Error Exponent Analysis for the Multiple-Access Channel with Correlated Sources

Arezou Rezazadeh

TESI DOCTORAL UPF / Year 2019

DIRECTORS DE LA TESI
Albert Guillén i Fàbregas and Alfonso Martinez
Department of Information and Communication Technologies
Universitat Pompeu Fabra

Acknowledgements

With my sincerest appreciation of all people who helped me in my scientific journey to reach this point, I would like to express my deepest gratitude to my advisers Prof. Albert Guillén i Fàbregas and Prof. Alfonso Martinez for their support, advisement and understanding. Their careful comments to avoid blind proving, emphasizing to have a profound understanding of the problem and technical guidance have given me invaluable insight into the problem solving and thinking procedure. I was impressed by their supervising style and the time they spent on me.

Special thanks to Dr. Josep Font-Segura for his shared views and guidance during my PhD career. His valuable suggestion made working with him a profitable experience. Our technical discussions with him left me with new ideas. His mind provoking questions and his high-level standard of presenting work are of precious value for me.

I am grateful to my colleague and my friend, Ivano Colombaro for his friendship and his continual encouragement to pass the research challenges. In times of dissatisfaction with the research progress, his inspiration prodded me into continuing on.

Many thanks towards the faculty members of the Department of Information and Communication Technologies for creating a unique friendship and creating a learning atmosphere. My deepest gratitude goes to my former supervisor at Ferdowsi University, Prof. Hodtani, for introducing me the beauty of information theory and his support up to now.

Finally, I am particularly indebted to my beloved family for their endless love, support and affection. Words are hardly sufficient to express my heartfelt gratitude to them for providing an environment for conveying their passion for mathematics and critical thinking to me. Their support and love in the difficult fluctuations of life, in my failures and success, in every step, regardless of my mistakes is a constant source of encouragement and energy to attempt and not giving up. I am fortuitous to have such a great family that without their support, help and love, I could not have achieved any of my goals.

Abstract

Due to delay constraints of modern communication systems, studying reliable communication with finite-length codewords is much needed. Error exponents are one approach to study the finite-length regime from the information-theoretic point of view. In this thesis, we study the achievable exponent for single-user communication and also multiple-access channel with both independent and correlated sources. By studying different coding schemes including independent and identically distributed, independent and conditionally distributed, message-dependent, generalized constant-composition and conditional constant-composition ensembles, we derive a number of achievable exponents for both single-user and multi-user communication, and we analyze them.

Resum

A causa de les restriccions de retard dels sistemes de comunicació moderns, estudiar la fiabilitat de la comunicació amb paraules de codis de longitud finita és important. Els exponents d'error són un mètode per estudiar el règim de longitud finita des del punt de vista de la teoria de la informació. En aquesta tesi, ens centrem en assolir l'exponent per a la comunicació d'un sol usuari i també per l'accés múltiple amb fonts independents i correlacionades. En estudiar els següents esquemes de codificació amb paraules independents i idènticament distribuïdes, independents i condicionalment distribuïdes, depenent del missatge, composició constant generalitzada, i conjunts de composició constant condicional, obtenim i analitzem diversos exponents d'error assolibles tant per a la comunicació d'un sol usuari com per la de múltiples usuaris.

Resumen

Las restricciones cada vez más fuertes en el retraso de transmisión de los sistemas de comunicación modernos hacen necesario estudiar la fiabilidad de la comunicación con palabras de códigos de longitud finita. Los exponentes de error son un método para estudiar el régimen de longitud finita desde el punto de vista la teoría de la información. En esta tesis, nos centramos en calcular el exponente para la comunicación tanto de un solo usuario como para el acceso múltiple con fuentes independientes y correladas. Estudiando diferentes familias de codificación, como son esquemas independientes e idénticamente distribuidos, independientes y condicionalmente distribuidos, que dependen del mensaje, de composición constante generalizada, y conjuntos de composición constante condicional, obtenemos y analizamos varios exponentes alcanzables tanto para la comunicación de un solo usuario como para la de múltiples usuarios.

Preface

The assumption of infinite length sequences usually performs the key role in proving the Shannon coding theorems. However, the delay and the complexity constraints of the modern communication systems demand analysis of the theoretical limits of communication with finite-length sequences. The error exponent approach provides a better way of evaluating the exponential decay of the error probability as a function of the sequence length. Not only does the error exponent establish the fundamental limits of reliable communication but it also gives an insight about constructing better codes whose error probability tends to zero more quickly. Considering the importance of the error exponent, this thesis studies a number of achievable exponents under different coding schemes and carries out analyses for both single-user and multi-user communication systems.

Contents

List of figures **xiv**

List of tables **xv**

1 Introduction **3**

1.1 System Setup . . . 3
1.2 Preliminaries . . . 5
1.2.1 Error Exponent . . . 5
1.2.1.1 Source Coding and Channel Coding . . . 6
1.2.1.2 Joint Source-Channel Coding . . . 9
1.2.2 Random Coding . . . 9
1.2.3 The Multiple-Access Channel . . . 10
1.2.3.1 Distributed Source Coding . . . 11
1.2.3.2 Transmissible Region . . . 12
1.3 Overview of Thesis . . . 15
1.4 Notation . . . 17

2 Single User Communication **19**

2.1 Statistical Dependency between the Messages and Coedwords 21
2.1.1 Message Dependent Ensemble with Statistical Dependency . . . 21
2.1.2 Independent and Conditionally Distributed Ensemble . 22
2.1.3 Conditional Constant-Composition Ensemble . . . 23
2.2 Source-Channel Coding without Statistical Dependency . . . 25
2.2.1 Message-Dependent Ensemble . . . 25
2.2.2 iid Random-Coding Exponent . . . 31
2.2.3 Generalized Constant-Composition Exponent . . . 33
2.3 Comparing the Exponents . . . 35
2.3.1 icd and iid Ensembles . . . 36
2.3.2 Conditional Constant-Composition Ensemble and Constant-Composition Ensemble . . . 38

2.4 Proofs . . . 39
2.4.1 Proof of Proposition 2.1 . . . 39
2.4.2 Proof of Proposition 2.2 . . . 43
2.4.3 Proof of Proposition 2.4 . . . 45
2.4.4 Proof of Lemma 2.1 . . . 50
2.4.5 Proof of Proposition 2.6 . . . 51
2.4.6 Proof of Proposition 2.7 . . . 56
2.4.6.1 Proving the Reverse Direction of (2.181) . . . 61
2.4.7 Proof of Lemma 2.2 . . . 62
2.4.8 Proof of (2.66) . . . 66
2.4.9 Proof of Proposition 2.9 . . . 69
2.4.10 Proof of Lemma 2.3 . . . 70

3 The Multiple-Access Channel with Independent Sources 73
3.1 System Setup . . . 74
3.2 Message-Dependent Exponent . . . 74
3.2.1 Optimal Thresholds . . . 76
3.2.1.1 The Intuition of Optimal Thresholds for Error Type $\{1, 2\}$ . . . 77
3.2.2 A Lower Bound for the Achievable Exponent . . . 87
3.2.3 An Upper Bound for the Achievable Exponent . . . 89
3.2.4 Numerical Example . . . 90
3.3 Generalizing to Multiple-Classes . . . 91
3.4 Proofs . . . 97
3.4.1 Proof of Proposition 3.1 . . . 97
3.4.2 Proof of Proposition 3.2 . . . 101
3.4.3 Proof of Proposition 3.3 . . . 102
3.4.4 Proof of Proposition 3.4 . . . 105

4 The Multiple-Access Channel with Correlated Sources 109
4.1 Message-Dependent Random-Coding Exponent with Statistical Dependency . . . 110
4.1.1 Statistically Independent Input Distributions . . . 111
4.1.1.1 iid Random-Coding Exponent . . . 117
4.1.1.2 Numerical Example . . . 118
4.1.1.3 On the Error Type $\tau \in \{\{1\}, \{2\}\}$ . . . 119
4.1.2 Statistical Dependency between Messages and Codewords . . . 121
4.2 Studying Generalized Conditional Constant-Composition Ensemble . . . 123
4.3 Proofs . . . 126

4.3.1 Proof of Lemma 4.1 126
4.3.2 Proof of Proposition 4.1 128
4.3.3 Proof of (4.4) 132
4.3.3.1 Proof of Equation (4.89) 134
4.3.4 Proof of Lemma 4.2 136
4.3.5 Proof of Proposition 4.2 137
4.3.6 Proof of Proposition 4.3 140
4.3.7 Proof of Proposition 4.4 142
4.3.8 Proof of (4.32) 143
4.3.9 Proof of Proposition 4.5 145

5 Conclusions and Future Work **149**

A General Lemmas **153**

List of Figures

1.1 Transmission of a source over the point-to-point channel. . . . 4
1.2 Transmission of two correlated sources over the MAC. 5
1.3 Communication channel. 8
1.4 Compression of two correlated sources. 12
1.5 Distributed source coding. 13
1.6 Transmissible region for iid ensemble. 14

2.1 $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$. 28
2.2 Message-dependent random coding ensemble gives larger exponent than iid ensemble. 31
2.3 Message-dependent random coding ensemble gives the same exponent as iid ensemble. 32
2.4 $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$ and $E_{s,i}(\rho, P_U, \gamma_2)$, $E_{s,i}(\rho, P_U, \gamma_3)$ for $i = 1, 2$ 58

3.1 Equations (3.19), (3.20), (3.21) and (3.22) with respect to ρ. . 80
3.2 $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_2)$ for $i_1, i_2 = 1, 2$. 80
3.3 The solution of system of equations is derived as $E^{1,1}_{0,s}(\gamma_1^\star, 0) = E^{2,1}_{0,s}(\gamma_1^\star, 0)$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,1}$ are located on $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + E_s(\rho, P_{U_2})$. See Figure 3.2 for more details. 81
3.4 The solution of systems of equations is derived as $E^{2,1}_{0,s}(1, \gamma_2^\star) = E^{2,2}_{0,s}(1, \gamma_2^\star)$. Both $\rho^\star_{2,1}$ and $\rho^\star_{2,2}$ are located on $E_s(\rho, P_{U_1}) + \hat{E}_s(\rho, P_{U_2}, \gamma_2^\star)$. See Figure 3.2 for more details. 83
3.5 The solution of system of equations is derived as $E^{1,2}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,2}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,1}_{0,s}(\gamma_1^\star, \gamma_2^\star)$. $\rho^\star_{1,2}$, $\rho^\star_{2,1}$ and $\rho^\star_{2,2}$, respectively are located on $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + E_s(\rho, P_{U_2})$, $E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2})$ and the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + \hat{E}_s(\rho, P_{U_1}, \gamma_1^\star)$. 84
3.6 $E^{1,1}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,2}_{0,s}(\gamma_1^\star, \gamma_2^\star)$, while $\rho_{\gamma_1^\star} \neq \rho_{\gamma_2^\star}$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,2}$ are located on the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + \hat{E}_s(\rho, P_{U_2}, \gamma_2^\star)$. 86
3.7 $E^{1,1}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,2}_{0,s}(\gamma_1^\star, \gamma_2^\star)$, and $\rho_{\gamma_1^\star} = \rho_{\gamma_2^\star}$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,2}$ are located on the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + \hat{E}_s(\rho, P_{U_2}, \gamma_2^\star)$. 87

