a set \mathcal{I} not containing k , and applying the total law of probability, we can write

$$\alpha_{k,n} = \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0] \quad (\text{B.144})$$

$$= \sum_{\mathcal{I} \in (\mathcal{P}(K) \cup \emptyset)} \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 0] \quad (\text{B.145})$$

$$= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \notin \mathcal{I}}} \Pr[Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 0] + \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 0] \quad (\text{B.146})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \notin \mathcal{I}}} \Pr[Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 0] + \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \Pr[\hat{\mathcal{H}}_{\mathcal{I},k} = 1 | \mathcal{H} = 0, Y_0^n \in \mathcal{D}_{\mathcal{I}}] \quad (\text{B.147})$$

Combining this inequality with (B.143), and by Inequalities (4.195), we conclude that the overall scheme satisfies the type-I error constraints:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{k,n} \leq \epsilon_k, \quad k \in \{1, \dots, K\}. \quad (\text{B.148})$$

For the type-II error exponent at a decision center k we observe:

$$\beta_{k,n} = \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1] \quad (\text{B.149})$$

$$= \sum_{\mathcal{I} \in (\mathcal{P}(K) \cup \emptyset)} \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 1] \quad (\text{B.150})$$

$$= \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \Pr[\hat{\mathcal{H}}_{\mathcal{I},k} = 0, Y_0^n \in \mathcal{D}_{\mathcal{I}} | \mathcal{H} = 1] \quad (\text{B.151})$$

$$\leq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \Pr[\hat{\mathcal{H}}_{\mathcal{I},k} = 0 | \mathcal{H} = 1, Y_0^n \in \mathcal{D}_{\mathcal{I}}]. \quad (\text{B.152})$$

Defining

$$\theta_{k,\mathcal{I}} := \lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_{\mathcal{I},k} = 0 | \mathcal{H} = 1, Y_0^n \in \mathcal{D}_{\mathcal{I}}], \quad (\text{B.153})$$

we conclude by (B.152) that the exponent

$$\min_{\substack{\mathcal{I} \in \mathcal{P}(K): \\ k \in \mathcal{I}}} \theta_{k,\mathcal{I}} \quad (\text{B.154})$$

is achievable at decision center k . This proves in particular that when applying an instance of the multi-hop scheme in [34] for each set $\mathcal{I} \in \mathcal{P}(K)$, the exponents $\theta_1, \dots, \theta_K$ in (12) are proved achievable.

B.8 Proof of Lemma 6

Note first that by (4.219):

$$D(P_{\tilde{Y}_0^n \dots \tilde{Y}_K^n} \| P_{Y_0 \dots Y_K}^n) \leq \log \Delta_n^{-1}, \quad (\text{B.155})$$

where we defined $\Delta_n \triangleq P_{Y_0^n \dots Y_K^n}(\mathcal{D})$.

Further define $\tilde{U}_{i,t} \triangleq (\tilde{\mathbf{M}}_i, \tilde{Y}_0^{t-1}, \dots, \tilde{Y}_K^{t-1})$ for $i \in \{1, \dots, K\}$ and notice:

$$H(\tilde{\mathbf{M}}_i) \geq I(\tilde{\mathbf{M}}_i; \tilde{Y}_0^n \dots \tilde{Y}_K^n) + D(P_{\tilde{Y}_0^n \dots \tilde{Y}_K^n} \| P_{Y_0^n \dots Y_K^n}^n) + \log \Delta_n \quad (\text{B.156})$$

$$= H(\tilde{Y}_0^n \dots \tilde{Y}_K^n) + D(P_{\tilde{Y}_0^n \dots \tilde{Y}_K^n} \| P_{Y_0^n \dots Y_K^n}^n) - H(\tilde{Y}_0^n \dots \tilde{Y}_K^n | \tilde{\mathbf{M}}_i) + \log \Delta_n \quad (\text{B.157})$$

$$\geq n[H(\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T}) + D(P_{\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T}} \| P_{Y_0 \dots Y_K}^n)] - \sum_{t=1}^n H(\tilde{Y}_{0,t} \dots \tilde{Y}_{K,t} | \tilde{U}_{i,t}) + \log \Delta_n \quad (\text{B.158})$$

$$= n[H(\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T}) + D(P_{\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T}} \| P_{Y_0 \dots Y_K}^n) - H(\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T} | \tilde{U}_{i,T}, T)] + \log \Delta_n \quad (\text{B.159})$$

$$\geq n[H(\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T}) - H(\tilde{Y}_{0,T} \dots \tilde{Y}_{K,T} | \tilde{U}_{i,T}, T)] + \log \Delta_n \quad (\text{B.160})$$

$$= n[I(\tilde{Y}_0 \dots \tilde{Y}_K; U_i)] + \log \Delta_n \quad (\text{B.161})$$

$$\geq n \left[I(\tilde{Y}_{i-1}; U_i) + \frac{1}{n} \log \Delta_n \right]. \quad (\text{B.162})$$

Here, (B.156) holds by (B.155); (B.158) holds by the super-additivity property in [36, Proposition 1], by the chain rule, by the definition of $\tilde{U}_{i,t}$ and by defining T uniform over $\{1, \dots, n\}$ independent of the previously defined random variables; and (B.161) by the definitions of $U_i, \tilde{Y}_i, \tilde{Y}_{i-1}$ in the lemma. This proves Inequality

(4.222) in the lemma.

We next upper bound the type-II error exponent at R_k , for $k \in \{1, \dots, K\}$. To this end, define the acceptance region at R_k

$$\mathcal{A}_{Y_k} \triangleq \{(\mathbf{m}_k, y_k^n) : g_k(\mathbf{m}_k, y_k^n) = 0\}. \quad (\text{B.163})$$

Since for any tuple of sequences $(y_0^n, \dots, y_{K-1}^n)$, the corresponding pair (\mathbf{m}_k, y_k^n) either lies inside or outside the acceptance region \mathcal{A}_{Y_k} , for any $k \in \{1, \dots, K-1\}$, Condition (4.224) implies

$$P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_{Y_k}) = 1, \quad (\text{B.164})$$

Define for any $k \in \{1, \dots, K\}$:

$$Q_{\tilde{M}_k}(\mathbf{m}_k) \triangleq \sum_{y_0^n, y_1^n, \dots, y_{k-1}^n} P_{\tilde{Y}_0^n}(y_0^n) \cdots P_{\tilde{Y}_{k-1}^n}(y_{k-1}^n) \cdot \mathbb{1}\{\mathbf{m}_k = \phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n) \cdots)), y_{k-1}^n)\}, \quad (\text{B.165})$$

and

$$Q_{M_k}(\mathbf{m}_K) \triangleq \sum_{y_0^n, y_1^n, \dots, y_{k-1}^n} P_{Y_0^n}(y_0^n) \cdots P_{Y_{k-1}^n}(y_{k-1}^n) \cdot \mathbb{1}\{\mathbf{m}_k = \phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n) \cdots)), y_{k-1}^n)\}. \quad (\text{B.166})$$

We then have:

$$Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_{Y_k}) \leq Q_{M_k} P_{Y_K^n}(\mathcal{A}_{Y_K}) \Delta_n^{-(k+1)} \quad (\text{B.167})$$

$$= \beta_{k,n} \Delta_n^{-(k+1)}. \quad (\text{B.168})$$

By (B.164), (B.168), and standard inequalities (see [20, Lemma 1]), we further obtain

$$-\frac{1}{n} \log \beta_{k,n} \leq -\frac{1}{n} \log \left(Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_{Y_k}) \right) - \frac{(k+1)}{n} \log \Delta_n \quad (\text{B.169})$$

$$\leq \frac{1}{n} D(P_{\tilde{M}_k \tilde{Y}_k^n} \| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}) + \delta'_n, \quad (\text{B.170})$$

where $\delta'_n \triangleq -\frac{(k+1)}{n} \log \Delta_n + \frac{1}{n}$ and tends to 0 as $n \rightarrow \infty$.

We continue to upper bound the divergence term as

$$D(P_{\tilde{M}_k \tilde{Y}_k^n} \| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}) = I(\tilde{M}_k; \tilde{Y}_k^n) + D(P_{\tilde{M}_k} \| Q_{\tilde{M}_k}) \quad (\text{B.171})$$

$$\leq I(\tilde{M}_k; \tilde{Y}_k^n) + D(P_{\tilde{Y}_{k-1}^n, \tilde{M}_{k-1}} \| P_{\tilde{Y}_{k-1}^n} Q_{\tilde{M}_{k-1}}) \quad (\text{B.172})$$

$$\leq I(\tilde{M}_k; \tilde{Y}_k^n) + I(\tilde{M}_{k-1}; \tilde{Y}_{k-1}^n) + D(P_{\tilde{Y}_{k-2}^n, \tilde{M}_{k-2}} \| P_{\tilde{Y}_{k-2}^n} Q_{\tilde{M}_{k-2}}) \quad (\text{B.173})$$

\vdots

$$\leq \sum_{i=1}^k I(\tilde{M}_i; \tilde{Y}_i^n) \quad (\text{B.174})$$

$$= \sum_{i=1}^k \sum_{t=1}^n I(\tilde{M}_i; \tilde{Y}_{i,t} | \tilde{Y}_i^{t-1}) \quad (\text{B.175})$$

$$\leq \sum_{i=1}^k \sum_{t=1}^n I(\tilde{M}_i \tilde{Y}_0^{t-1} \cdots \tilde{Y}_K^{t-1}; \tilde{Y}_{i,t}) \quad (\text{B.176})$$

$$= \sum_{i=1}^k \sum_{t=1}^n I(\tilde{U}_{i,t}; \tilde{Y}_{i,t}) \quad (\text{B.177})$$

$$= \sum_{i=1}^k n I(\tilde{U}_{i,T}; \tilde{Y}_{i,T} | T) \quad (\text{B.178})$$

$$\leq n \sum_{i=1}^k I(U_i; \tilde{Y}_i). \quad (\text{B.179})$$

Here (B.172) is obtained by the data processing inequality for KL-divergence; (B.175) by the chain rule; and (B.177)–(B.179) by the definitions of $\tilde{U}_{i,t}$, U_i , \tilde{Y}_i and T .