4.1 The function $E_{s,\tau,i_1,i_2}(\cdot)$ given by (4.14) for two independent sources $P_{U_1}(u_1)P_{U_2}(u_2)$ versus ρ, for the fixed γ_1 and γ_2 where $i_1, i_2 = 1, 2$. For error type $\tau \in \{\{1\},\{2\}\}$, the solid red and blue curves are respectively $E_s(\rho, P_{U_\tau})$ and $E_s(\rho, P_{U_\tau}) + E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$. 114
4.2 The function $E_{s,\tau,i_1,i_2}(\cdot)$ given by (4.12) for two correlated sources versus ρ, for the fixed γ_1 and γ_2 where $i_1, i_2 = 1, 2$. The solid red and blue curves are respectively given by (1.24) and (4.16). 115
4.3 $\min_{i_1,i_2} f_{i_1,i_2}(\gamma_1, \gamma_2)$ with respect to γ_1 and γ_2. 119
4.4 Single user transmission in the presence of channel state. . . . 121

List of Tables

1.1 The ensemble of all codes for two messages with block length $n = 2$, and $\mathcal{X} = \{0, 1\}$. Each column shows one possible code. 10

3.1 Values of $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ in (3.7) with optimal thresholds $\gamma_1^\star = 0.8159$ $\gamma_2^\star = 0.7057$, for types of error τ, and user classes i_τ and i_{τ^c}. 92
3.2 Values of $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$ in (3.50) for types of error τ, and input distribution Q_{1,i_1}, Q_{2,i_2}. 92
3.3 Values of F^{U}_{τ} in (3.53) for types of error τ. 92
3.4 The behaviour of $f_{\ell_1,\ell_2}(\gamma_{1,2}, \gamma_{1,3}, \gamma_{2,2})$ given by (3.67) for a two-user MAC. 95

4.1 Values of (4.22) with optimal thresholds $\gamma_1^\star = 0.8469$ $\gamma_2^\star = 0.6581$, for types of error τ, and user classes i_τ and i_{τ^c}. 120
4.2 Values of $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$ in (4.27) for types of error τ, and input distribution Q_{1,i_1}, Q_{2,i_2}. 120

Nomenclature

Abbreviations

cc	Constant-composition ensemble
ccc	Conditional constant-composition ensemble
gcc	Generalized constant-composition ensemble
gccc	Generalized conditional constant-composition ensemble
icd	Independent and conditionally distributed ensemble
iid	Independent and identically distributed ensemble
MAC	Multiple-access channel
MAP	Maximum a posteriori
md	Message-dependent ensemble
mds	Message-dependent ensemble with statistical dependency

Chapter 1

Introduction

Shannon in his well known paper [1] answered two fundamental questions of information theory. He considered the system consisting of one transmitter and one receiver, and studied the problem of reliable communication over it. Unlike single-user communication, for multi-user communication containing multiple senders and receivers, the problem of reliable communication is not solved in many cases. In this thesis, we focus on point-to-point and multiple-access channels presented in the following.

1.1 System Setup

Here, we describe the system model of both single-user and multi-user communication systems that will be assumed throughout the thesis (except where stated otherwise). We follow the notation presented in Section 1.4.

Figure 1.1 shows a point-to-point communication system consisting of a source, an encoder, a channel and a decoder. A discrete source over a finite alphabet $\mathcal{U}$ is defined as a sequence of n-dimensional random variables $\boldsymbol{U}$, where each $\boldsymbol{U}$ takes values in $\mathcal{U}^n$. The jth element of the sequence $\boldsymbol{U}$ is denoted by U_j, where $j \in \{1, ..., n\}$. In this thesis, we only consider memoryless sources, and we say that the source P_U is memoryless if and only if, the U_js are iid and their distribution is given by P_U. Thus, the discrete memoryless source P_U generates length-n messages $\boldsymbol{u} = (u_1, ..., u_n) \in \mathcal{U}^n$ according to probability distribution $P^n_{\boldsymbol{U}}(\boldsymbol{u}) = \prod_{j=1}^n P_U(u_j)$.

The output of the source is processed by an encoder where each message is assigned to a codeword with block length n. The encoder maps the message $\boldsymbol{u} \in \mathcal{U}^n$ into the length-n codeword $\boldsymbol{x}(\boldsymbol{u}) = (x_1, ..., x_n) \in \mathcal{X}^n$ drawn from the codebook $\mathcal{C} = \{\boldsymbol{x}(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{U}^n\}$.

Then, codewords are transmitted over a channel with finite input al-

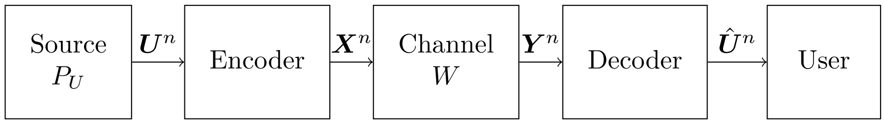

Figure 1.1: Transmission of a source over the point-to-point channel.

phabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$. A discrete channel is a sequence of n-dimension transition matrices W^n, where $W^n(\boldsymbol{y}|\boldsymbol{x})$ is the conditional probability of $\boldsymbol{y} \in \mathcal{Y}^n$ given $\boldsymbol{x} \in \mathcal{X}^n$. Here, we only consider discrete memoryless channels where the output sequence $\boldsymbol{y} = (y_1, ..., y_n) \in \mathcal{Y}^n$ is randomly generated from the input sequence $\boldsymbol{x} \in \mathcal{X}^n$ according to $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{j=1}^n W(y_j|x_j)$. Due to the randomness inherent of the channel, the received sequence at the output of channel differs from the original one. Hence, a decoder is used to estimate the transmitted message based on a specific criterion. Here, we use the maximum a posteriori (MAP) decoder. More precisely, by receiving the sequence $\boldsymbol{y}$, the MAP decoder estimates the transmitted message $\hat{\boldsymbol{u}}$ based on the following criterion

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u} \in \mathcal{U}^n} P_{\boldsymbol{U}}^n(\boldsymbol{u}) W^n(\boldsymbol{y} \,|\, \boldsymbol{x}(\boldsymbol{u})). \tag{1.1}$$

An error occurs if the decoded message $\hat{\boldsymbol{u}}$ differs from the transmitted $\boldsymbol{u}$. The error probability for a given codebook $\mathcal{C}$ is given by

$$\epsilon^n(\mathcal{C}) \triangleq \mathbb{P}[\hat{\boldsymbol{U}} \neq \boldsymbol{U}]. \tag{1.2}$$

In addition, we say that the source P_U is transmissible over the channel, if there exists a sequence of codebooks $\mathcal{C}_n$ such that we have $\lim_{n\to\infty} \epsilon^n(\mathcal{C}_n) = 0$.

In this thesis, we also study the transmission over the multiple-access channel (MAC). Figure 1.2 shows the transmission of two correlated sources over the MAC. The discrete memoryless sources are characterized by a probability distribution $P_{U_1U_2}$ on the alphabet $\mathcal{U}_1 \times \mathcal{U}_2$, where $\mathcal{U}_1$ and $\mathcal{U}_2$ are the respective alphabets of the two sources, and the source messages $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ have length n.

For user $\nu = 1, 2$, the source message $\boldsymbol{u}_\nu$ is mapped onto codeword $\boldsymbol{x}_\nu(\boldsymbol{u}_\nu)$, which also has length n and is drawn from the codebook $\mathcal{C}^\nu = \{\boldsymbol{x}_\nu(\boldsymbol{u}_\nu); \boldsymbol{u}_\nu \in \mathcal{U}_\nu^n\}$. Both terminals send the codewords over a discrete memoryless multiple access channel with transition probability $W(y|x_1, x_2)$, input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and output alphabet $\mathcal{Y}$.

Given the received sequence $\boldsymbol{y}$, the decoder estimates the transmitted pair messages $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ based on the maximum a posteriori criterion,

$$(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) = \arg\max_{(\boldsymbol{u}_1, \boldsymbol{u}_2) \in \mathcal{U}_1^n \times \mathcal{U}_2^n} P_{\boldsymbol{U}_1\boldsymbol{U}_2}^n(\boldsymbol{u}_1, \boldsymbol{u}_2) W^n\Big(\boldsymbol{y}|\boldsymbol{x}_1(\boldsymbol{u}_1), \boldsymbol{x}_2(\boldsymbol{u}_2)\Big). \tag{1.3}$$

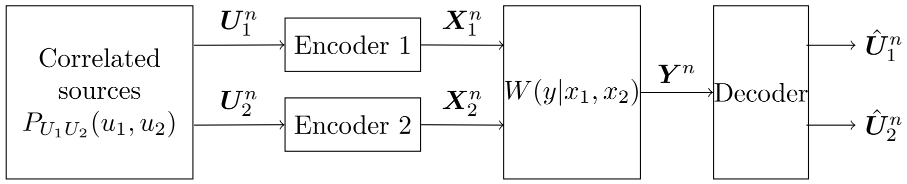

Figure 1.2: Transmission of two correlated sources over the MAC.

An error occurs if the decoded messages $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2)$ differ from the transmitted $(\boldsymbol{u}_1, \boldsymbol{u}_2)$; the error probability for a given pair of codebooks is thus given by

$$\epsilon^n(\mathcal{C}^1, \mathcal{C}^2) \triangleq \mathbb{P}\left[(\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)\right]. \tag{1.4}$$

Like single-user communication, the pair of sources (U_1, U_2) is transmissible over the channel if there exists a sequence of codebooks $(\mathcal{C}_n^1, \mathcal{C}_n^2)$ such that $\lim_{n\to\infty} \epsilon^n(\mathcal{C}_n^1, \mathcal{C}_n^2) = 0$.

1.2 Preliminaries

The goal of communication is transmitting a message from a source to a destination, reliably. By reliable transmission, we mean that one can reconstruct the message at the destination with a very low probability of error.

In [1], Shannon studied the problem of reliable communication in a single point-to-point connection. He answered two fundamental questions in communication. Firstly, he found the fundamental limits on data compression as source entropy and showed the impossibility of compressing a source without information loss, if the average number of bits per source symbol is less than the source entropy. Secondly, he also proved channel coding theorem stating that for a given channel, there exists a quantity known as channel capacity such that reliable communication is only possible at rates below the channel capacity.

In fact, Shannon by introducing the idea of random coding, and using the jointly typical decoder, found the fundamental limits of communication. In this section, we review these fundamental theorems for both point-to-point and multiple-access channels.

1.2.1 Error Exponent

A large number of information theoretic problems can be modelled as transmitting a message over a noisy channel with very small error probability. However, having small error probability corresponds to an increasingly large block length, and large block length n imposes delay on the system. In addition, due to delay requirements in practical systems, it is needed to use codewords with finite block length.