Combined with (B.170) this establishes Inequality (4.225) for any $k \in \{1, \dots, K\}$. Finally, we proceed to prove that for any $k \in \{1, \dots, K\}$ the Markov chain $U_k \rightarrow \tilde{Y}_{k-1} \rightarrow \tilde{Y}_k$ holds in the limit as $n \rightarrow \infty$. We start by noticing the Markov chain $\tilde{M}_1 \rightarrow \tilde{Y}_0^n \rightarrow (\tilde{Y}_1^n, \dots, \tilde{Y}_K^n)$, and thus:

$$0 = I(\tilde{M}_1; \tilde{Y}_1^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n) \quad (\text{B.180})$$

$$\geq H(\tilde{Y}_1^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n) - H(\tilde{Y}_1^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \tilde{M}_1) + D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n}^n) + \log \Delta_n \quad (\text{B.181})$$

$$\geq n[H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}})] + \log \Delta_n - H(\tilde{Y}_1^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \tilde{M}_1) \quad (\text{B.182})$$

$$\geq n[H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{1,t} \cdots \tilde{Y}_{K,t} | \tilde{Y}_{0,t} \tilde{Y}_0^{t-1} \cdots \tilde{Y}_K^{t-1} \tilde{Y}_{0,t+1}^n \tilde{M}_1) \quad (\text{B.183})$$

$$= n[H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}})] + \log \Delta_n - nH(\tilde{Y}_{1,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T} \tilde{Y}_0^{T-1} \cdots \tilde{Y}_K^{T-1} \tilde{Y}_{0,T+1}^n \tilde{M}_1 T) \quad (\text{B.184})$$

$$\geq nI(\tilde{Y}_{1,T} \cdots \tilde{Y}_{K,T}; \tilde{Y}_0^{T-1} \cdots \tilde{Y}_K^{T-1} \tilde{Y}_{0,T+1}^n \tilde{M}_1 T | \tilde{Y}_{0,T}) + \log \Delta_n \quad (\text{B.185})$$

$$\geq nI(\tilde{Y}_1 \cdots \tilde{Y}_K; \tilde{U}_1 | Y_0) + \log \Delta_n, \quad (\text{B.186})$$

where (B.182) holds by the super-additivity property in [36, Proposition 1]; (B.183) by the chain rule; (B.185) by the non-negativity of the Kullback-Leibler divergence.

Since Δ_n is bounded, $\frac{1}{n} \log \Delta_n$ tends to 0 as $n \rightarrow \infty$, and we can conclude that

$$\lim_{n \rightarrow \infty} I(\tilde{Y}_1 \cdots \tilde{Y}_K; \tilde{U}_1 | \tilde{Y}_0) = 0, \quad (\text{B.187})$$

thus proving (4.223) for $k = 1$.

Notice next that for any $k \in \{2, \dots, K\}$:

$$I(U_k; \tilde{Y}_k | \tilde{Y}_{k-1}) \leq I(U_k \tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1}) \quad (\text{B.188})$$

$$= I(U_k; \tilde{Y}_k | \tilde{Y}_0 \cdots \tilde{Y}_{k-1}) + I(\tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1}). \quad (\text{B.189})$$

In the following we show that both quantities $I(U_k; \tilde{Y}_k | \tilde{Y}_0 \cdots \tilde{Y}_{k-1})$ and $I(\tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1})$ tend to 0 as $n \rightarrow \infty$, which establishes (4.223) for $k \in \{2, \dots, K\}$.

To prove that $I(\tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1})$ tends to 0 as $n \rightarrow \infty$, we notice that for any $k \in \{2, \dots, K\}$:

$$D(P_{\tilde{Y}_0 \cdots \tilde{Y}_K} \| P_{Y_0 \cdots Y_K}) \geq D(P_{\tilde{Y}_0 \cdots \tilde{Y}_k} \| P_{Y_0 \cdots Y_k}) \quad (\text{B.190})$$

$$= D(P_{\tilde{Y}_0 \cdots \tilde{Y}_k} \| P_{Y_0 \cdots Y_{k-1}} P_{Y_k | Y_{k-1}}) \quad (\text{B.191})$$

$$= D(P_{\tilde{Y}_0 \cdots \tilde{Y}_k} \| P_{\tilde{Y}_0 \cdots \tilde{Y}_{k-1}} P_{\tilde{Y}_k | \tilde{Y}_{k-1}}) + \mathbb{E}_{P_{\tilde{Y}_{k-1}}} \left[D(P_{\tilde{Y}_k | \tilde{Y}_{k-1}} \| P_{Y_k | Y_{k-1}}) \right] + D(P_{\tilde{Y}_0 \cdots \tilde{Y}_{k-1}} \| P_{Y_0 \cdots Y_{k-1}}) \quad (\text{B.192})$$

$$\geq D(P_{\tilde{Y}_0 \cdots \tilde{Y}_k} \| P_{\tilde{Y}_0 \cdots \tilde{Y}_{k-1}} P_{\tilde{Y}_k | \tilde{Y}_{k-1}}) \quad (\text{B.193})$$

$$= I(\tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1}). \quad (\text{B.194})$$

Since the tuple $(\tilde{Y}_0 \cdots \tilde{Y}_K)$ lies in the jointly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \cdots Y_K})$:

$$|P_{\tilde{Y}_0 \cdots \tilde{Y}_K} - P_{Y_0 \cdots Y_K}| \leq \mu_n. \quad (\text{B.195})$$

Recalling that $\mu_n \downarrow 0$ as $n \rightarrow \infty$, and by the continuity of the KL-divergence, we conclude that $D(P_{\tilde{Y}_0 \cdots \tilde{Y}_K} \| P_{Y_0 \cdots Y_K})$ tends to 0 as $n \rightarrow \infty$, and thus by (B.194) and the nonnegativity of mutual information:

$$\lim_{n \rightarrow \infty} I(\tilde{Y}_0 \cdots \tilde{Y}_{k-2}; \tilde{Y}_k | \tilde{Y}_{k-1}) = 0. \quad (\text{B.196})$$

Following similar steps to (B.180)–(B.185), we further obtain:

$$0 = I(\tilde{\mathbf{M}}_k; \tilde{Y}_k^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \cdots \tilde{Y}_{k-1}^n) \quad (\text{B.197})$$

$$\geq H(\tilde{Y}_k^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \cdots \tilde{Y}_{k-1}^n) - H(\tilde{Y}_k^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \cdots \tilde{Y}_{k-1}^n \tilde{\mathbf{M}}_k) + \log \Delta_n + D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{Y_0^n \cdots Y_K^n}) \quad (\text{B.198})$$

$$\geq n[H(\tilde{Y}_{k,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{k-1,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{Y_{0,T} \cdots Y_{K,T}})] + \log \Delta_n - H(\tilde{Y}_k^n \cdots \tilde{Y}_K^n | \tilde{Y}_0^n \cdots \tilde{Y}_{k-1}^n \tilde{\mathbf{M}}_k) \quad (\text{B.199})$$

$$\geq n[H(\tilde{Y}_{k,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{k-1,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{Y_{0,T} \cdots Y_{K,T}})] + \log \Delta_n - \sum_{t=1}^n H(\tilde{Y}_{k,t} \cdots \tilde{Y}_{K,t} | \tilde{Y}_{0,t} \cdots \tilde{Y}_{k-1,t} \tilde{Y}_0^{t-1} \cdots \tilde{Y}_K^{t-1} \tilde{Y}_{0,t+1}^n \cdots \tilde{Y}_{k-1,t+1}^n \tilde{\mathbf{M}}_k) \quad (\text{B.200})$$

$$\geq nH(\tilde{Y}_{k,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{k-1,T}) + \log \Delta_n - nH(\tilde{Y}_{k,T} \cdots \tilde{Y}_{K,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{k-1,T} \tilde{Y}_0^{T-1} \cdots \tilde{Y}_K^{T-1} \tilde{Y}_{0,T+1}^n \cdots \tilde{Y}_{k-1,T+1}^n \tilde{\mathbf{M}}_k T) \quad (\text{B.201})$$

$$\geq nI(\tilde{Y}_{k,T} \cdots \tilde{Y}_{K,T}; \tilde{Y}_0^{T-1} \cdots \tilde{Y}_K^{T-1} \tilde{M}_k T | \tilde{Y}_{0,T} \cdots \tilde{Y}_{k-1,T}) + \log \Delta_n \quad (\text{B.202})$$

$$= nI(\tilde{Y}_k \cdots \tilde{Y}_K; U_k | \tilde{Y}_0 \cdots \tilde{Y}_{k-1}) + \log \Delta_n. \quad (\text{B.203})$$

Since Δ_n is bounded, $\frac{1}{n} \log \Delta$ tends to 0 as $n \rightarrow \infty$, and we can conclude that

$$\lim_{n \rightarrow \infty} I(\tilde{Y}_k \cdots \tilde{Y}_K; \tilde{U}_k | Y_k - 1) = 0. \quad (\text{B.204})$$

Combined with (4.66), (B.196), and the nonnegativity of mutual information, this proves (4.223) for $k \in \{2, \dots, K\}$.

B.9 Proof of Lemma 7

To show sufficiency of (4.246), start by fixing any set of nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$, and $\{R_{\mathcal{I},1}, \dots, R_{\mathcal{I},\ell_{\mathcal{I}}^*}\}_{\mathcal{I} \in \mathcal{P}(3)}$ satisfying (4.202) for $K = 3$, (and possibly violating (4.246)). Choose new nonnegative numbers $\tilde{\sigma}_{\{1,2,3\}}, \tilde{\sigma}_{\{\pi(1),\pi(2)\}}, \tilde{\sigma}_{\{\pi(1),\pi(3)\}}, \tilde{\sigma}_{\{\pi(1)\}}$ satisfying

$$\tilde{\sigma}_{\mathcal{I}} \leq \sigma_{\mathcal{I}}, \quad \forall \mathcal{I}: \pi(1) \in \mathcal{I}, \quad (\text{B.205})$$

$$\tilde{\sigma}_{\{1,2,3\}} + \tilde{\sigma}_{\{\pi(1),\pi(2)\}} \geq 1 - \epsilon_{\pi(1)} - \epsilon_{\pi(2)} \quad (\text{B.206})$$

$$\tilde{\sigma}_{\{1,2,3\}} + \tilde{\sigma}_{\{\pi(1),\pi(3)\}} \geq 1 - \epsilon_{\pi(1)} - \epsilon_{\pi(3)} \quad (\text{B.207})$$

and

$$\tilde{\sigma}_{\{1,2,3\}} + \tilde{\sigma}_{\{\pi(1),\pi(2)\}} + \tilde{\sigma}_{\{\pi(1),\pi(3)\}} + \tilde{\sigma}_{\{\pi(1)\}} = 1 - \epsilon_{\pi(1)}. \quad (\text{B.208})$$

The existence of the desired numbers can be checked by applying the Fourier-Motzkin Elimination algorithm [52] and by noting Constraints (4.202). Further choose for any set \mathcal{I} containing $\pi(1)$ and $\ell \in \{1, \dots, \ell_{\mathcal{I}}^*\}$ the rate:

$$\tilde{R}_{\mathcal{I},\ell} := R_{\mathcal{I},\ell}, \quad (\text{B.209})$$

and for any set \mathcal{I} not containing $\pi(1)$ and $\ell \in \{1, \dots, \ell_{\mathcal{I}}^*\}$:

$$\tilde{\sigma}_{\mathcal{I}} := \sigma_{\mathcal{I}} + \sigma_{\mathcal{I}_{\pi(1)}} - \tilde{\sigma}_{\mathcal{I}_{\pi(1)}} \quad (\text{B.210})$$

$$\tilde{R}_{\mathcal{I},\ell} := \frac{\sigma_{\mathcal{I}}}{\tilde{\sigma}_{\mathcal{I}}} R_{\mathcal{I},\ell} + \frac{\sigma_{\mathcal{I}_{\pi(1)}} - \tilde{\sigma}_{\mathcal{I}_{\pi(1)}}}{\tilde{\sigma}_{\mathcal{I}}} R_{\mathcal{I}_{\pi(1)},\ell}, \quad (\text{B.211})$$

where we defined $\mathcal{I}_{\pi(1)} := \mathcal{I} \cup \{\pi(1)\}$.