Hence, it is important to study error probability in presence of codewords with finite block length n. By using the maximum likelihood decoder, it has been shown that for discrete memoryless channel, there exist a coding scheme which error probability tends to zero exponentially as block length tends to infinity [2], [3, Theorem 3], [4, Ch. 9] and [5, Ch. 5].

In fact, by expressing error probability as an exponential function of block length n, we can not only determine reliable transmission conditions that under them error probability tends to zero, but we can also construct better coding schemes leading error probability tends to zero more quickly.

Roughly speaking, we are interested expressing the error probability as an exponential function of block length n such as e^{-nE}, where the exponent E shows how quickly the error probability drops as a function of n. Recalling that for point-to-point communication, the error probability is given by (1.2), we define the error exponent as follows. An exponent E is said to be achievable if there exists a sequence of codebooks such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\left(\epsilon^n(\mathcal{C}_n)\right) \geq E, \tag{1.5}$$

and the supremum of all achievable exponents is defined as error exponent.

In fact, we can measure the performance of two different coding schemes by comparing their corresponding achievable exponents. Particularly, the one that has larger exponent, its error probability tends to zero more quickly. Throughout this thesis, we propose different coding schemes and by comparing their exponent, we evaluate their performance.

For the multiple-access channel, the error probability given by (1.4). Similarly, an exponent E is achievable if there exists a sequence of codebooks such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log\left(\epsilon^n(\mathcal{C}_n^1,\mathcal{C}_n^2)\right) \geq E, \tag{1.6}$$

and error exponent is defined as the supremum of all achievable exponents given by (1.6).

1.2.1.1 Source Coding and Channel Coding

As mentioned, the source coding theorem deals with the ultimate limit of compressing a source sequence. Here, we review it. Consider a discrete memoryless source P_U on the alphabet $\mathcal{U}$, where an encoder compresses all $|\mathcal{U}|^n$ messages into a set with 2^{nR} elements, i. e.

$$\mathcal{U}^n \rightarrow \left\{1, 2, ..., 2^{nR}\right\}, \tag{1.7}$$

where the rate R is the average number of bits per source symbol.

In [6, Eq. (5.18)], it is showed that the average error probability of decompressing is upper bounded as an exponential function with respect to n, and its exponent is given by

$$\max_{\rho \geq 0} \rho R - E_s(\rho, P_U), \tag{1.8}$$

where

$$E_s(\rho, P_U) = (1+\rho) \log \left(\sum_u P_U(u)^{\frac{1}{1+\rho}} \right), \tag{1.9}$$

is known as Gallager source function.

Since the error probability is upper bounded by $e^{-n\left(\max_{\rho \geq 0} \rho R - E_s(\rho, P_U)\right)}$, it tends to zero if its exponent given by (1.8) is greater than zero. By using a similar argument to [6, Theorem 5.1], it can be proved that reliable compression is possible, if we have

$$R > \frac{\partial E_s(\rho, P_U)}{\partial \rho}\Big|_{\rho=0} = H(P_U), \tag{1.10}$$

where $H(P_U)$ is known as source entropy, and is given by

$$H(P_U) = -\sum_{u \in \mathcal{U}} P_U(u) \log \left(P_U(u)\right), \tag{1.11}$$

and conversely, if the rate R is less than the source entropy, the probability of error tends to one.

Next, to review the channel coding theorem, we consider the Figure 1.3. Message $m \in \left\{1, 2, ..., 2^{nR}\right\}$ is mapped into the length-n codeword $\boldsymbol{x}(m)$ drawn from the codebook $\mathcal{C} = \left\{\boldsymbol{x}(1), ..., \boldsymbol{x}(2^{nR})\right\}$. All codewords are generated independently according to an identical product distribution $Q^n(\boldsymbol{x}) = \prod_{j=1}^n Q(x_j)$. The decoder receives the sequence $\boldsymbol{y}$ at the channel output, and estimates the transmitted message based on the maximum likelihood criterion,

$$\hat{m} = \underset{m \in \{1,2,...,2^{nR}\}}{\arg\max} \; W^n\left(\boldsymbol{y}|\boldsymbol{x}(m)\right). \tag{1.12}$$

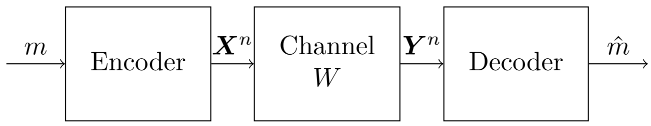

Figure 1.3: Communication channel.

An error occurs if the decoded message $\hat{m}$ differs from the transmitted m.

In [6, Eq. (6.31)], it has been shown that the iid exponent

$$E(R, W) = \max_{Q \in \mathcal{P}_\mathcal{X}} \max_{\rho \in [0,1]} E_0(\rho, Q, W) - \rho R, \tag{1.13}$$

is achievable where

$$E_0(\rho, Q, W) = -\log \sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{1.14}$$

is known as Gallager channel function.

To find the reliable transmission conditions, by setting $E(R, W)$ in (1.13) positive, and following the approach given by [6, Theorem 6.2], we find that

$$R < \max_{Q \in \mathcal{P}_\mathcal{X}} \frac{\partial E_0(\rho, Q, W)}{\partial \rho}\Big|_{\rho=0} = \max_{Q \in \mathcal{P}_\mathcal{X}} I(X; Y) = C, \tag{1.15}$$

where $I(X; Y)$ is known as mutual information between input and output of channel W, and C is the channel capacity. As mention in (1.15), the channel capacity is the maximum of all $I(X; Y)$ which is defined as

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} Q(x) W(y|x) \log \frac{W(y|x)}{\sum_{\bar{x}} Q(\bar{x}) W(y|\bar{x})}. \tag{1.16}$$

Conversely, if the rate of information is higher than the channel capacity, then the probability of error is bounded away from zero.

Combining (1.10) and (1.15) provides us the possibility of the reliable transmission of the discrete memoryless source P_U over a point-to-point memoryless channel, if its entropy is less than the capacity of the channel, i. e.

$$H(P_U) < C. \tag{1.17}$$

Similarly, by combining (1.8) and (1.13), we find that by the concatenation of source and channel codes, the error probability exponentially tends to zero by the exponent

$$\min \left\{ \max_{\rho \geq 0} \rho R - E_s(\rho, P_U), \max_{\rho \in [0,1]} \max_{Q \in \mathcal{P}_\mathcal{X}} E_0(\rho, Q, W) - \rho R \right\}. \tag{1.18}$$

1.2.1.2 Joint Source-Channel Coding

For joint source-channel coding, by considering Figure 1.1, the encoder maps a length-n source message $\boldsymbol{u}$ to the length-n codeword $\boldsymbol{x}(\boldsymbol{u})$, where all codewords generated independently according to an identical input distribution Q. In [5, Prob. 5.16], Gallager derived the following exponent for an iid ensemble

$$\max_{\rho\in[0,1]} E_0(\rho, Q, W) - E_s(\rho, P_U), \tag{1.19}$$

where $E_s(\cdot)$ and $E_0(\cdot)$, respectively are given by (1.9) and (1.14).

Comparing (1.18) and (1.19), we find that joint source-channel coding leads to larger exponent than the concatenation of source and channel coding, and hence its performance in terms of error exponent is better than the separate source-channel coding. In [7, Theorem 3], it was proved that the error exponent of joint design is may be up to twice that of the separate design. However, by a similar argument to Gallager [5, Theorem 5.6.3], we find (1.17) as a reliable transmission condition for joint source-channel coding.

The rest of this thesis is devoted to joint source-channel coding exponent for both point-to-point and multiple-access channels when different coding schemes are applied.

1.2.2 Random Coding

Shannon in [1] used the random coding idea to prove the possibility of reliable transmission over a noisy channel. We recall that, if there exists a sequence of codebooks whose error probability tends to zero, reliable communication is possible.

Consider an iid ensemble, where for a given input distribution Q, all codewords of each codebook are generated independently according to the identical distribution Q. In order to compute the average error probability over the ensemble of all codebooks, it suffices that for each message, we sum over all length-n codewords $\boldsymbol{x} \in \mathcal{X}^n$ when the input distribution Q is applied. Table 1.1 shows the ensemble of all possible codebooks for two messages m_1 and m_2, when the length of codewords is $n = 2$ and $\mathcal{X} = \{0, 1\}$. Each column of Table 1.1 shows one possible codebook and the probability of occurring each of them depends on the input distribution Q. If the error probability averaged over an ensemble of codebooks tends to zero, then there exists at least one sequence of codebooks whose error probability also tends to zero.

In this thesis, we frequently compute the average error probability by applying random coding technique, and then we find an achievable exponent

Table 1.1: The ensemble of all codes for two messages with block length $n = 2$, and $\mathcal{X} = \{0, 1\}$. Each column shows one possible code.

m_1	00	00	00	00	01	01	01	01	10	10	10	10	11	11	11	11
m_2	00	01	10	11	00	01	10	11	00	01	10	11	00	01	10	11

for our problems. We usually use two different forms of solution to compute error exponent. The first way of solution suggested by Csiszár and Körner [8], is built based on the method of types properties, and the results are given in the form of Kullback-Leibler divergence minimization. The final expression of the exponent is known as the exponent in the primal domain, i. e. as a multidimensional optimization problem over distributions. In fact, the exponent in the primal domain gives us a good insight about the characterize of the error event; however, since it contains a multidimensional optimization problem, it is difficult to analyze.

On the other hand, for obtaining the error exponent, there is another way of solution suggested by Gallager [5], where the exponent is derived by applying Chernoff bound. Results are given in the form of optimization over an scaler usually shown by ρ. The final expression of the exponent is known as the exponent in the dual domain, i. e. a lower dimensional problem over parameters in terms of Gallager source and channel functions. Comparing to the exponent in the primal domain, the exponent in the dual domain is easier to analyze; however, sometimes it is difficult to understand the operational meanings behind the derivations.

Usually, by applying Lagrange duality theorem on the exponent in the primal domain, we can get the exponent in the dual domain. In the thesis, we apply both methods for different problems, and we use Lagrange duality theorem to have the results both in primal domain and dual domain.

1.2.3 The Multiple-Access Channel

As shown in Figure 1.2, the multiple access channel considers the problem of transmitting information from two or more sources to one receiver. To answer the fundamental limits of communication over the MAC, like point-to-point case, reliable compression and reliable transmission of it were studied. Here, we study the MAC with two correlated sources. We proceed by formally presenting the compression of two sources, and then discussing the reliable transmission over the MAC.

1.2.3.1 Distributed Source Coding

Distributed source coding studies the compression of multiple correlated information sources that do not communicate with each other.