By Lemma 5, the new set of numbers $\{\tilde{\sigma}_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$, and $\{\tilde{R}_{\mathcal{I},1}, \dots, \tilde{R}_{\mathcal{I},\ell_{\mathcal{I}}^*}\}_{\mathcal{I} \in \mathcal{P}(3)}$ also satisfies Constraints (4.202), which proves that one can restrict to numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$ satisfying (4.246). Since $\epsilon_{\pi(1)} \geq \epsilon_{\pi(2)}$

and

$$\sigma_{\{1,2,3\}} + \sigma_{\{\pi(1),\pi(2)\}} + \sigma_{\{\pi(2),\pi(3)\}} + \sigma_{\{\pi(2)\}} \geq 1 - \epsilon_{\pi(2)}, \quad (\text{B.212})$$

this further implies that one can restrict to numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I} \in \mathcal{P}(3)}$ satisfying

$$\sigma_{\{\pi(2),\pi(3)\}} \geq \sigma_{\{\pi(1),\pi(3)\}} + \sigma_{\{\pi(1)\}} - \sigma_{\{\pi(2)\}} \quad (\text{B.213})$$

$$\geq \sigma_{\{\pi(1),\pi(3)\}} - \sigma_{\{\pi(2)\}} - \sigma_{\{\pi(1),\pi(2)\}}. \quad (\text{B.214})$$

We next show that one can further restrict to nonnegative numbers satisfying also (4.247). To this end, assume that (4.247) is violated and define

$$a := \tilde{\sigma}_{\{\pi(1),\pi(3)\}} - \tilde{\sigma}_{\{\pi(2)\}} - \tilde{\sigma}_{\{\pi(1),\pi(2)\}} > 0. \quad (\text{B.215})$$

Define also the new parameters

$$\sigma'_{\{1,2,3\}} := \tilde{\sigma}_{\{1,2,3\}} + a \quad (\text{B.216})$$

$$\sigma'_{\{\pi(3)\}} := \tilde{\sigma}_{\{\pi(3)\}} + a \quad (\text{B.217})$$

$$\sigma'_{\{\pi(1),\pi(3)\}} := \tilde{\sigma}_{\{\pi(1),\pi(3)\}} - a \quad (\text{B.218})$$

$$\sigma'_{\{\pi(2),\pi(3)\}} := \tilde{\sigma}_{\{\pi(2),\pi(3)\}} - a \quad (\text{B.219})$$

$$\sigma'_{\mathcal{I}} := \tilde{\sigma}_{\mathcal{I}}, \quad \pi(3) \notin \mathcal{I}, \quad (\text{B.220})$$

and the new rates

$$R'_{\{1,2,3\},\ell} = \frac{a \left(\lambda_{\ell} \tilde{R}_{\{\pi(1),\pi(3)\},\ell} + (1 - \lambda_{\ell}) \tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right)}{\sigma'_{\{1,2,3\}}} + \frac{\tilde{\sigma}_{\{1,2,3\}} \tilde{R}_{\{1,2,3\},\ell}}{\sigma'_{\{1,2,3\}}}, \quad \ell \in \{1, 2, 3\}, \quad (\text{B.221a})$$

$$R'_{\{\pi(3)\},\ell} = \frac{a \left((1 - \lambda_{\ell}) \tilde{R}_{\{\pi(1),\pi(3)\},\ell} + \lambda_{\ell} \tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right)}{\sigma'_{\{\pi(3)\}}} + \frac{\tilde{\sigma}_{\{\pi(3)\}} \tilde{R}_{\{\pi(3)\},\ell}}{\sigma'_{\{\pi(3)\}}}, \quad \ell \in \{1, \dots, \pi(3)\} \quad (\text{B.221b})$$

$$R'_{\mathcal{I},\ell} = \tilde{R}_{\mathcal{I},\ell}, \quad \mathcal{I} \in \mathcal{P}(3) \setminus \{\{1, 2, 3\}, \{\pi(3)\}\}. \quad (\text{B.221c})$$

Notice that by the definition of a in (B.215) and by (B.214), the parameters $\{\sigma'_{\mathcal{I}}\}$ are all nonnegative, and it is easily verified that they continue to satisfy (4.202) for any choice of $\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$.

We next choose the parameters $\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$ in function of the rates $\{\tilde{R}_{\mathcal{I},\ell}\}$ and the ordering $\pi(\cdot)$, and show that for the proposed choice of rates in (B.221), the exponents $\theta_1, \theta_2, \theta_3$ are only increased. We distinguish three cases.

For notational simplicity we assume $\pi(1) < \pi(2)$. (The proof for $\pi(1) > \pi(2)$ is analogous.) This implies that

$$1 = \pi(1) < \pi(3) \quad \text{or} \quad 1 = \pi(3) < \pi(2) = 2 \quad (\text{B.222})$$

and

$$2 = \pi(2) < \pi(3) = 3 \quad \text{or} \quad \pi(3) < \pi(2) = 3. \quad (\text{B.223})$$

Case 1: If

$$\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) \leq \eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right), \quad (\text{B.224})$$

choose

$$\lambda_\ell = 0, \quad \ell \in \{1, \dots, \pi(3)\}, \quad (\text{B.225})$$

$$\lambda_\ell = \mathbb{1} \left\{ \tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \geq \tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right\}, \quad \ell \in \{\pi(3) + 1, \dots, 3\}. \quad (\text{B.226})$$

Using the same proof steps as in Lemma 5, it can be shown that for this choice of the λ s the new rates in (B.221) still satisfy Constraint (4.202a) for $\theta_{\pi(3)}$ because $\lambda_1 = \dots = \lambda_{\pi(3)} = 0$.

To see that they satisfy (4.202a) also for $\theta_{\pi(1)}$, notice that:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1, 2, 3\}}}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \quad (\text{B.227}) \end{aligned}$$

$$\begin{aligned} & \leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1, 2, 3\}}} \eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right) \right. \\ & \quad \left. + \frac{a}{\sigma'_{\{1, 2, 3\}}} \mathbb{1} \{ \pi(1) = 2 \} \cdot \max \left\{ \eta_2 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 2} \right), \eta_2 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 2} \right) \right\} \right. \\ & \quad \left. + \frac{\tilde{\sigma}_{\{1, 2, 3\}}}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \quad (\text{B.228}) \end{aligned}$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(R'_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_\ell \left(R'_{\{1, 2, 3\}, \ell} \right) \right\}. \quad (\text{B.229})$$

where the second inequality holds by Assumption (B.224) and the third inequality holds by the definitions of the rates $\{R'_{\{1, 2, 3\}, \ell}\}$ and by the concavity and monotonicity of the functions $\{\eta_\ell(\cdot)\}$.

Similarly, we notice for $\theta_{\pi(2)}$:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1, 2, 3\}}}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \quad (\text{B.230}) \end{aligned}$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_\ell \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\min\{\pi(2), \pi(3)\}} \eta_\ell \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) \right\}$$

$$\begin{aligned}
& + \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=\pi(3)+1}^{\pi(2)} \max \left\{ \eta_{\ell} \left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell} \right), \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right) \right\} \\
& \left. + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\},\ell} \right) \right\} \tag{B.231}
\end{aligned}$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{\pi(2),\pi(3)\},\ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{1,2,3\},\ell} \right) \right\}, \tag{B.232}$$

where notice that the sum in the second line of (B.231) is empty when $\pi(2) \leq \pi(3)$. Here, the last inequality holds by the definitions of the rates $\{R'_{\{1,2,3\},\ell}\}$ and by the choice of the λ s and the concavity and monotonicity of the functions $\{\eta_{\ell}(\cdot)\}_{\ell}$.

Case 2: If

$$\sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right) \leq \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell} \right), \tag{B.233}$$

choose

$$\lambda_{\ell} = 1, \quad \ell \in \{1, \dots, \max\{2, \pi(3)\}\}, \tag{B.234}$$

$$\lambda_{\ell} = \mathbb{1} \left\{ \tilde{R}_{\{\pi(1),\pi(3)\},\ell} \geq \tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right\}, \quad \ell \in \{\max\{2, \pi(3)\} + 1, \dots, 3\}. \tag{B.235}$$

Using similar arguments as in the previous case, one can conclude that the new rates in (B.221) still satisfy (4.202a). More specifically, since $\lambda_1 = \dots = \lambda_{\pi(3)} = 1$ by (B.234), similar proof steps as in Lemma 5 can be used to show that (4.202a) holds for $\theta_{\pi(3)}$.

To see that (4.202a) holds for $\theta_{\pi(2)}$, recall that $\pi(2) \geq 2$ and notice:

$$\begin{aligned}
& \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\},\ell} \right) \right\} \\
& \leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\},\ell} \right) \right\} \tag{B.236}
\end{aligned}$$

$$\begin{aligned}
& \leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell} \right) \right. \\
& \quad \left. + \frac{a}{\sigma'_{\{1,2,3\}}} \mathbb{1}\{\pi(2) = 3\} \cdot \max \left\{ \eta_3 \left(\tilde{R}_{\{\pi(1),\pi(3)\},3} \right), \eta_3 \left(\tilde{R}_{\{\pi(2),\pi(3)\},3} \right) \right\} \right. \\
& \quad \left. + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\},\ell} \right) \right\} \tag{B.237}
\end{aligned}$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{\pi(2),\pi(3)\},\ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{1,2,3\},\ell} \right) \right\}, \tag{B.238}$$

where the second inequality holds by our assumption (B.233) and since $\pi(2) \geq 2$.