As shown in Figure 1.4, we consider two correlated sources with joint distribution $P_{U_1U_2}$. For $\nu = 1, 2$, the encoder ν compresses all $\mathcal{U}_\nu^n$ messages into a set with 2^{nR_ν} elements, i. e. for two encoder mapping, we have

$$\mathcal{U}_1^n \to \left\{1, 2, .., 2^{nR_1}\right\}, \tag{1.20}$$

$$\mathcal{U}_2^n \to \left\{1, 2, .., 2^{nR_2}\right\}, \tag{1.21}$$

and the decoder decodes $\left(\hat{\boldsymbol{U}}_1^n, \hat{\boldsymbol{U}}_2^n\right)$ as the transmitted messages, i. e. the decoder mapping is

$$\left\{1, 2, .., 2^{nR_1}\right\} \times \left\{1, 2, .., 2^{nR_2}\right\} \to \mathcal{U}_1^n \times \mathcal{U}_2^n. \tag{1.22}$$

As pointed in (1.4), an error occurs if $\left(\hat{\boldsymbol{U}}_1^n, \hat{\boldsymbol{U}}_2^n\right) \neq (\boldsymbol{U}_1^n, \boldsymbol{U}_2^n)$. In fact, the error event $\hat{\boldsymbol{U}} \neq \hat{\boldsymbol{U}}$ can be split into three disjoint types of error events, namely $(\hat{\boldsymbol{U}}_1, \boldsymbol{U}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$, $(\boldsymbol{U}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$ and $(\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$. These events are respectively labelled by τ, with $\tau \in \{\{1\}, \{2\}, \{1, 2\}\}$. The notation of index τ is explained in Section 1.4, specifically in (1.43) and (1.44). In Chapter 4, we present the exponent of the error probability as

$$\min_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \max_{\rho \in [0,1]} \rho R_\tau - E_{s,\tau}(\rho, P_{U_1U_2}), \tag{1.23}$$

where for $\tau = \{1, 2\}$, $R_\tau = R_1 + R_2$. Moreover, $E_{s,\tau}(\cdot)$ is the generalized Gallager's source functions for error type τ, and is given by

$$E_{s,\tau}(\rho, P_{U_1U_2}) = \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{U_1U_2}(u_1, u_2)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{1.24}$$

Considering source coding theorem for single-user communication, if these two sources are encoded and decoded separately, the reliable rate for the first source is $R_1 > H\left(P_{U_1}\right)$, and similarly the reliable rate for the second source is $R_2 > H\left(P_{U_2}\right)$. Thus, to have a reliable compression, we have $R_1 + R_2 > H\left(P_{U_1}\right) + H\left(P_{U_2}\right)$. Slepian and Wolf have shown that it is possible to code these two sources with lower rate as [9]

$$R_\tau > \frac{\partial E_{s,\tau}\left(\rho, P_{U_1U_2}\right)}{\partial \rho}\Big|_{\rho=0} = H\left(P_{U_\tau|U_{\tau^c}}\right), \tag{1.25}$$

where as presented in Section 1.4, for $\tau = \{1, 2\}$, (1.25) is equal to $R_1 + R_2 > H(P_{U_1U_2})$.

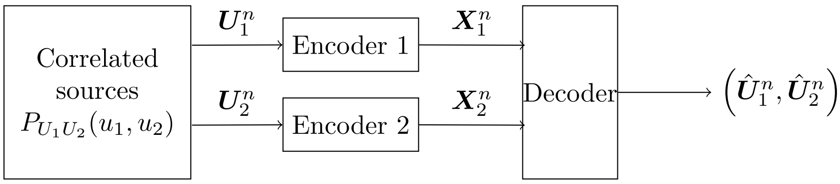

Figure 1.4: Compression of two correlated sources.

Considering Figure 1.5, when two sources are encoded and decoded separately, the achievable region is shown by blue; however, by separately encoding and jointly decoding, the achievable region increases to the red region. Roughly speaking, instead of $H\left(P_{U_1}\right)+H\left(P_{U_2}\right)$, a rate $R > H(P_{U_1U_2})$ would be sufficient to accurately reconstruct both P_{U_1} and P_{U_2} at the decoder. As a result, instead of $H\left(P_{U_1}\right)+H\left(P_{U_2}\right)$, we only need $H\left(P_{U_1U_2}\right)$ to describe joint sources error freely.

1.2.3.2 Transmissible Region

Many studies found the reliable transmission conditions for the multiple-access channel [10], [11], [12]. Separate source-channel coding for the MAC with independent sources was studied in [10] and [13], where for $\nu = 1, 2$, encoder ν takes as input a message m_ν uniformly distributed on the set $\left\{1, 2, ..., 2^{nR_\nu}\right\}$, and transmits its corresponding codeword $\boldsymbol{x}_\nu(m_\nu)$ from the codebook $\mathcal{C}_\nu = \left\{\boldsymbol{x}_\nu(1), ..., \boldsymbol{x}_\nu(2^{nR_\nu})\right\}$ over the MAC with transition probability of W. Considering iid ensemble, for given input distributions Q_1 and Q_2, by receiving the output sequence, the decoder estimates the transmitted messages based on the maximum likelihood decoder,

$$(\hat{m}_1, \hat{m}_2) = \underset{(m_1,m_2):m_1\in\left\{1,...,2^{nR_1}\right\},m_2\in\left\{1,...,2^{nR_2}\right\}}{\arg\max} W^n\left(\boldsymbol{y} \,|\, \boldsymbol{x}_1, \boldsymbol{x}_2\right). \tag{1.26}$$

In [13], the exponent of the error probability was derived as

$$\min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \max_{\rho\in[0,1]} E_0\left(\rho, Q_\tau, WQ_{\tau^c}\right) - \rho R_\tau \tag{1.27}$$

where $E_0(\cdot)$ is given by (1.14), and is the Gallager channel function for channel WQ_{τ^c} and input distribution Q_τ, i. e.

$$E_0\left(\rho, Q_\tau, WQ_{\tau^c}\right) =$$
$$-\log \sum_{x_{\tau^c}\in\mathcal{X}_{\tau^c}, y\in\mathcal{Y}} \left(\sum_{x_\tau\in\mathcal{X}_\tau} Q_\tau(x_\tau) W(y|x_1,x_2)^{\frac{1}{1+\rho}} Q_{\tau^c}(x_{\tau^c})^{\frac{1}{1+\rho}}\right)^{1+\rho}, \tag{1.28}$$

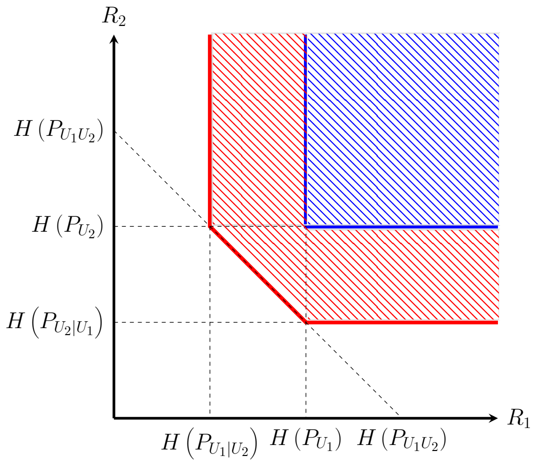

Figure 1.5: Distributed source coding.

and for types of error $\tau = \{1\}$ and $\tau = \{2\}$, WQ_{τ^c} denotes as a point-to-point channel with input and output alphabets given by $\mathcal{X}_\tau$ and $\mathcal{X}_{\tau^c} \times \mathcal{Y}$, respectively, and transition probability $W(y|x_1,x_2)Q_{\tau^c}(x_{\tau^c})$. For $\tau = \{1,2\}$, the input distribution $Q_\tau = Q_1Q_2$ is the product distribution $Q_1(x_1)Q_2(x_2)$ over the alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, and $WQ_{\tau^c,i} = W$. In [10] and [13], by using the similar argument to the single-user communication, the reliable transmission conditions for the MAC was derived as

$$R_\tau < \frac{\partial E_0\left(\rho, Q_\tau, WQ_{\tau^c}\right)}{\partial \rho}\Big|_{\rho=0} = I\left(X_\tau; Y|X_{\tau^c}\right). \tag{1.29}$$

The transmissible region derived by conditions in (1.29), is shown in Figure 1.6. Like before, if the decoder decodes the message of each user separately, only the blue region in Figure 1.6 is achieved. In fact, the blue region of Figure 1.6 can be interpreted as the region that obtained by considering the assumption that each user send its message over its channel. Assuming no interference between the two users, the channel model can be simplified as $W(y|x_1,x_2) = W(y|x_1)W(y|x_2)$, and hence we only can achieve the blue region. However, by decoding simultaneously the messages of both users, the transmissible region increases to the red region.

Combining (1.29) with (1.25), we can say that correlated sources $P_{U_1U_2}$ can be transmitted reliably over the MAC, if

$$H\left(P_{U_\tau|U_{\tau^c}}\right) < I(X_\tau; Y|X_{\tau^c}). \tag{1.30}$$

However, by an example in [12], it was shown that the above strategy is suboptimal for correlated sources. Suboptimal means one can send pair mes-

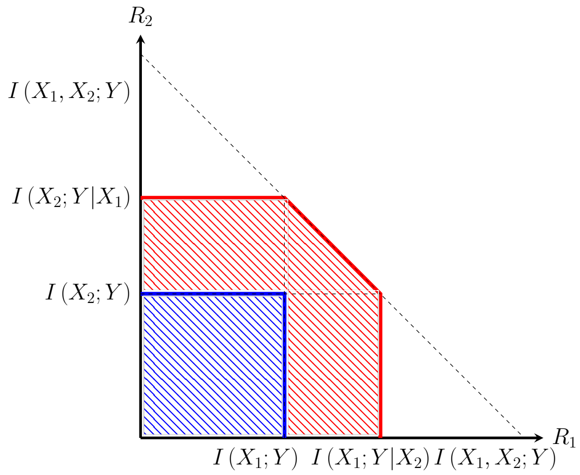

Figure 1.6: Transmissible region for iid ensemble.

sages reliably over the MAC without satisfying (1.30). Here, we review the presented example in [12].

Based on the condition in (1.30), to have a reliable communication, the rate should satisfies $H(P_{U_1U_2}) < I(X_1, X_2; Y)$. Now, assume U_1 and U_2 as two binary random variables with joint probability distribution $P_{U_1U_2}(u_1, u_2)$, where $u_1, u_2 \in \{0, 1\}$, and

$$\begin{cases} P_{U_1U_2}(0,0) = \frac{1}{3}, & P_{U_1U_2}(0,1) = 0, \\ P_{U_1U_2}(1,0) = \frac{1}{3}, & P_{U_1U_2}(1,1) = \frac{1}{3}. \end{cases} \tag{1.31}$$

From (1.31), we find that $H(P_{U_1U_2}) = 1.58$ bits. By considering a MAC with characterization $Y = X_1 + X_2$, we can verified that for iid ensemble, the maximum of mutual information $I(X_1, X_2; Y)$ is 1.5. Consequently,

$$H(P_{U_1U_2}) = 1.58 > 1.5 = I(X_1, X_2; Y), \tag{1.32}$$

which means that these two correlated sources do not satisfies conditions in (1.30) and cannot be transmitted reliably over the given channel.