Finally, (4.202a) holds for $\theta_{\pi(1)}$, because:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1, 2, 3\}}}{\sigma'_{\{1, 2, 3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1, 2, 3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.239})$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(R'_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(R'_{\{1, 2, 3\}, \ell} \right) \right\}. \quad (\text{B.240})$$

where the second inequality holds by the assumption $\pi(1) < \pi(2)$ and thus $\pi(1) \leq 2$.

Case 3: Else, i.e., if

$$\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) > \eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right) \quad (\text{B.241})$$

and

$$\sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) < \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right). \quad (\text{B.242})$$

Choose

$$\lambda_1 = 1, \quad \lambda_2 = \lambda, \quad \text{and} \quad \lambda_3 = 0, \quad (\text{B.243})$$

for a value of $\lambda \in [0, 1]$ so that the auxiliary rates

$$\bar{R}_{\{\pi(1), \pi(3)\}, 2} := \lambda R_{\{\pi(1), \pi(3)\}, 2} + (1 - \lambda) R_{\{\pi(2), \pi(3)\}, 2} \quad (\text{B.244})$$

$$\bar{R}_{\{\pi(2), \pi(3)\}, 2} := (1 - \lambda) R_{\{\pi(1), \pi(3)\}, 2} + \lambda R_{\{\pi(2), \pi(3)\}, 2} \quad (\text{B.245})$$

satisfy

$$\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) + \eta_2 \left(\bar{R}_{\{\pi(1), \pi(3)\}, 2} \right) = \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) \quad (\text{B.246})$$

$$\eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right) + \eta_2 \left(\bar{R}_{\{\pi(2), \pi(3)\}, 2} \right) \geq \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right). \quad (\text{B.247})$$

Existence of the desired choice of λ can be seen as follows. Notice first that for $\lambda = 0$, relation (B.246) holds with a “ $>$ ” sign because of Assumption (B.241). For $\lambda = 1$, relation (B.246) holds with a “ $<$ ” sign because of Assumption (B.242). By the continuity of the functions $\{\eta_{\ell}(\cdot)\}$ and the intermediate value theorem, there is thus a value $\lambda \in (0, 1)$ such that (B.246) holds with equality. Let λ be this value and notice that by the concavity of the functions $\{\eta_{\ell}(\cdot)\}$:

$$\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) + \eta_2 \left(\bar{R}_{\{\pi(1), \pi(3)\}, 2} \right) + \eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right) + \eta_2 \left(\bar{R}_{\{\pi(2), \pi(3)\}, 2} \right)$$

$$\geq \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) + \sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \quad (\text{B.248})$$

which combined with (B.246) implies (B.247).

Now that we established the existence of the desired value λ , we continue to show that for the choice in (B.243), Constraints (4.202a) remain valid. For $\theta_{\pi(1)}$ this can be verified through the following steps, where recall that $\pi(1) \leq 2$:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.249})$$

$$\begin{aligned} & = \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \left(\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) + \mathbb{1} \{ \pi(1) = 2 \} \cdot \eta_2 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 2} \right) \right) \right. \\ & \quad \left. + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.250})$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(R'_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(1)} \eta_{\ell} \left(R'_{\{1,2,3\}, \ell} \right) \right\}, \quad (\text{B.251})$$

where the second inequality holds since $\pi(1) \leq 2$, and by (B.242) and (B.246), and the last inequality holds by the definitions of the rates $\{R'_{\{1,2,3\}, \ell}\}$, the choice of the λ s, and the concavity and monotonicity of the functions $\{\eta_{\ell}(\cdot)\}$.

To verify that Constraint (4.202a) remains valid for $\theta_{\pi(2)}$, recall that $\pi(2) \geq 2$ and notice:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.252})$$

$$\begin{aligned} & = \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \left(\sum_{\ell=1}^2 \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) + \mathbb{1} \{ \pi(2) = 3 \} \cdot \eta_3 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 3} \right) \right) \right. \\ & \quad \left. + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.253})$$

$$\begin{aligned} & = \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \left(\eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) + \eta_2 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 2} \right) \right) \right. \\ & \quad \left. + \frac{a}{\sigma'_{\{1,2,3\}}} \mathbb{1} \{ \pi(2) = 3 \} \cdot \eta_3 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 3} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned}$$

(B.254)

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{\pi(2), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(2)} \eta_{\ell} \left(R'_{\{1,2,3\}, \ell} \right) \right\}. \quad (\text{B.255})$$

Here, the second equality holds by (B.246).

Finally, to see that Constraint (4.202a) is also satisfied for $\theta_{\pi(3)}$, we distinguish two cases. If $\pi(3) = 1$, the proof is similar to the proof of Lemma 5 because $\lambda_1 = 1$. For the proof in the case $\pi(3) \geq 2$, notice first:

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.256})$$

$$\begin{aligned} & = \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{1,2,3\}}} \eta_1 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 1} \right) + \frac{a}{\sigma'_{\{1,2,3\}}} \eta_2 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 2} \right) \right. \\ & \quad \left. + \frac{a}{\sigma'_{\{1,2,3\}}} \mathbb{1}\{\pi(3) = 3\} \cdot \eta_{\ell} \left(\tilde{R}_{\{\pi(2), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{1,2,3\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.257})$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(R'_{\{\pi(2), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(R'_{\{1,2,3\}, \ell} \right) \right\}, \quad (\text{B.258})$$

where the equality holds by Assumption (B.246).

Notice further that

$$\begin{aligned} & \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(3)\}, \ell} \right) \right\} \\ & \leq \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{\pi(3)\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right) + \frac{\tilde{\sigma}_{\{\pi(3)\}}}{\sigma'_{\{\pi(3)\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(3)\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.259})$$

$$\begin{aligned} & \leq \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(1), \pi(3)\}, \ell} \right), \frac{a}{\sigma'_{\{\pi(3)\}}} \eta_1 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 1} \right) + \frac{a}{\sigma'_{\{\pi(3)\}}} \eta_2 \left(\tilde{R}_{\{\pi(2), \pi(3)\}, 2} \right) \right. \\ & \quad \left. + \frac{a}{\sigma'_{\{\pi(3)\}}} \mathbb{1}\{\pi(3) = 3\} \cdot \eta_3 \left(\tilde{R}_{\{\pi(1), \pi(3)\}, 3} \right) + \frac{\tilde{\sigma}_{\{\pi(3)\}}}{\sigma'_{\{\pi(3)\}}} \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(\tilde{R}_{\{\pi(3)\}, \ell} \right) \right\} \end{aligned} \quad (\text{B.260})$$

$$\leq \min \left\{ \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(R'_{\{\pi(1), \pi(3)\}, \ell} \right), \sum_{\ell=1}^{\pi(3)} \eta_{\ell} \left(R'_{\{\pi(3)\}, \ell} \right) \right\}, \quad (\text{B.261})$$

where the second inequality holds by (B.247).

Appendix C

Proofs for Chapter 5

C.1 Analysis of the coding scheme in Subsection 5.5.1

Let $\hat{\mathcal{H}}_{\mathcal{I},i}$ denote the hypothesis guessed by R_i , for $i \in \{1, 2\}$, and $R_{\mathcal{I},i}$ the maximum rate of message \mathbf{M}_i , for $i \in \{0, 1, 2\}$, when the scheme in [28] is employed with auxiliaries $(U_{\mathcal{I},0}, U_{\mathcal{I},1}, U_{\mathcal{I},2})$, for $\mathcal{I} \in \mathcal{P}(2)$, to the present setup. By assumption, the type-I error probabilities of these schemes tend to 0 as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1 | \mathcal{H} = 0] = 0, \quad (\text{C.1a})$$

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1 | \mathcal{H} = 0] = 0, \quad (\text{C.1b})$$

$$\lim_{n \rightarrow \infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1, 2\}. \quad (\text{C.1c})$$

We can then write:

$$\alpha_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \quad (\text{C.2})$$

$$\begin{aligned} &= \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] \\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \end{aligned} \quad (\text{C.3})$$

$$\begin{aligned} &= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] \\ &\quad + \Pr[Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \end{aligned} \quad (\text{C.4})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset] + \Pr[Y_0^n \in \mathcal{D}_{\{2\}}] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1 | \mathcal{H} = 0] \quad (\text{C.5})$$

Combining this inequality with (C.1), and because in the limit $n \rightarrow \infty$ Inequalities (5.30) turn into equalities, we conclude by (5.28) that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{1,n} \leq 1 - \sigma_{\{1,2\}} - \sigma_{\{1\}} \leq \epsilon_1. \quad (\text{C.6})$$

Analogously, we have

$$\alpha_{2,n} = \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0] \quad (\text{C.7})$$

$$\begin{aligned} &= \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] \\ &\quad + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \end{aligned} \quad (\text{C.8})$$

$$\begin{aligned} &= \Pr[Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \\ &\quad + \Pr[Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \end{aligned} \quad (\text{C.9})$$

$$\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset] + \Pr[Y_0^n \in \mathcal{D}_{\{1\}}] + \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1 | \mathcal{H} = 0] \quad (\text{C.10})$$

Again combining this inequality with (C.1), and because in the limit $n \rightarrow \infty$ Inequalities (5.30) turn into equalities, we conclude by (5.28) that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{2,n} \leq 1 - \sigma_{\{1,2\}} - \sigma_{\{2\}} \leq \epsilon_2. \quad (\text{C.11})$$

For the type-II error probabilities we obtain

$$\beta_{1,n} = \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \quad (\text{C.12})$$

$$\begin{aligned} &= \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_\emptyset | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] \\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{2\}} | \mathcal{H} = 0] \end{aligned} \quad (\text{C.13})$$

$$= \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}} | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1, Y_0^n \in \mathcal{D}_{\{1\}} | \mathcal{H} = 0] \quad (\text{C.14})$$

$$\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1 | \mathcal{H} = 0] \quad (\text{C.15})$$

and analogously

$$\beta_{2,n} \leq \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1 | \mathcal{H} = 0] + \Pr[\hat{\mathcal{H}}_{\{2\},2} = 1 | \mathcal{H} = 0]. \quad (\text{C.16})$$

Taking logarithms, dividing by the blocklength n , and letting $n \rightarrow \infty$, we then obtain for $i \in \{1, 2\}$:

$$\underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{i,n} = \min \left\{ \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_{\{1,2\},i} = 1 | \mathcal{H} = 0], \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_{\{i\},i} = 1 | \mathcal{H} = 0] \right\} \quad (\text{C.17})$$

$$= \min \left\{ I(U_{\{1,2\},0} U_{\{1,2\},i}; Y_i), I(U_{\{i\},0} U_{\{i\},i}; Y_i) \right\}, \quad (\text{C.18})$$

where the last equality holds by the error exponents derived in [28] for each employment under the event $Y_0^n \in \mathcal{D}_{\mathcal{I}}$ ($\mathcal{I} \in \mathcal{P}(2)$) of the optimal scheme with vanishing type-I error probabilities and respecting maximum-rate constraints $(R_{\mathcal{I},0}, R_{\mathcal{I},1}, R_{\mathcal{I},2})$. Note that we set $R_{\{1\},2}$ and $R_{\{2\},1}$ to be approximately 0 (two bits do not impact the communication rate as the blocklength n grows largely).