On the other hand, with simple encoding as $X_1 = U_1$ and $X_2 = U_2$, we have

$$Y = X_1 + X_2 = U_1 + U_2 = \begin{cases} 0 \overset{\text{decoded as}}{\Longrightarrow} \hat{U}_1 = 0, & \hat{U}_2 = 0, \\ 1 \overset{\text{decoded as}}{\Longrightarrow} \hat{U}_1 = 1, & \hat{U}_2 = 0, \\ 2 \overset{\text{decoded as}}{\Longrightarrow} \hat{U}_1 = 1, & \hat{U}_2 = 1, \end{cases} \tag{1.33}$$

where obviously, one can decode the messages correctly, and hence the probability of error is equal to zero. In other words, the above example shows the sub-optimality of (1.30), where the example does not satisfy the reliable transmission conditions which are derived by considering iid ensemble; however, with a simple encoding and decoding, the messages are decoded with zero error probability.

In fact, in [12], MAC with correlated sources was studied, where system model is the same with the one discussed in Section 1.1. In [12], it was proved that in presence of correlated sources, codes generated according to the conditional probability distributions statistically depending on the source messages, leads to larger transmissible region than that of achieved by iid ensemble. In fact, the following transmissible region was derived for the MAC with correlated sources [12]

$$H\left(P_{U_\tau|U_{\tau^c}}\right) < I\left(X_\tau; Y|X_{\tau^c}, U_{\tau^c}\right). \tag{1.34}$$

Although, for correlated sources, for error type $\tau = \{1, 2\}$, we have $I\left(X_1, X_2; Y\right)$ in both (1.30) and (1.34), the values of $I\left(X_1, X_2; Y\right)$ in (1.30) and (1.34) are different. More specifically, for the mentioned example in [12] reviewed above, in (1.30), $I\left(X_1, X_2; Y\right) = 1.5$ which is derived by considering the joint distribution

$$P_{X_1X_2Y}(x_1, x_2, y) = Q_1(x_1)Q_2(x_2)W(y|x_1, x_2), \tag{1.35}$$

while the mutual information $I\left(X_1, X_2; Y\right) = 1.58$ in (1.34) is derived by considering the joint distribution

$$P_{X_1X_2Y}(x_1, x_2, y) = \sum_{u_1,u_2} P_{U_1U_2}(u_1, u_2)Q_1(x_1)Q_2(x_2)W(y|x_1, x_2). \tag{1.36}$$

Thus, $I\left(X_1, X_2; Y\right)$ for the iid ensemble in (1.30) is different with that of for icd ensemble in (1.34).

However, the conditions in (1.34) are not still sufficient. An example presented in [14], shows that one can transmit information through the MAC reliably, without satisfying (1.34). It means that in contrast to single-user communication, the problem of reliable transmission of two correlated sources has not been solved yet and just the sufficient conditions of a reliable transmission has been derived.

1.3 Overview of Thesis

The thesis is structured as follows:

- In Chapter 2, we study single-user communication under various coding scheme including iid, icd, md, gcc, cc and ccc ensembles. We re-obtain a number of achievable random coding exponents for joint source channel coding in both primal and dual domains. For message-dependent random-coding exponent, we presented another proof to show that in terms of error exponent, there is no benefit to generate codewords with more than two input distributions. We also find that for point-to-point channel, there is no penalty in the error exponent if the messages are assumed to be statistically independent of the codewords.

- In Chapter 3, we derive an achievable error exponent for the multiple-access channel with two independent sources. For each user, the source messages are partitioned into two classes and codebooks are generated by drawing codewords from an input distribution depending on the class index of the source message. The partitioning thresholds that maximize the achievable exponent are given by the solution of a system of equations. We also derive both lower and upper bounds for the achievable exponent in terms of Gallager's source and channel functions. By using the results obtained by Chapter 4, we can conclude that considering statistical dependency between messages and codewords may not improve error exponent for the MAC with independent sources.

- In Chapter 4, after discussing about the Gallager's source function for two correlated sources, we study the random-coding exponent of joint source-channel coding for the multiple-access channel with correlated sources. For each user, by defining a threshold, the messages of each source are partitioned into two classes. The achievable exponent for correlated sources with two message-dependent input distributions for each user is determined and shown to be larger than that achieved using only one input distribution for each user. A system of equations is presented to determine the optimal thresholds maximizing the achievable exponent. We show that the obtained achievable exponent is ensemble tightness. We also generalize the result to constant-composition families.

- In Chapter 5, we discuss about the future works.

- Appendix A provide general Lemmas used frequently throughout this thesis.

1.4 Notation

Throughout this thesis, the following notations are used.

Sets and variables:
Sets are denoted by calligraphic upper case letter, e.g., $\mathcal{X}$, and the n-Cartesian product set of $\mathcal{X}$ is denoted by $\mathcal{X}^n$. A scalar random variable is denoted by a capital letter, e.g., X, lower case letters is used as a particular realisation, e.g., $x \in \mathcal{X}$, capital bold letter denotes the random vector, e.g., $\boldsymbol{X}$, small bold letter $\boldsymbol{x} \in \mathcal{X}^n$ is the deterministic vector. The cardinality of a set such as $\mathcal{X}$, or equivalently the number of elements in $\mathcal{X}$ is shown by $|\mathcal{X}|$.

Probability distribution and types:
Except where stated otherwise, the probability distribution of a random variable is denoted by placing the random variable as a subscript, e.g. P_U. A joint distribution of a pair of random variables (U_1, U_2) is denoted by $P_{U_1U_2}$ and the conditional distribution is denoted by $P_{U_1|U_2}$ or $P_{U_2|U_1}$. The distribution of random vector $\boldsymbol{U}$ with length n is shown by $P^n_{\boldsymbol{U}}$. The set of all possible distributions of single letter U is denoted by $\mathcal{P}_{\mathcal{U}}$, and the set of all empirical distributions on a vector in $\mathcal{U}^n$ (i.e. types) is denoted by $\mathcal{P}^n_{\mathcal{U}}$.

Given $\hat{P}_X \in \mathcal{P}^n_{\mathcal{X}}$, the type class $\mathcal{T}^n(\hat{P}_X)$ is the set of all sequences in $\mathcal{X}^n$ with type $\hat{P}_X$. If $\boldsymbol{x} \in \mathcal{T}^n(\hat{P}_X)$, for any probability distribution $Q^n(\boldsymbol{x}) = \prod_{i=1}^n Q(x_i)$, we have the following facts [15]

$$Q^n(\boldsymbol{x}) = e^{n\sum_{x\in\mathcal{X}} \hat{P}_X(x)\log Q(x)}, \tag{1.37}$$

$$\frac{e^{nH(\hat{P}_X)}}{(n+1)^{|\mathcal{X}|}} \le \left|\mathcal{T}^n(\hat{P}_X)\right| \le e^{nH(\hat{P}_X)}. \tag{1.38}$$

Considering (1.37) and (1.38), we have

$$\mathbb{P}\Big[\mathcal{T}^n(\hat{P}_X)\Big] = \sum_{\boldsymbol{x}\in\mathcal{T}^n(\hat{P}_X)} Q^n(\boldsymbol{x}) \le e^{-nD(\hat{P}_X\|Q)}. \tag{1.39}$$

Given $\hat{P}_{XY} \in \mathcal{P}^n_{\mathcal{X}\times\mathcal{Y}}$ and $\boldsymbol{y} \in \mathcal{T}^n(\hat{P}_Y)$, the conditional type class $\mathcal{T}^n_{\boldsymbol{y}}(\hat{P}_{XY})$ is defined to be the set of all sequences $\boldsymbol{x} \in \mathcal{X}^n$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{XY})$. It can be proved that [15]

$$\left|\mathcal{T}^n_{\boldsymbol{y}}(\hat{P}_{XY})\right| = \frac{\left|\mathcal{T}^n(\hat{P}_{XY})\right|}{\left|\mathcal{T}^n(\hat{P}_Y)\right|}. \tag{1.40}$$

Similarly, for given $y \in \mathcal{Y}$, the set $\mathcal{T}^n_y(\hat{P}_{X|Y})$ is defined to be the set of all sequences $\boldsymbol{x}$ such that $\boldsymbol{x}$ contains $n\hat{P}_{X|Y}(x|y)$ occurrence of letter $x \in \mathcal{X}$.

To distinguish between conditional input distribution and input distribution independent of the source probability, the conditional input distribution is denoted by bar, i. e. $\bar{Q}$, while Q denotes an input distribution statistically independent of message. More specifically,

$$P_X(x)P_U(u) = Q(x)P_U(u) \tag{1.41}$$

$$P_{XU}(x,u) = P_U(u)\bar{Q}(x|u). \tag{1.42}$$

The MAC notation:

Throughout this thesis, except where stated otherwise, we use the following notation for the MAC. The symbol ν denotes the user 1 or 2, i. e. $\nu \in \{1,2\}$, and ν^c denotes the complement index of ν among set $\{1,2\}$. To simplify some expressions, we use underline to represent a pair of quantities for users 1 and 2, such as $\underline{u} = (u_1, u_2)$, $\underline{\boldsymbol{u}} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $\underline{\mathcal{U}} = \mathcal{U}_1 \times \mathcal{U}_2$, $P_{\underline{U}}(\underline{u}) = P_{U_1U_2}(u_1, u_2)$ or the transition probability of the MAC as $W(y|\underline{x}) = W(y|x_1, x_2)$. We also frequently use the symbol $\tau \in \{\{1\},\{2\},\{1,2\}\}$ to denote the error event type of the error probability (1.4), i. e. $\tau = \{1\}$ as the error event $(\hat{\boldsymbol{U}}_1, \boldsymbol{U}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$, $\tau = \{2\}$ as the error event $(\boldsymbol{U}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$ and $\tau = \{1,2\}$ for the error event $(\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)$. The complement of τ is denoted by τ^c among the subsets of $\{1,2\}$. For example, $\tau^c = \{2\}$ for $\tau = \{1\}$ and $\tau^c = \emptyset$ for $\tau = \{1,2\}$. In order to simplify some expressions, we adopt the following notational convention,

$$u_\tau = \begin{cases} \emptyset & \tau = \emptyset \\ u_1 & \tau = \{1\} \\ u_2 & \tau = \{2\} \\ \underline{u} & \tau = \{1,2\} \end{cases}, \quad Q_\tau(x_\tau) = \begin{cases} \emptyset & \tau = \emptyset \\ Q_1(x_1) & \tau = \{1\} \\ Q_2(x_2) & \tau = \{2\} \\ Q_1(x_1)Q_2(x_2) & \tau = \{1,2\} \end{cases}, \tag{1.43}$$

when X_1 and X_2 are independent from each other, and also

$$P_{U_\tau}(u_\tau) = \begin{cases} \emptyset & \tau = \emptyset \\ P_{U_1}(u_1) & \tau = \{1\} \\ P_{U_2}(u_2) & \tau = \{2\} \\ P_{\underline{U}}(\underline{u}) & \tau = \{1,2\} \end{cases}, \mathcal{U}_\tau = \begin{cases} \emptyset & \tau = \emptyset \\ \mathcal{U}_1 & \tau = \{1\} \\ \mathcal{U}_2 & \tau = \{2\} \\ \mathcal{U}_1 \times \mathcal{U}_2 & \tau = \{1,2\} \end{cases}, \tag{1.44}$$

when U_1 and U_2 are correlated to each other.