In fact, the expected lengths of the messages are given by

$$\mathbb{E}[\text{len}(\mathbf{M}_0)] \leq 2 + \sum_{\mathcal{I} \in \mathcal{P}(2)} \sigma_{\mathcal{I}} \cdot nR_{\mathcal{I},0} \quad (\text{C.19})$$

and for $i \in \{1, 2\}$

$$\mathbb{E}[\text{len}(\mathbf{M}_i)] \leq \sigma_{\{1,2\}} nR_{\{1,2\},i} + \sigma_{\{i\}} nR_{\{i\},i}. \quad (\text{C.20})$$

Since the chosen random variables $U_{\{1,2\},0}, U_{\{1,2\},1}, U_{\{1,2\},2}, U_{\{1\},0}, U_{\{1\},1}, U_{\{2\},0}, U_{\{2\},2}$ and probabilities $\sigma_{\{1,2\}}, \sigma_{\{1\}}, \sigma_{\{2\}}$ satisfy Conditions (5.23)–(5.24), and since by [28], for $i \in \{1, 2\}$ and $\mathcal{I} \in \mathcal{P}(2)$

$$\tilde{R}_{\mathcal{I},0} = I(U_{\mathcal{I},0}; Y_0) + \mu, \quad (\text{C.21})$$

$$\tilde{R}_{\{1,2\},i} = I(U_{\{1,2\},i}; Y_0 | U_{\{1,2\},0}) + \mu, \quad (\text{C.22})$$

$$\tilde{R}_{\{i\},i} = I(U_{\{i\},i}; Y_0 | U_{\{i\},0}) + \mu, \quad (\text{C.23})$$

for an arbitrary small $\mu > 0$, we conclude that in the limit $n \rightarrow \infty$ and $\mu \downarrow 0$ the expected lengths of the messages satisfy the rate constraints (5.3).

C.2 Proof of Lemma 9

Throughout this section, let $h_b(\cdot)$ denote the binary entropy function, and $D(P\|Q)$ the Kullback-Leibler divergence between two pmfs on the same alphabet. Note first that by (5.41):

$$D(P_{\tilde{Y}_0^n} \| P_{Y_0^n}) \leq \log \Delta_n^{-1}, \quad (\text{C.24})$$

where we defined $\Delta_n \triangleq P_{Y_0^n}(\mathcal{D})$.

Further define $\tilde{U}_{0,t} \triangleq (\tilde{\mathbf{M}}_0, \tilde{Y}_0^{t-1})$ and $\tilde{U}_{1,t} \triangleq \tilde{\mathbf{M}}_1$, $\tilde{U}_{2,t} \triangleq \tilde{\mathbf{M}}_2$ and notice:

$$H(\tilde{\mathbf{M}}_0) \geq I(\tilde{\mathbf{M}}_0; \tilde{Y}_0^n) + D(P_{\tilde{Y}_0^n} \| P_{Y_0^n}) + \log \Delta_n \quad (\text{C.25})$$

$$= H(\tilde{Y}_0^n) + D(P_{\tilde{Y}_0^n} \| P_{Y_0^n}) - H(\tilde{Y}_0^n | \tilde{\mathbf{M}}_0) + \log \Delta_n \quad (\text{C.26})$$

$$\geq n[H(\tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T}} \| P_{Y_0})] - \sum_{t=1}^n H(\tilde{Y}_{0,t} | \tilde{U}_{0,t}) + \log \Delta_n \quad (\text{C.27})$$

$$= n[H(\tilde{Y}_{0,T}) + D(P_{\tilde{Y}_{0,T}} \| P_{Y_0}) - H(\tilde{Y}_{0,T} | \tilde{U}_{0,T}, T)] + \log \Delta_n \quad (\text{C.28})$$

$$\geq n[H(\tilde{Y}_{0,T}) - H(\tilde{Y}_{0,T} | \tilde{U}_{0,T}, T)] + \log \Delta_n \quad (\text{C.29})$$

$$= n[I(\tilde{Y}_0; U_0) + \frac{1}{n} \log \Delta_n]. \quad (\text{C.30})$$

Here, (C.25) holds by (C.24); (C.27) holds by the super-additivity property in [36, Proposition 1], by the chain rule, and by the definition of $\tilde{U}_{0,t}$; (C.28) by defining T uniform over $\{1, \dots, n\}$ independent of all

other random variables; and (C.30) by the definitions of U_0 and \tilde{Y}_0 in the lemma.

We lower bound the entropy of \tilde{M}_1 and \tilde{M}_2 for $i \in \{1, 2\}$:

$$H(\tilde{M}_i) \geq I(\tilde{M}_i; \tilde{Y}_0^n | \tilde{M}_0) \quad (\text{C.31})$$

$$\geq \sum_{t=1}^n I(\tilde{M}_i; \tilde{Y}_{0,t} | \tilde{M}_0 \tilde{Y}_0^{t-1}) \quad (\text{C.32})$$

$$= nI(U_i; \tilde{Y}_{0,T} | \tilde{U}_{0,T}, T) \quad (\text{C.33})$$

$$= nI(U_i; \tilde{Y}_0 | U_0) \quad (\text{C.34})$$

where (C.31) holds since conditioning can only reduce entropy and since \tilde{M}_i is a function of \tilde{Y}_0^n , and (C.33)–(C.34) hold by the definitions of $\tilde{U}_{0,T}$, U_1 , U_2 , \tilde{Y}_0 , and U_0 .

We next upper bound the error exponents at the decision centers. In the following, we note that the pair $(\mathbf{m}_0, \mathbf{m}_i)$ is always determined as a function of y_0^n .

Define for $i \in \{1, 2\}$

$$\mathcal{A}_{Y_i, n}(\mathbf{m}_0, \mathbf{m}_i) \triangleq \{y_i^n : g_i^{(n)}(\mathbf{m}_0, \mathbf{m}_i, y_i^n) = 0\}, \quad (\text{C.35})$$

and its Hamming neighborhood:

$$\hat{\mathcal{A}}_{Y_i, n}^{\ell_n}(\mathbf{m}_0, \mathbf{m}_i) \triangleq \{\tilde{y}_i^n : \exists y_i^n \in \mathcal{A}_{Y_i, n}(\mathbf{m}_0, \mathbf{m}_i) \text{ s.t. } d_H(y_i^n, \tilde{y}_i^n) \leq \ell_n\} \quad (\text{C.36})$$

for some real number ℓ_n satisfying $\lim_{n \rightarrow \infty} \ell_n/n = 0$ and $\lim_{n \rightarrow \infty} \ell_n/\sqrt{n} = \infty$, and where $d_H(\cdot, \cdot)$ represents the hamming distance between two sequences. Since by Condition (5.44),

$$P_{\tilde{Y}_i^n | \tilde{Y}_0^n}(\mathcal{A}_{Y_i, n}(\mathbf{m}_0, \mathbf{m}_i) | y_0^n) \geq \eta, \quad \forall y_0^n \in \mathcal{D}, \quad (\text{C.37})$$

by the blowing-up lemma [40]:

$$P_{\tilde{Y}_i^n | \tilde{Y}_0^n}(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n}(\mathbf{m}_0, \mathbf{m}_i) | y_0^n) \geq 1 - \zeta_n, \quad \forall y_0^n \in \mathcal{D}, \quad (\text{C.38})$$

for a real number $\zeta_n > 0$ such that $\lim_{n \rightarrow \infty} \zeta_n = 0$.

Define

$$\mathcal{A}_{Y_i, n} \triangleq \bigcup_{(\mathbf{m}_0, \mathbf{m}_i) \in \mathcal{M}_0 \times \mathcal{M}_i} \{\mathbf{m}_0, \mathbf{m}_i\} \times \mathcal{A}_{Y_i, n}(\mathbf{m}_0, \mathbf{m}_i) \quad (\text{C.39})$$

and

$$\hat{\mathcal{A}}_{Y_i, n}^{\ell_n} \triangleq \bigcup_{(\mathbf{m}_0, \mathbf{m}_i) \in \mathcal{M}_0 \times \mathcal{M}_i} \{\mathbf{m}_0, \mathbf{m}_i\} \times \hat{\mathcal{A}}_{Y_i, n}^{\ell_n}(\mathbf{m}_0, \mathbf{m}_i), \quad (\text{C.40})$$

and notice that

$$P_{\tilde{M}_0 \tilde{M}_i \tilde{Y}_i^n}(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n}) = \sum_{y_0^n \in \mathcal{D}} P_{\tilde{Y}_0^n}(y_0^n) \cdot P_{\tilde{Y}_i^n | \tilde{Y}_0^n}(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n}(\mathbf{m}_0, \mathbf{m}_i) | y_0^n) \quad (\text{C.41})$$

$$\geq (1 - \zeta_n). \quad (\text{C.42})$$

Finally, we can write

$$P_{\tilde{M}_0 \tilde{M}_i} P_{\tilde{Y}_i^n} \left(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n} \right) \leq P_{M_0 M_i} P_{Y_i}^n \left(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n} \right) \Delta_n^{-2} \quad (\text{C.43})$$

$$= \sum_{\substack{(\mathbf{m}_0, \mathbf{m}_i) \in \\ \mathcal{M}_0 \times \mathcal{M}_i}} P_{M_0 M_i}(\mathbf{m}_0, \mathbf{m}_i) P_{Y_i}^n \left(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n}(\mathbf{m}_0, \mathbf{m}_i) \right) \Delta_n^{-2} \quad (\text{C.44})$$

$$\leq \sum_{\substack{(\mathbf{m}_0, \mathbf{m}_i) \in \\ \mathcal{M}_0 \times \mathcal{M}_i}} P_{M_0 M_i}(\mathbf{m}_0, \mathbf{m}_i) P_{Y_i}^n(\mathcal{A}_{Y_i, n}(\mathbf{m}_0, \mathbf{m}_i)) \cdot e^{nh_b(\ell_n/n)} |\mathcal{Y}_i|^{\ell_n} k_n^{\ell_n} \Delta_n^{-2} \quad (\text{C.45})$$

$$= \beta_{i, n} e^{n\delta_n}, \quad (\text{C.46})$$

where $\delta_n \triangleq h_b(\ell_n/n) + \frac{\ell_n}{n} \log(|\mathcal{Y}_i| \cdot k_n) - \frac{2}{n} \log \Delta_n$ and $k_n \triangleq \min_{\substack{y_i, y'_i: \\ P_{Y_i}(y'_i) > 0}} \frac{P_{Y_i}(y_i)}{P_{Y_i}(y'_i)}$. Here, (C.46) holds by [44, Proof of Lemma 5.1].