Chapter 2

Single User Communication

As mentioned in Chapter 1, for point-to-point communication, many studies show that joint source channel coding might be expected to have a better error exponent than separate source channel coding [5, 7, 8, 16]. To have a better insight about error exponent for single-user communication, we quickly review some previous works.

Since finding an exact expression for error probability is very difficult, many works investigated upper and lower bounds on the average error probability, or equivalently lower and upper bounds for the error exponent. Using random-coding technique leads (1.13) as a lower bound of error exponent [5, Ch. 5], [4, Ch. 9]. However, since at low rates, the error probability of poor codes in the ensemble dominates the average error probability, the performance of (1.13) at low rate is weak [5, Sec. 5.7]. In [5], through an expurgation process, a tight exponent at low rates was proposed [5, Eq. 5.7.10].

In fact, both random-coding and expurgating methods gives lower bounds on the error exponent. On the other hand, sometimes finding an upper bound for error exponent satisfied by every code is challenging. Generally, hypothesis-testing method [17] is utilized to derive upper bound for error exponent. The two well-known upper bounds of error exponent are sphere-packing exponent [18] and minimum-distance exponent [19].

For the rates greater than critical rate [5, Sec. 5.6], the random-coding and sphere-packing bounds are coincide to each other, while the expurgate and minimum-distance bounds are coincide at rate $R = 0$. In terms of error exponent, at low rates, an upper bound known as straight-line bound has better performance than sphere-packing bound. The straight-line bound is obtained by connecting any two points of sphere-packing and minimum-distance bounds. Thus, for the rates smaller than critical rate, the error exponent is greater than random-coding exponent, and is smaller than sphere-packing exponent.

As mentioned in Chapter 1, considering joint source-channel coding, the random-coding exponent is derived as (1.19). However, by partitioning message sets into source-type classes and assigning constant-composition codes [20], to map messages within a source type onto sequences within a channel-input type, the following achievable exponent is derived [8]

$$\min_{H(U)\leq R\leq \log|\mathcal{U}|} \sup_{\rho\geq 0} \{\rho R - E_s(\rho, P_U)\} + \max_{\rho\in[0,1]} \left\{\max_{Q} E_0(\rho, Q, W) - \rho R\right\}. \quad (2.1)$$

In addition, in [8, Lemma 2], the sphere-packing bound on the exponent is obtained as

$$\min_{H(U)\leq R\leq \log|\mathcal{U}|} \sup_{\rho\geq 0} \{\rho R - E_s(\rho, P_U)\} + \max_{\rho\geq 0} \left\{\max_{Q} E_0(\rho, Q, W) - \rho R\right\}. \quad (2.2)$$

By applying Fenchel's duality, it was shown that [7]

$$\begin{aligned} \min_{H(U)\leq R\leq \log|\mathcal{U}|} \sup_{\rho\geq 0} \{\rho R - E_s(\rho, P_U)\} + \max_{\rho\in[0,1]} \left\{\max_{Q} E_0(\rho, Q, W) - \rho R\right\} \\ = \max_{\rho\in[0,1]} \bar{E}_0(\rho, \mathcal{Q}, W) - E_s(\rho, P_U), \end{aligned} \quad (2.3)$$

where, $\bar{E}_0$ is the point-wise supremum over all convex combinations of any two values of the function $E_0(\rho, \mathcal{Q}, W)$, i. e.

$$\bar{E}_0(\rho, \mathcal{Q}, W) \triangleq \sup_{\substack{\rho_1,\rho_2,\theta\in[0,1]:\\ \theta\rho_1+(1-\theta)\rho_2=\rho}} \Big\{\theta E_0(\rho_1, \mathcal{Q}, W) + (1-\theta)E_0(\rho_2, \mathcal{Q}, W)\Big\}, \quad (2.4)$$

and $\mathcal{Q}$ is a set of distributions.

Finally, in [16], it is proved that joint source-channel random coding where source messages are assigned to different classes and codewords are generated according to a distribution that depends on the class index of source message, achieves the following exponent

$$\max_{\rho\in[0,1]} \bar{E}_0(\rho, \mathcal{Q}, W) - E_s(\rho, P_U), \quad (2.5)$$

which coincides with the sphere-packing exponent [8, Lemma 2] whenever it is tight.

To summarize the results, using codewords with a composition dependent on the source message leads to a better exponent than the case where codewords are drawn according to a fixed product distribution [8]. In addition, considering the scheme where source messages are assigned to disjoint classes and encoded by codes that depend on the class index, attains the sphere-packing exponent in those cases where it is tight [16].

In this chapter, we study the random-coding exponent of joint source channel coding under various ensembles. In Section 2.1, we consider the scheme where codewords are generated according to a conditional distribution that depends on the instantaneous source symbol. However, in Section 2.2, we simplify the coding scheme to the case where codewords and messages are statistically independent of each other. Finally, in Section 2.3, by comparing the results obtained in Sections 2.1 and 2.2, we show that there is no penalty in the error exponent if the messages are assumed to be statistically independent of the codewords.

Throughout this chapter, to distinguish between conditional input distribution and input distribution independent of the source probability, we denote $\bar{Q}$ as the conditional input distribution and Q as the marginal input distribution.

2.1 Statistical Dependency between the Messages and Coedwords

As mentioned in Chapter 1, for the MAC with correlated sources, considering statistical dependency between the messages and codewords leads to a larger error exponent. To examine whether this idea can also improve the error exponent of single-user communication or not, we study various ensembles. In all studied ensembles in this section, we assume that codewords are generated by a conditional probability distribution of the codeword symbol that depends on the instantaneous source symbol.

2.1.1 Message Dependent Ensemble with Statistical Dependency

Here, we derive an achievable random-coding error exponent for joint source channel coding where codebooks are generated by a conditional probability distribution of the codeword symbol that depends both on the instantaneous source symbol and on the type of the source sequence. In other words, for every message $\boldsymbol{u} \in \mathcal{U}^n$, we randomly generate a codeword $\boldsymbol{x}(\boldsymbol{u})$ according to the probability distribution $\bar{Q}^n_{\pi(\boldsymbol{u})}(\boldsymbol{x}|\boldsymbol{u}) = \prod_{j=1}^n \bar{Q}_{\pi(\boldsymbol{u})}(x_j|u_j)$, where $\bar{Q}_{\pi(\boldsymbol{u})}$ is a conditional probability distribution that depends on the type of $\boldsymbol{u}$, denoted by $\pi(\boldsymbol{u})$.

Proposition 2.1. *For a given channel* W *and source* P_U, E^{mds} *is an achievable exponent where*

$$E^{\text{mds}} = \min_{\hat{P}_{UXY} \in \mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U \bar{Q}_{\hat{P}_U} W) + \Big[D(\hat{P}_{UXY}||\hat{P}_U \bar{Q}_{\hat{P}_U} \hat{P}_Y) - H(\hat{P}_U)\Big]^+, \quad (2.6)$$

and $[x]^+ = \max\{0, x\}$.

Proof. See Section 2.4.1. □

In Section 2.2.1, we will study a simplified version of (2.6) known as message-dependent exponent. By showing that message-dependent exponent is equal with (2.6), we will prove that removing statistical dependency between the messages and codewords, will not effect on the error exponent.

2.1.2 Independent and Conditionally Distributed Ensemble

In this section, we study icd ensemble, a simpler ensemble than the one described in Section 2.1.1. Here, statistically codewords depend on the source messages, and are generated independently according to a conditional distribution denoted by $\bar{Q}$.

In fact, we randomly generate a codeword $\boldsymbol{x}(\boldsymbol{u}) \in \mathcal{X}^n$ according to the conditional probability distribution $\bar{Q}^n(\boldsymbol{x}|\boldsymbol{u}) = \prod_{j=1}^n \bar{Q}(x_j|u_j)$. By setting $\bar{Q}_{\hat{P}_U}(x|u) = \bar{Q}(x|u)$ in Proposition 2.1, it can be proved that the following exponent is achievable

$$E^{\text{icd}} = \min_{\hat{P}_{UXY} \in \mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U \bar{Q} W) + \Big[D(\hat{P}_{UXY}||\hat{P}_U \bar{Q} \hat{P}_Y) - H(\hat{P}_U)\Big]^+, \quad (2.7)$$

where $\hat{P}_U$ and $\hat{P}_Y$ are marginal distributions of the $\hat{P}_{UXY}$.

Proposition 2.2. *The optimal joint distribution* $\hat{P}_{UXY}$ *minimizing* (2.7)*, is given by*

$$\hat{P}^*_{UXY}(u,x,y) = \frac{P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(y|\bar{x})^{\frac{1}{1+\rho}}\Big)^{\rho}}{\sum_{\bar{y}} \Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\Big)^{1+\rho}}, \quad (2.8)$$

where in view of (2.8)*, an equal expression of* (2.7) *can be expressed as*

$$E^{\text{icd}} = \max_{\rho \in [0,1]} E_{0,s}(\rho, P_U, \bar{Q}, W), \tag{2.9}$$

where

$$E_{0,s}(\rho, P_U, \bar{Q}, W) = -\log \Big(\sum_y \Big(\sum_x \sum_u P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} \Big)^{1+\rho} \Big). \tag{2.10}$$

Proof. See Section 2.4.2. □

2.1.3 Conditional Constant-Composition Ensemble

In this section, we describe conditional constant-composition ensemble which can be considered as a generalization of the constant-composition ensemble. Consider a discrete memoryless source which is characterized by a distribution $P_{\boldsymbol{U}}^n(\boldsymbol{u}) = \prod_{k=1}^n P_U(u_k)$. For a given message $\boldsymbol{u} = (u_1, u_2, ..., u_n)$, we consider the sub-sequences of $\boldsymbol{u}$ which have the same symbols. We define $j_u(\boldsymbol{u})$ as the set of all positions where the symbol u appears in $\boldsymbol{u}$, i.e. for all $u \in \mathcal{U}$

$$j_u(\boldsymbol{u}) = \{i \in \{1, 2, ..., n\}, \text{such that } u_i = u\}. \tag{2.11}$$

The subsequence can be represented by $\boldsymbol{u}(j_u(\boldsymbol{u}))$. Let $\bar{Q}(x|u)$ be a conditional input distribution. We approximate the conditional distribution $\bar{Q}(x|u)$ with a type-p conditional distribution $\bar{Q}_p$ that satisfies

$$\bar{Q}_p(x|u) \in \left\{0, \frac{1}{p}, \frac{2}{p}, ..., 1\right\}, \tag{2.12}$$

for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$. We set p depends on u and $\boldsymbol{u}$, $p = |j_u(\boldsymbol{u})|$ and choose this distribution such that the variational distance between $\bar{Q}$ and $\bar{Q}_p$ satisfies

$$\left|\bar{Q}_p(x|u) - \bar{Q}(x|u)\right| < \frac{1}{p}. \tag{2.13}$$

For every $u \in \mathcal{U}$, we randomly pick a sequence $\boldsymbol{x}_u$ of length $|j_u(\boldsymbol{u})|$ from the set $\mathcal{T}_u^p(\bar{Q}_p)$ and set $\boldsymbol{x}(j_u(\boldsymbol{u})) = \boldsymbol{x}_u$. Codebook $\mathcal{C}$ is called a conditional constant-composition codebook with distribution $\bar{Q}$, if $\boldsymbol{x}(\boldsymbol{u}) = (\boldsymbol{x}_u)_{u \in \mathcal{U}}$, where $\boldsymbol{x}_u \in \mathcal{T}_u^p(\bar{Q}_p)$.