Combining (C.46) with (C.42) and standard inequalities (see [20, Lemma 1]), we then obtain:

$$-\frac{1}{n} \log \beta_{i, n} \leq -\frac{1}{n} \log \left(P_{\tilde{M}_0 \tilde{M}_i} P_{\tilde{Y}_i^n} \left(\hat{\mathcal{A}}_{Y_i, n}^{\ell_n} \right) \right) + \delta_n \quad (\text{C.47})$$

$$\leq \frac{1}{n(1 - \zeta_n)} D(P_{\tilde{M}_0 \tilde{M}_i \tilde{Y}_i^n} \| P_{\tilde{M}_0 \tilde{M}_i} P_{\tilde{Y}_i^n}) + \delta_n + \frac{1}{n}, \quad (\text{C.48})$$

where ζ_n and δ_n both tend to 0 as $n \rightarrow \infty$. We continue to upper bound the divergence term as

$$D(P_{\tilde{M}_0 \tilde{M}_i \tilde{Y}_i^n} \| P_{\tilde{M}_0 \tilde{M}_i} P_{\tilde{Y}_i^n}) = I(\tilde{M}_0 \tilde{M}_i; \tilde{Y}_i^n) \quad (\text{C.49})$$

$$= \sum_{t=1}^n I(\tilde{M}_0 \tilde{M}_i; \tilde{Y}_{i, t} | \tilde{Y}_i^{t-1}) \quad (\text{C.50})$$

$$\leq \sum_{t=1}^n I(\tilde{M}_0 \tilde{M}_i \tilde{Y}_0^{t-1}; \tilde{Y}_{i, t}) \quad (\text{C.51})$$

$$= n[I(\tilde{U}_{0, T} U_i; \tilde{Y}_{i, T} | T)] \quad (\text{C.52})$$

$$\leq n[I(\tilde{U}_{0, T} T U_i; \tilde{Y}_{i, T})] \quad (\text{C.53})$$

$$= n[I(U_0 U_i; \tilde{Y}_i)]. \quad (\text{C.54})$$

Here, (C.50) holds by the chain rule; (C.51) by the Markov chain $\tilde{Y}_i^{t-1} \rightarrow (\tilde{Y}_0^{t-1} \tilde{M}_0 \tilde{M}_i) \rightarrow \tilde{Y}_{i, t}$; and (C.52)–(C.54) by the definitions of $T, \tilde{U}_{0, t}, U_0, U_i, \tilde{Y}_i$.

C.3 Strong Converse Proof to Theorem 15

Fix an exponent pair in $\mathcal{E}_{\max}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ and a sequence (in n) of encoding and decision functions $\{(\phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ satisfying the constraints on the rate and the error probabilities in (5.7a), (5.7b), (5.7c).

Fix a positive $\eta > 0$ and a blocklength n and choose $\mu_n = n^{-2/3}$. Define for $i \in \{1, 2\}$, the sets

$$\mathcal{B}_i(\eta) \triangleq \{y_0^n \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0}) : \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \geq \eta\}, \quad i \in \{1, 2\}, \quad (\text{C.55})$$

$$\mathcal{D}(\eta) \triangleq \mathcal{B}_1(\eta) \cap \mathcal{B}_2(\eta). \quad (\text{C.56})$$

Further define the probability

$$\Delta \triangleq P_{Y_0^n}(\mathcal{D}(\eta)), \quad (\text{C.57})$$

and notice that by the laws of probability

$$\Delta \geq \max\{P_{Y_0^n}(\mathcal{B}_1(\eta)) + P_{Y_0^n}(\mathcal{B}_2(\eta)) - 1, 0\}. \quad (\text{C.58})$$

By (5.7a), it can further be shown that for $i \in \{1, 2\}$:

$$1 - \epsilon_i \leq \sum_{y_0^n} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \cdot P_{Y_0^n}(y_0^n) \quad (\text{C.59})$$

$$\leq \sum_{y_0^n \in \mathcal{T}_{\mu_n}^{(n)}} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \cdot P_{Y_0^n}(y_0^n) + \sum_{y_0^n \notin \mathcal{T}_{\mu_n}^{(n)}} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \cdot P_{Y_0^n}(y_0^n) \quad (\text{C.60})$$

$$\leq \sum_{y_0^n \in \mathcal{T}_{\mu_n}^{(n)} \cap \bar{\mathcal{B}}_i(\eta)} \Pr[\hat{\mathcal{H}}_i = 0 | Y_0^n = y_0^n, \mathcal{H} = 0] \cdot P_{Y_0^n}(y_0^n) + \sum_{y_0^n \in \mathcal{B}_i(\eta)} P_{Y_0^n}(y_0^n) + (1 - P_{Y_0^n}(\mathcal{T}_{\mu_n}^{(n)})) \quad (\text{C.61})$$

$$\leq \eta(1 - P_{Y_0^n}(\mathcal{B}_i(\eta))) + P_{Y_0^n}(\mathcal{B}_i(\eta)) + P_{Y_0^n}(\bar{\mathcal{T}}_{\mu_n}^{(n)}). \quad (\text{C.62})$$

Thus, by (C.62) and [44, Lemma 2.12]:

$$\Delta \geq \max\left\{\frac{1 - \epsilon_1 - \epsilon_2 - \eta}{1 - \eta} - \frac{|\mathcal{Y}_0|}{(1 - \eta)\mu_n n}, 0\right\}, \quad (\text{C.63})$$

and we conclude that in the limit as $n \rightarrow \infty$ and $\eta \downarrow 0$:

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}. \quad (\text{C.64a})$$

We proceed by applying Lemma 9 to the set $\mathcal{D}(\eta)$ with positive probability $\Delta > 0$. By initial condition, $\epsilon_1 + \epsilon_2 < 1$, and thus for $\eta > 0$ sufficiently small and n sufficiently large, by (C.64) Δ is positive and we can apply Lemma 4 to the set $\mathcal{D}(\eta)$. By this Lemma 9, and using also the maximum-rate constraints (5.2), alongside the trivial inequality $nR_i \geq H(\tilde{\mathbf{M}}_i)$, for all $i \in \{0, 1, 2\}$, we conclude that there is a tuple (U_0, U_1, U_2) satisfying

$$nR_0 \geq H(\tilde{\mathbf{M}}_0) \geq nI(U_0; \tilde{Y}_0) + \log P_{Y_0^n}(\mathcal{D}(\eta)), \quad (\text{C.65})$$

$$nR_i \geq H(\tilde{\mathbf{M}}_i) \geq nI(U_i; \tilde{Y}_0 | U_0), \quad i \in \{1, 2\}, \quad (\text{C.66})$$

and for $i \in \{1, 2\}$:

$$-\frac{1}{n} \log \beta_{i,n} \leq I(U_0 U_i; \tilde{Y}_i) + \phi_i(n), \quad (\text{C.67})$$

where for each i , the function $\phi_i(n) \downarrow 0$ as $n \rightarrow \infty$ and the random variables $\tilde{Y}_0, \tilde{Y}_i, \tilde{M}_0, \tilde{M}_i$ are defined as in the lemma applied to the set $\mathcal{D}(\eta)$.

Thus we have proved so far that for all $n \geq 1$ there exists joint pmf $P_{U_0 U_1 U_2 \tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2} = P_{\tilde{Y}_0} P_{\tilde{Y}_1 \tilde{Y}_2 | \tilde{Y}_0} P_{U_0 U_1 U_2 | \tilde{Y}_0}$ (abbreviated as $P^{(n)}$) so that the following conditions hold for $i \in \{1, 2\}$

$$R_0 \geq I_{P^{(n)}}(U_0; \tilde{Y}_0) + g_1(n), \quad (\text{C.68a})$$

$$R_i \geq I_{P^{(n)}}(U_i; \tilde{Y}_0 | U_0), \quad (\text{C.68b})$$

$$\theta_i \leq I_{P^{(n)}}(U_0 U_i; \tilde{Y}_i) + g_{2,i}(n), \quad (\text{C.68c})$$

for some nonnegative functions $g_1(n), g_{2,i}(n)$ with the following asymptotic behaviors:

$$\lim_{n \rightarrow \infty} g_1(n) = 0, \quad (\text{C.69})$$

$$\lim_{n \rightarrow \infty} g_{2,i}(n) = 0, \quad \forall i \in \{1, 2\}. \quad (\text{C.70})$$

The rest of the proof follows the same steps as the proof of the converse in Section 5.7. By Carathéodory's theorem [48, Appendix C], there exist for each n random variables U_0, U_1, U_2 satisfying (C.68) over alphabets of sizes

$$|\mathcal{U}_0| \leq |\mathcal{Y}_0| + 3, \quad (\text{C.71})$$

$$|\mathcal{U}_i| \leq |\mathcal{U}_0| \cdot |\mathcal{Y}_0| + 1, \quad i \in \{1, 2\}. \quad (\text{C.72})$$

Invoke the Bolzano-Weierstrass theorem and consider a sub-sequence $P_{U_0 U_1 U_2 \tilde{Y}_0 \tilde{Y}_1 \tilde{Y}_2}^{(n_k)}$ that converges to a limiting pmf $P_{U_0 U_1 U_2 Y_0 Y_1 Y_2}^*$. For this limiting pmf, which we abbreviate by P^* , we conclude by (C.68a)–(C.68c) that for all $i \in \{1, 2\}$:

$$R_0 \geq I_{P^*}(U_0; Y_0), \quad (\text{C.73})$$

$$R_i \geq I_{P^*}(U_i; Y_0 | U_0), \quad (\text{C.74})$$

$$\theta_i \leq I_{P^*}(U_0 U_i; Y_i). \quad (\text{C.75})$$

Notice further that since for any k , the sequence $\tilde{Y}_0^{n_k}$ lies in the typical set $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{Y_0})$, we have $|P_{\tilde{Y}_0}^{(n_k)} - P_{Y_0}| \leq \mu_{n_k}$ and thus the limiting pmf satisfies $P_{\tilde{Y}_0}^* = P_{Y_0}$. Moreover, since for each n_k the pair of random variables $(\tilde{Y}_1, \tilde{Y}_2)$ is drawn according to $P_{Y_1 Y_2 | Y_0}$ given \tilde{Y}_0 , the limiting pmf also satisfies $P_{\tilde{Y}_1 \tilde{Y}_2 | Y_0}^* = P_{Y_1 Y_2 | Y_0}$. We also notice that under P^* the Markov chain $(U_0, U_1, U_2) \rightarrow Y_0 \rightarrow (Y_1, Y_2)$ holds.

This concludes the strong converse proof.

C.4 Proof of Corollary 1

By Theorem 16, $\mathcal{E}^*(R_0, 0, 0, \epsilon_1, \epsilon_2)$ is the set of all (θ_1, θ_2) pairs satisfying

$$\theta_i \leq \min \{I(U_0; Y_i), \eta_i(R_{\{i\},0})\}, \quad i \in \{1, 2\}. \quad (\text{C.76a})$$

for some nonnegative numbers $\sigma_{\{1,2\}}, \sigma_{\{1\}}, \sigma_{\{2\}}$, with sum ≤ 1 and satisfying (5.35d) and (5.35e), a conditional pmf $P_{U_0|Y_0}$, and nonnegative rates $R_{\{1\},0}, R_{\{2\},0}$ such that

$$R_0 \geq \sigma_{\{1,2\}}I(U_0; Y_0) + \sigma_{\{1\}}R_{\{1\},0} + \sigma_{\{2\}}R_{\{2\},0}. \quad (\text{C.76b})$$

Notice that without loss in optimality, in the evaluation of above region, we can restrict to tuples $(P_{U_0|Y_0}, R_{\{1\},0}, R_{\{2\},0})$ satisfying

$$I(U_0; Y_i) \geq \eta_i(R_{\{i\},0}), \quad (\text{C.77})$$

which by the maximum in the definition and by the non-decreasing monotonicity of function $\eta_i(\cdot)$ implies

$$I(U_0; Y_0) \geq R_{\{i\},0}, \quad i \in \{1, 2\}. \quad (\text{C.78})$$

In fact, if (C.77) is violated, rates $R_{\{1\},0}$ and/or $R_{\{2\},0}$ can be reduced without changing (C.76a) and so that (C.77) holds.

We next show that any exponent pair (θ_1, θ_2) and tuple $(P_{U_0|Y_0}, R_{\{1\},0}, R_{\{2\},0})$ satisfying (C.76), (C.77), and

$$I(U_0; Y_0) \leq R_{\{\pi(1)\},0} + R_{\{\pi(2)\},0} \quad (\text{C.79})$$

also satisfies (5.40). The constraints on the exponents (5.40a) and (5.40b) are easily verified. To verify (5.40c), notice that when $\sigma_{\{1,2\}} > 1 - \epsilon_{\pi(1)}$:

$$R_0 \geq \sigma_{\{1,2\}}I(U_0; Y_0) + \sigma_{\{\pi(1)\}}R_{\{\pi(1)\},0} + \sigma_{\{\pi(2)\}}R_{\{\pi(2)\},0} \quad (\text{C.80})$$

$$= (1 - \epsilon_{\pi(1)})I(U_0; Y_0) + \sigma_{\{\pi(1)\}}R_{\{\pi(1)\},0} + (\sigma_{\{1,2\}} - 1 + \epsilon_{\pi(1)})I(U_0; Y_0) + \sigma_{\{\pi(2)\}}R_{\{\pi(2)\},0} \quad (\text{C.81})$$

$$\geq (1 - \epsilon_{\pi(1)})I(U_0; Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})R_{\{\pi(2)\},0} \quad (\text{C.82})$$

where (C.82) holds since $\sigma_{\{\pi(1)\}}R_{\{\pi(1)\},0} \geq 0$, $I(U_0; Y_0) \geq R_{\{\pi(2)\},0}$ by (C.78), and $\sigma_{\{1,2\}} + \sigma_{\{\pi(2)\}} \geq 1 - \epsilon_{\pi(2)}$ by (5.35d).

For $\sigma_{\{1,2\}} \leq 1 - \epsilon_{\pi(1)}$, the rate constraint (5.40c) can be verified as follows:

$$R_0 \geq \sigma_{\{1,2\}}I(U_0; Y_0) + \sigma_{\{\pi(1)\}}R_{\{\pi(1)\},0} + \sigma_{\{\pi(2)\}}R_{\{\pi(2)\},0} \quad (\text{C.83})$$

$$\geq \sigma_{\{1,2\}}I(U_0; Y_0) + (1 - \epsilon_{\pi(1)} - \sigma_{\{1,2\}})R_{\{\pi(1)\},0} + \sigma_{\{\pi(2)\}}R_{\{\pi(2)\},0} \quad (\text{C.84})$$

$$\geq \sigma_{\{1,2\}}I(U_0; Y_0) + (1 - \epsilon_{\pi(1)} - \sigma_{\{1,2\}})(R_{\{\pi(1)\},0} + R_{\{\pi(2)\},0}) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})R_{\{\pi(2)\},0} \quad (\text{C.85})$$

$$\geq (1 - \epsilon_{\pi(1)})I(U_0; Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})R_{\{\pi(2)\},0} \quad (\text{C.86})$$

where (C.84) holds by (5.35d), indicating $\sigma_{\{\pi(1)\}} \geq 1 - \epsilon_{\pi(1)} - \sigma_{\{1,2\}}$ (C.85) holds because $\sigma_{\{\pi(2)\}} \geq 1 - \epsilon_{\pi(2)} - \sigma_{\{1,2\}}$ by (5.35d), and (C.86) holds by (C.79) and by $\sigma_{\{1,2\}} \leq 1 - \epsilon_{\pi(1)}$. This establishes that (C.76) holds under condition (C.79).

The proof is concluded by showing that for any tuple $(\theta_1, \theta_2, P_{U_0|Y_0}, R_{\{1\},0}, R_{\{2\},0})$ satisfying (C.76), (C.77), and

$$I(U_0; Y_0) > R_{\{\pi(1)\},0} + R_{\{\pi(2)\},0}, \quad (\text{C.87})$$

we can find a pmf $P_{\tilde{U}_0|Y_0}$, satisfying (5.40) when U_0 is replaced by \tilde{U}_0 .

To this end, choose a bivariate $\tilde{U}_0 = (\tilde{U}_{\{1\},0}, \tilde{U}_{\{2\},0})$ such that $\tilde{U}_{\{1\},0} \rightarrow Y_0 \rightarrow \tilde{U}_{\{2\},0}$ forms a Markov chain and for each $i \in \{1, 2\}$ the new random-variable $\tilde{U}_{\{i\},0}$ achieves $\eta_i(R_{\{i\},0})$, i.e.,

$$R_{\{i\},0} = I(\tilde{U}_{\{i\},0}; Y_0) \quad \text{and} \quad \eta_i(R_{\{i\},0}) = I(\tilde{U}_{\{i\},0}; Y_i). \quad (\text{C.88})$$

Since for any $i \in \{1, 2\}$ we have $I(\tilde{U}_0; Y_i) \geq I(\tilde{U}_{\{i\},0}; Y_i) = \eta_i(R_{\{i\},0})$, the exponents satisfy

$$\theta_{\pi(1)} \leq \min \{I(U_0; Y_{\pi(1)}), \eta_{\pi(1)}(R_{\{\pi(1)\},0})\} = \eta_{\pi(1)}(R_{\{\pi(1)\},0}) \leq I(\tilde{U}_0; Y_{\pi(1)}) \quad (\text{C.89})$$

$$\theta_{\pi(2)} \leq \min \{I(U_0; Y_{\pi(2)}), \eta_{\pi(2)}(R_{\{\pi(2)\},0})\} = \eta_{\pi(2)}(R_{\{\pi(2)\},0}) \quad (\text{C.90})$$

$$= \min \{I(\tilde{U}_0; Y_{\pi(2)}), \eta_{\pi(2)}(R_{\{\pi(2)\},0})\}, \quad (\text{C.91})$$

where the equalities in (C.89) and (C.90) hold by (C.77). Similarly,

$$R_0 \geq \sigma_{\{1,2\}}I(U_0; Y_0) + \sigma_{\{\pi(1)\}}R_{\{\pi(1)\},0} + \sigma_{\{\pi(2)\}}R_{\{\pi(2)\},0} \quad (\text{C.92})$$

$$> (1 - \epsilon_{\pi(1)})R_{\{\pi(1)\},0} + (1 - \epsilon_{\pi(2)})R_{\{\pi(2)\},0} \quad (\text{C.93})$$

$$= (1 - \epsilon_{\pi(1)})I(\tilde{U}_{\{\pi(1)\},0}; Y_0) + (1 - \epsilon_{\pi(2)})I(\tilde{U}_{\{\pi(2)\},0}; Y_0) \quad (\text{C.94})$$

$$\geq (1 - \epsilon_{\pi(1)})I(\tilde{U}_{\{\pi(1)\},0}; Y_0) + (1 - \epsilon_{\pi(1)})I(\tilde{U}_{\{\pi(2)\},0}; Y_0 | \tilde{U}_{\{\pi(1)\},0}) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})I(\tilde{U}_{\{\pi(2)\},0}; Y_0) \quad (\text{C.95})$$

$$= (1 - \epsilon_{\pi(1)})I(\tilde{U}_0; Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)})I(\tilde{U}_{\{\pi(2)\},0}; Y_0) \quad (\text{C.96})$$

where inequality (C.93) holds by the assumption that $I(U_0; Y_0) > R_{\{1\},0} + R_{\{2\},0}$ and by condition (5.35d); equality (C.94) holds by (C.88); inequality (C.95) holds by the Markov chain $\tilde{U}_{\{1\},0} \rightarrow Y_0 \rightarrow \tilde{U}_{\{2\},0}$; and (C.96) holds by the chain rule and by the definition of \tilde{U}_0 .

Bibliography

- [1] R. Ahlswede and I. Csiszár, “Hypothesis testing with communication constraints,” *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.
- [2] T. S. Han, “Hypothesis testing with multiterminal data compression,” *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.
- [3] T. Han and K. Kobayashi, “Exponential-type error probabilities for multiterminal hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.
- [4] H. Shalaby and A. Papamarcou, “Multiterminal detection with zero-rate data compression,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.
- [5] H. Shimokawa, T. Han, and S. I. Amari, “Error bound for hypothesis testing with data compression,” in *Proc. ISIT*, p. 114, Jul. 1994.
- [6] M. S. Rahman and A. B. Wagner, “On the optimality of binning for distributed hypothesis testing,” *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.
- [7] Y. Xiang and Y.-H. Kim, “Interactive hypothesis testing with communication constraints,” in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1065–1072, IEEE, 2012.
- [8] Y. Xiang and Y. H. Kim, “Interactive hypothesis testing against independence,” in *Proc. ISIT*, pp. 2840–2844, Jun. 2013.
- [9] E. Haim and Y. Kochman, “Binary distributed hypothesis testing via korner-marton coding,” in *Proc. IEEE Info. Theory Work. (ITW)*, 2016.