Now, we apply constant-composition random coding to determine an achievable exponent.

Proposition 2.3. *Consider a conditional constant-composition ensemble generated by conditional distribution $\bar{Q}$. For the point-to-point channel with transition probability W and source probability distribution P_U, the following exponent is achievable*

$$E^{\text{ccc}} = \min_{\hat{P}_{UXY} \in \mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \Big[I\left(U, X; Y\right) - H(\hat{P}_U) \Big]^+, \qquad (2.14)$$

where the set S^{ccc} is defined as

$$\mathcal{S}^{\text{ccc}} \triangleq \Big\{ \hat{P}_{UXY} : \hat{P}_{UXY} = \hat{P}_U \bar{Q} \hat{P}_{Y|XU}, \hat{P}_U \in \mathcal{P}_{\mathcal{U}}, \hat{P}_{Y|XU} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}\times\mathcal{U}} \Big\}. \qquad (2.15)$$

Proof. One simple way to prove Proposition 2.3, is using Proposition 2.1. In fact, for conditional constant-composition ensemble, we ought to consider only $\hat{P}_{UXY}$ that their conditional distribution $\hat{P}_{X|U}$ is $\bar{Q}$. By applying the identity $\hat{P}_{UXY} = \hat{P}_U \bar{Q} \hat{P}_{Y|X}$ in $D(\hat{P}_{UXY}||P_U\bar{Q}W)$ in (2.6), noticing the fact that $D(\hat{P}_U \bar{Q} \hat{P}_{Y|XU}||\hat{P}_U \bar{Q} \hat{P}_Y) = I(U, X; Y)$, (2.6) is derived as

$$E^{\text{ccc}} = \min_{\hat{P}_{UXY} \in \mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \Big[I\left(U, X; Y\right) - H(\hat{P}_U) \Big]^+, \qquad (2.16)$$

which proves Proposition 2.3. The long way is very similar to the proof presented in Section 4.3.9. □

Next, we determine an equivalent expression for the derived exponent in (2.14). We continue by proving the following proposition.

Proposition 2.4. *An equivalent dual expression for the exponent given in* (2.14) *can be expressed as*

$$E^{\text{ccc}} = \max_{\rho \in [0,1]} E_{0,s}^{\text{ccc}}(\rho, P_U, \bar{Q}, W), \qquad (2.17)$$

where

$$E_{0,s}^{\text{ccc}}(\rho, P_U, \bar{Q}, W) = \max_{\substack{\bar{\beta}(u,x): \\ \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0}} -\log \left(\sum_y \left(\sum_{u,x} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right). \qquad (2.18)$$

Proof. See Section 2.4.3. □

2.2 Source-Channel Coding without Statistical Dependency

As mentioned in Chapter 1, joint source-channel coding leads larger exponent than separate source-channel coding. In this section, by considering joint source-channel coding, we re-obtain achievable exponents for message-dependent, iid and generalized constant-composition ensembles.

2.2.1 Message-Dependent Ensemble

Message-dependent ensemble was studied in [16] and [21] and the derivation of achievable exponent were done in the dual domain. Here, by doing the analysis in the primal domain and applying the Lagrange duality theory, we re-obtain the message-dependent random-coding exponent obtained in [16].

In this section, we consider a simplified version of the ensemble described in Section 2.1.1. In fact, we consider a case where codebooks are generated according to a distribution that only depends on the type of the whole source sequence. For every message $\boldsymbol{u} \in \mathcal{U}^n$, codeword $\boldsymbol{x}(\boldsymbol{u})$ is generated randomly according to the probability distribution $Q^n_{\pi(\boldsymbol{u})}(\boldsymbol{x}) = \prod_{j=1}^n Q_{\pi(\boldsymbol{u})}(x_j)$, where $Q_{\pi(\boldsymbol{u})}$ is a probability distribution that depends on the type of $\boldsymbol{u}$, denoted by $\pi(\boldsymbol{u})$.

Proposition 2.5. *For point-to-point channel W with source P_U, an achievable random-coding exponent for joint source channel coding is*

$$E^{\mathrm{md}} = \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} D(\hat{P}_U || P_U) + \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \Big[D(\hat{P}_{XY} || Q_{\hat{P}_U} \hat{P}_Y) - H(\hat{P}_U) \Big]^+, \qquad (2.19)$$

where $[x]^+ = \max\{0, x\}$.

Proof. Let in Proposition 2.1, condewords are generated according to $Q_{\hat{P}_U}(x)$ rather than $\bar{Q}_{\hat{P}_U}(x|u)$. By setting $\bar{Q}_{\hat{P}_U}(x|u) = Q_{\hat{P}_U}(x)$ in Proposition 2.1, and noting to the fact that when there is no statistical dependency between random variables U and X, we have

$$D(\hat{P}_{UXY} || P_U Q_{\hat{P}_U} W) = D(\hat{P}_U || P_U) + D(\hat{P}_{XY} || Q_{\hat{P}_U} W), \qquad (2.20)$$

Proposition 2.5 is proved. □

Now, to determine an equivalent expression for the derived exponent given in (2.19), we start by proving the following Lemma.

Lemma 2.1. *Let* $0 = \gamma_{L+1} \le \gamma_L \le \dots \le \gamma_2 \le \gamma_1 = 1$ *be* $L+1$ *positive ordered numbers such that* $\gamma_L > \min P_U(u)$ *and* $\gamma_2 \le \max P_U(u)$*. For each* $\hat{P}_U \in \mathcal{P}_\mathcal{U}$*, there exists a unique* $\ell \in \{1, \dots, L\}$ *such that*

$$\log(\gamma_{\ell+1}) < \sum_u \hat{P}_U(u) \log(P_U(u)) \le \log(\gamma_\ell). \tag{2.21}$$

Proof. See Section 2.4.4. □

In fact, Lemma 2.1 describes a situation where the set $\mathcal{P}_\mathcal{U}$ is partitioned by L classes given by (2.21) and the classes are indexed by ℓ. Roughly speaking, the equivalent expression of Lemma 2.1 in the dual domain is that the source-message set $\mathcal{U}^n$ is partitioned into L classes where the class $\ell \in \{1, \dots, L\}$ is

$$\mathcal{D}^\ell = \Big\{ \boldsymbol{u} \in \mathcal{U}^n \,:\, \gamma_{\ell+1}^n < P_{\boldsymbol{U}}^n(\boldsymbol{u}) \le \gamma_{\ell_\nu}^n \Big\}. \tag{2.22}$$

Now, we consider L input distributions as $\{Q_1, \dots, Q_L\}$, and we assign the distribution Q_ℓ to the class ℓ. Briefly, in primal domain, if $\hat{P}_U$ belongs to the class ℓ, then we let $Q_{\hat{P}_U} = Q_\ell$. Given $\ell \in \{1, \dots, L\}$, depending on the class index of $\hat{P}_U$, the input distribution is determined. In addition, the interpretation in the dual domain is that for the messages belonging to $\mathcal{D}^\ell$, input distribution Q_ℓ is assigned to generate codewords.

Proposition 2.6. *Consider* L *input distributions as* $\{Q_1, \dots, Q_L\}$ *and let* $0 = \gamma_{L+1} \le \gamma_L \le \dots \le \gamma_2 \le \gamma_1 = 1$ *be* $L+1$ *positive ordered numbers such that* $\gamma_L > \min P_U(u)$ *and* $\gamma_2 \le \max P_U(u)$*. An equivalent dual expression for the exponent given in* (2.19) *can be expressed as*

$$E^{\mathrm{md}} = \max_{\gamma_2, \dots, \gamma_L} \min_{\ell \in \{1, \dots, L\}} \max_{\rho \in [0,1]} E_0(\rho, Q_\ell, W) - E_{s,\ell}(\rho, P_{U_\nu}, \gamma_{\ell+1}, \gamma_\ell), \tag{2.23}$$

where $E_0(\cdot)$ *given by* (1.14)*, and*

$$E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell) = \begin{cases} E_s(\rho_{\gamma_{\ell+1}}, P_U) + E_s'(\rho_{\gamma_{\ell+1}})(\rho - \rho_{\gamma_{\ell+1}}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}}, \\ E_s(\rho, P_{U_\nu}) & \frac{1}{1+\rho_{\gamma_{\ell+1}}} \le \frac{1}{1+\rho} \le \frac{1}{1+\rho_{\gamma_\ell}}, \\ E_s(\rho_{\gamma_\ell}, P_U) + E_s'(\rho_{\gamma_\ell})(\rho - \rho_{\gamma_\ell}) & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_\ell}}, \end{cases} \tag{2.24}$$

where in (2.24)*,* $E_s'(\rho_\gamma) = \frac{\partial E_s(\rho, P_U)}{\partial \rho}\Big|_{\rho = \rho_\gamma}$*. For* $0 = \gamma_{L+1} \le \gamma_L \le \dots \le \gamma_2 < \gamma_1 = 1$*, the parameter* ρ_{γ_ℓ} *for* $\ell = 1, \dots, L$ *is the solution of the implicit equation*

$$\frac{\sum_u P_U(u)^{\frac{1}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1}{1+\rho}}} = \log(\gamma_\ell), \tag{2.25}$$

when $\min_u P_U(u) \le \gamma_\ell \le \max_u P_U(u)$ *is satisfied. However, when* $\gamma_\ell \ge \max_u P_U(u)$, $\rho_{\gamma_\ell} = -1_+$ *and when* $\gamma_\ell \le \min_u P_U(u)$, $\rho_{\gamma_\ell} = -1_-$ *[21].*

Proof. See Section 2.4.5. □

We remark that since we considered L classes an L input distributions, there are L^L possible assignments. In (2.23), the optimal assignment of input distributions to source classes is considered.

Assuming $L = 3$, for $\ell = 1, 2, 3$, Figure 2.1 shows $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ defined in (2.24), versus ρ where in view of Proposition 2.6, we have $\gamma_1 = 1$ and $\gamma_4 = 0$. Consider $E_{s,2}(\cdot)$, for an interval of ρ, $E_{s,2}(\cdot)$ given in (2.24), is a tangent line to $E_s(\cdot)$ function; however, for the ρs between ρ_{γ_2} and ρ_{γ_3} is the Gallager source function.