-
- [10] G. Katz, P. Piantanida, and M. Debbah, “Collaborative distributed hypothesis testing with general hypotheses,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1705–1709, IEEE, 2016.
- [11] G. Katz, P. Piantanida, and M. Debbah, “Distributed binary detection with lossy data compression,” *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5207–5227, 2017.
- [12] S. Sreekumar and D. Gündüz, “Distributed hypothesis testing over noisy channels,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 983–987, 2017.
- [13] S. Sreekumar, D. Gündüz, and A. Cohen, “Distributed hypothesis testing under privacy constraints,” in *2018 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2018.
- [14] S. Sreekumar and D. Gündüz, “Hypothesis testing over a noisy channel,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 2004–2008, 2019.
- [15] N. Weinberger, Y. Kochman, and M. Wigger, “Exponent trade-off for hypothesis testing over noisy channels,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1852–1856, 2019.
- [16] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. F. Tan, “Distributed hypothesis testing with privacy constraints,” *Entropy*, vol. 21, p. 478, May 2019.
- [17] S. Sreekumar and D. Gündüz, “Strong converse for testing against independence over a noisy channel,” in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1283–1288, 2020.
- [18] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing based on unequal-error protection codes,” *IEEE Trans. Inf. Theory*, vol. 66, pp. 4150–41820, Jul. 2020.
- [19] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing with variable-length coding,” in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–5, 2020.
- [20] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing with variable-length coding,” *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 681–694, 2020.
- [21] S. Salehkalaibar and V. Y. Tan, “Distributed sequential hypothesis testing with zero-rate compression,” in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–5, IEEE, 2021.
-

-
- [22] S. Watanabe, “On sub-optimality of random binning for distributed hypothesis testing.” [Online]. Available: <https://arxiv.org/abs/2201.13005>, 2022.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.
- [24] W. Zhao and L. Lai, “Distributed detection with vector quantizer,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 105–119, 2016.
- [25] W. Zhao and L. Lai, “Distributed testing with cascaded encoders,” *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.
- [26] S. Sreekumar and D. Gündüz, “Distributed hypothesis testing over discrete memoryless channels,” *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2044–2066, 2020.
- [27] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing over multi-access channels,” in *2018 Information Theory and Applications Workshop (ITA)*, pp. 1–5, 2018.
- [28] M. Wigger and R. Timo, “Testing against independence with multiple decision centers,” in *2016 International Conference on Signal Processing and Communications (SPCOM)*, pp. 1–5, 2016.
- [29] S. Salehkalaibar, M. Wigger, and R. Timo, “On hypothesis testing against independence with multiple decision centers,” *IEEE Trans. on Comm.*, vol. 66, pp. 2409–2420, Jan. 2018.
- [30] P. Escamilla, M. Wigger, and A. Zaidi, “Distributed hypothesis testing with concurrent detection,” in *Proc. ISIT*, Jun. 2018.
- [31] P. Escamilla, A. Zaidi, and M. Wigger, “Distributed hypothesis testing with collaborative detection,” in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 512–518, 2018.
- [32] P. Escamilla, M. Wigger, and A. Zaidi, “Distributed hypothesis testing: cooperation and concurrent detection,” *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.
- [33] S. Salehkalaibar, M. Wigger, and L. Wang, “Hypothesis testing over the two-hop relay network,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 4411–4433, Jul. 2019.
- [34] S. Salehkalaibar, M. Wigger, and L. Wang, “Hypothesis testing in multi-hop networks.” [Online]. Available: <https://arxiv.org/abs/1708.05198v1>, 2017.
-

- [35] D. Cao, L. Zhou, and V. Y. F. Tan, “Strong converse for hypothesis testing against independence over a two-hop network,” *Entropy (Special Issue on Multiuser Information Theory II)*, vol. 21, Nov. 2019.
- [36] H. Tyagi and S. Watanabe, “Strong converse using change of measure arguments,” *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.
- [37] W. Gu and M. Effros, “A strong converse for a collection of network source coding problems,” in *2009 IEEE International Symposium on Information Theory*, pp. 2316–2320, IEEE, 2009.
- [38] W. Gu and M. Effros, “A strong converse in source coding for super-source networks,” in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 395–399, IEEE, 2011.
- [39] J. Liu, R. Van Handel, and S. Verdú, “Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 943–947, IEEE, 2017.
- [40] K. Marton, “A simple proof of the blowing-up lemma,” *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.
- [41] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [42] Y. Oohama, “Exponential strong converse for source coding with side information at the decoder,” *Entropy*, vol. 20, no. 5, p. 352, 2018.
- [43] M. Hamad, M. Wigger, and M. Sarkiss, “Strong converses using change of measure and asymptotic Markov chains.” [Online]. Available: <https://arxiv.org/abs/2205.08910>, May 2022.
- [44] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [45] Y. Inan, M. Kayaalp, A. H. Sayed, and E. Telatar, “A fundamental limit of distributed hypothesis testing under memoryless quantization,” in *2022 IEEE International Conference on Communications (ICC)*, pp. 4824–4829, IEEE, 2022.
- [46] J. Körner and K. Marton, “How to encode the modulo-two sum of binary sources (corresp.),” *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, 1979.

-
- [47] M. Hamad, M. Wigger, and M. Sarkiss, “Cooperative multi-sensor detection under variable-length coding,” in *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2021.
- [48] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [49] M. Hamad, M. Wigger, and M. Sarkiss, “Optimal exponents in cascaded hypothesis testing under expected rate constraints,” in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–6, 2021.
- [50] M. Hamad, M. Wigger, and M. Sarkiss, “Two-hop network with multiple decision centers under expected-rate constraints,” in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.
- [51] M. Hamad, M. Wigger, and M. Sarkiss, “Multi-hop network with multiple decision centers under expected-rate constraints.” [Online]. Available: <https://arxiv.org/abs/2208.14243>, 2022.
- [52] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, “Fourier-motzkin elimination software for information theoretic inequalities.” [Online]. Available: <https://arxiv.org/abs/1610.03990>, 2016.
- [53] M. Hamad, M. Sarkiss, and M. Wigger, “Benefits of rate-sharing for distributed hypothesis testing.” [Online]. Available: <https://arxiv.org/pdf/2202.02282.pdf>, 2022.
- [54] S. Sreekumar, A. Cohen, and D. Gündüz, “Privacy-aware distributed hypothesis testing,” *Entropy*, vol. 22, no. 6, p. 665, 2020.
- [55] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. Tan, “Distributed hypothesis testing with privacy constraints,” *Entropy*, vol. 21, no. 5, p. 478, 2019.
- [56] Z. Li, T. J. Oechtering, and D. Gündüz, “Privacy against a hypothesis testing adversary,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1567–1581, 2019.

Titre: Partage des ressources pour l'amélioration des tests d'hypothèses distribués

Mots clés: Tests d'hypothèses distribués; Exposants d'erreur; Contraintes de taux moyen; Capteurs; Centres de décision; Partage du taux

Résumé: Les tests d'hypothèses distribués ont de nombreuses applications dans la sécurité, la surveillance de la santé, le contrôle automobile ou la détection d'anomalies. À l'aide de capteurs distribués, les centres de décision de ces systèmes visent à distinguer une situation normale (hypothèse nulle) d'une situation d'alerte (hypothèse alternative). Nous nous concentrons sur la maximisation de la décroissance exponentielle des probabilités d'erreur de type-II (correspondant aux détections manquées), avec un nombre croissant d'observations, tout en maintenant les probabilités d'erreur de type-I (correspondant aux fausses alertes) en dessous de seuils fixés. Dans cette thèse, nous supposons que différents systèmes ou applications partagent les ressources limitées du réseau et imposent des contraintes de taux moyen sur les liens de communication. Nous caractérisons les premières limites fondamentales de la théorie de l'information sous des contraintes de taux moyen pour les systèmes

avec capteurs multiples et centres de décision multiples. Notre caractérisation révèle un nouveau compromis entre les exposants maximaux d'erreur de type-II aux différents centres de décision qui découle des différentes marges à exploiter sous des contraintes de taux moyen correspondant aux différents seuils d'erreur de type-I des centres de décision. Nous proposons une nouvelle stratégie de multiplexage et de partage du taux pour atteindre ces exposants d'erreur. Notre stratégie se généralise également à toute configuration avec des contraintes de taux moyen et permet d'obtenir des gains prometteurs par rapport aux résultats sur la même configuration avec des contraintes de taux maximal. La méthode de preuve de "converse" que nous utilisons pour caractériser ces limites théoriques peut également être utilisée pour dériver de nouveaux résultats de "converse forte" sous des contraintes de taux maximal. Elle est même applicable à d'autres problèmes tels que la compression ou le calcul distribué.

Title: Sharing Resources for Enhanced Distributed Hypothesis Testing

Keywords: Distributed hypothesis testing; Error exponents; Expected-rate constraints; Sensors; Decision centers; Rate-sharing

Abstract: Distributed hypothesis testing has many applications in security, health monitoring, automotive car control, or anomaly detection. With the help of distributed sensors, the decision centers (DCs) in such systems aim to distinguish between a normal situation (null hypothesis) and an alert situation (alternative hypothesis). Our focus will be on maximizing the exponential decay of the type-II error probabilities (corresponding to missed detections), with increasing numbers of observations, while keeping the type-I error probabilities (corresponding to false alarms) below given thresholds. In this thesis, we assume that different systems or applications share the limited network resources and impose expected-rate constraints on the system's communication links. We characterize the first information-theoretic fundamental limits

under expected-rate constraints for multi-sensor multi-DC systems. Our characterization reveals a new trade-off between the maximum type-II error exponents at the different DCs that stems from different margins to exploit under expected-rate constraints corresponding to the DCs' different type-I error thresholds. We propose a new multiplexing and rate-sharing strategy to achieve the error-exponents. Our strategy also generalizes to any setup with expected-rate constraints with promising gains compared to the results on the same setup under maximum-rate constraints. The converse proof method that we use to characterize the information-theoretic limits can also be used to derive new strong converse results under maximum-rate constraints. It is even applicable to other problems such as distributed compression or computation.