The reason why straight lines are appeared is explained in the following. Firstly, we recall the primal form of $E_s(\cdot)$ function, i. e.

$$-E_s(\rho, P_U) = \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} D\left(\hat{P}_U || P_U\right) - H\left(\hat{P}_U\right). \tag{2.26}$$

Now, consider the empirical distribution $\hat{P}_U$. From the proof of Lemma 2.1, it can be verified that for $\boldsymbol{u} \in \mathcal{T}^n(\hat{P}_U)$, we have $P^n_{\boldsymbol{U}}(\boldsymbol{u}) = e^{n\sum_u \hat{P}_U(u)\log(P_U(u))}$. Noting that all messages in $\mathcal{T}^n(\hat{P}_U)$ have the same probability, there exist a $\gamma \in [0, 1]$ such that $\gamma^n = P^n_{\boldsymbol{U}}(\boldsymbol{u}) = e^{n\sum_u \hat{P}_U(u)\log(P_U(u))}$, or equivalently

$$\sum_u \hat{P}_U(u) \log(P_U(u)) = \log(\gamma). \tag{2.27}$$

Roughly speaking, (2.27) represents the primal form of an empirical distribution. Thus, the minimization problem of (2.26) with the constraint given by (2.27), can be interpreted as the the Gallager's source function when the empirical distribution of the source is fixed.

Applying Lagrange duality theory to (2.26) with the constrain given by (2.27), we find that

$$-\hat{E}_s(\rho, P_U, \gamma) = \min_{\substack{\hat{P}_U \in \mathcal{P}_\mathcal{U}: \\ \sum_u \hat{P}_U(u)\log(P_U(u)) = \log(\gamma)}} D\left(\hat{P}_U || P_U\right) - H\left(\hat{P}_U\right), \tag{2.28}$$

where

$$\hat{E}_s(\rho, P_U, \gamma) = E_s\left(\rho_\gamma, P_U\right) + E'_s\left(\rho_\gamma\right)\left(\rho - \rho_\gamma\right), \tag{2.29}$$

is the straight line tangent to $E_s(\cdot)$ function at ρ_γ and ρ_γ is derived by (2.25) when $\gamma_\ell = \gamma$.

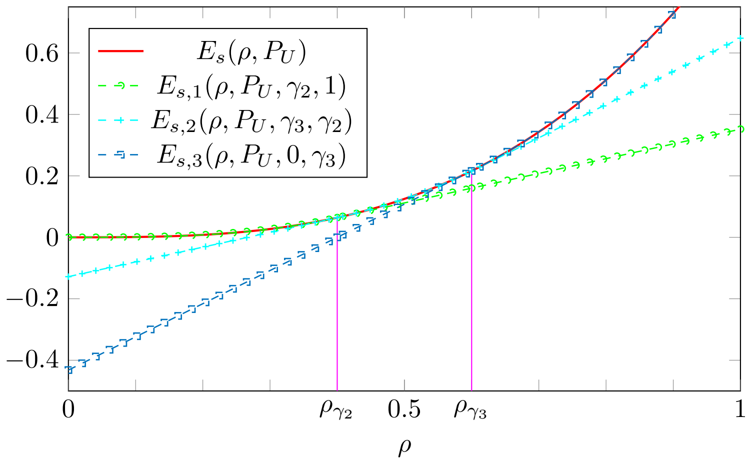

Figure 2.1: $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$.

Next, as a special case, we partition $\mathcal{P}_{\mathcal{U}}$ into two classes, i.e. $L = 2$. In fact, the two-class partitioning was widely studied in [16]. Here, to emphasize two-class partitioning, we rename the set $\mathcal{D}^\ell$ when $\ell \in \{1, 2\}$ as $\mathcal{A}^i(\gamma)$ for $i = 1, 2$. In other words, setting $\gamma_3 = 0$, $\gamma_2 = \gamma$ and $\gamma_1 = 1$ in Lemma 2.1, the classes $\mathcal{D}^1$ and $\mathcal{D}^2$ are respectively renamed as

$$\mathcal{A}^1(\gamma) = \Big\{\boldsymbol{u} \in \mathcal{U}^n \ : \ P^n_{\boldsymbol{U}}(\boldsymbol{u}) \geq \gamma\Big\}, \tag{2.30}$$

$$\mathcal{A}^2(\gamma) = \Big\{\boldsymbol{u} \in \mathcal{U}^n \ : \ P^n_{\boldsymbol{U}}(\boldsymbol{u}) < \gamma\Big\}, \tag{2.31}$$

where $0 \leq \gamma \leq 1$ is a fixed threshold.

As mentioned, (2.30) and (2.31) are the interpretations of Lemma 2.1 in the dual domain. It means that for the messages belonging to the class $\mathcal{A}^i(\gamma)$, codewords are generated according to the input distribution Q_i, where $i = 1, 2$.

Setting $\gamma_3 = 0$, $\gamma_2 = \gamma$ and $\gamma_1 = 1$ in Proposition 2.6, the achievable exponent (2.23), when two classes are assigned is given by

$$E^{\text{md}}(P_U, \{Q_1, Q_2\}, W) = \max_{\gamma\in[0,1]} \min_{i\in\{1,2\}} \max_{\rho\in[0,1]} E_0(\rho, Q_i, W) - E_{s,i}(\rho, P_U, \gamma), \tag{2.32}$$

where from (2.24), by setting $\gamma_3 = 0$, $\gamma_2 = \gamma$ and $\gamma_1 = 1$, we have

$$E_{s,1}(\rho, P_U, \gamma) = \begin{cases} E_s(\rho, P_U) & \frac{1}{1+\rho} \geq \frac{1}{1+\rho_\gamma}, \\ E_s(\rho_\gamma, P_U) + E_s'(\rho_\gamma)(\rho - \rho_\gamma) & \frac{1}{1+\rho} < \frac{1}{1+\rho_\gamma}, \end{cases} \tag{2.33}$$

and

$$E_{s,2}(\rho, P_U, \gamma) = \begin{cases} E_s(\rho, P_U) & \frac{1}{1+\rho} < \frac{1}{1+\rho_\gamma}, \\ E_s(\rho_\gamma, P_U) + E_s'(\rho_\gamma)(\rho - \rho_\gamma) & \frac{1}{1+\rho} \geq \frac{1}{1+\rho_\gamma}. \end{cases} \quad (2.34)$$

In (2.33) and (2.34), the parameter ρ_γ is the solution of the implicit equation given by (2.25) when $\min_u P_U(u) \leq \gamma \leq \max_u P_U(u)$ is satisfied. We observe that $E_{s,1}(\rho, \cdot)$ follows the Gallager $E_s(\rho, \cdot)$ function for an interval of ρ, while it is the straight line tangent to $E_s(\rho, \cdot)$ beyond that interval, and similarly for $E_{s,2}(\rho, \cdot)$.

When $\gamma \in [0, \min_u P_U(u))$, we have that $\rho_\gamma = -1_-$ and hence $E_{s,1}(\rho, \cdot) = E_s(\rho, \cdot)$ and $E_{s,2}(\rho, \cdot) = -\infty$. Otherwise, when $\gamma \in (\max_u P_U(u), 1]$, we have that $\rho_\gamma = -1_+$ and hence $E_{s,1}(\rho, \cdot) = -\infty$ and $E_{s,2}(\rho, \cdot) = E_s(\rho, \cdot)$. In our analysis, it suffices to consider $\gamma = 0$ or $\gamma = 1$ to represent the cases where $E_{s,1}(\rho, \cdot)$ or $E_{s,2}(\rho, \cdot)$ are infinity. For such cases, we have

$$E_{s,1}(\rho, P_U, 0) = E_s(\rho, P_U), \quad E_{s,2}(\rho, P_U, 0) = -\infty, \quad (2.35)$$

$$E_{s,1}(\rho, P_U, 1) = -\infty, \quad E_{s,2}(\rho, P_U, 1) = E_s(\rho, P_U). \quad (2.36)$$

It has been shown that partitioning the source messages into two classes leads larger exponent than (1.19); however, there is no benefit in terms of error exponent, if we partition the source message-set into more than two classes [16]. Here, we present another proof for this fact.

Proposition 2.7. *For the optimal assignment of three input distributions* $\{Q_1, Q_2, Q_3\}$*, we denote* E^{md} *given by* (2.23) *as* $E^{\mathrm{md}}(P_U, \{Q_1, Q_2, Q_3\}, W)$*. Then, we have*

$$E^{\mathrm{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) = \max \Big\{ E^{\mathrm{md}}(P_U, \{Q_1, Q_2\}, W), \\ E^{\mathrm{md}}(P_U, \{Q_2, Q_3\}, W), E^{\mathrm{md}}(P_U, \{Q_1, Q_3\}, W) \Big\}, \quad (2.37)$$

where the right hand side of (2.37) *is defined by* (2.32).

Proof. See Section 2.4.6. □

Since it was proved that in terms of error exponent, partitioning the source messages into two classes is sufficient, from now we focus on (2.32). In view of (2.33) and (2.34), we recall that $E_{s,i}(\cdot)$ for $i = 1, 2$ is either $E_s(\cdot)$ function given by (1.9), or $\hat{E}_s(\cdot)$ given by (2.29).

To find the optimal γ maximizing (2.32), we use Lemma A.6. Setting $E(\rho, Q_1) = E_0(\rho, Q_1, W)$ and $E(\rho, Q_2) = E_0(\rho, Q_2, W)$ in Lemma A.6, the optimal $\gamma^\star$ maximizing (2.32), is obtained at the point where $\max_\rho E_0(\cdot, Q_1) - E_{s,1}(\cdot)$ equals to $\max_\rho E_0(\cdot, Q_2) - E_{s,2}(\cdot)$. Figure 2.2 shows this equality where ρ_1 and ρ_2 are given by

$$\rho_i = \arg\max_{\rho\in[0,1]} E_0(\rho, Q_i, W) - E_{s,i}(\rho, P_U, \gamma^\star), \tag{2.38}$$

for $i = 1, 2$. In view of Lemma A.6, from (A.37) we conclude that

$$E^{\text{md}}(P_U, \{Q_1, Q_2\}, W) = \max_{\rho\in[0,1]} \bar{E}_0(\rho, \{Q_1, Q_2\}, W) - E_s(\rho, P_U), \tag{2.39}$$

where $\bar{E}_0(\cdot)$ is given by (2.4). Since $\bar{E}_0(\cdot)$ is greater than the Gallager's channel function, we conclude that message-dependent exponent given by (2.32) or (2.39) is greater than iid exponent presented in (1.19).

Coming back to Figure 2.2, we note that both ρ_1 and ρ_2 derived by (2.38), are located on the straight line and also on the both sides of ρ_γ. Otherwise, the message-dependent exponent will equal to iid exponent.

For example, Figure 2.3 shows an example where both ρ_1 and ρ_2 are located one side of ρ_γ. As a result, one of them is located on the $E_s(\cdot)$ function. In the example of Figure 2.3, ρ_2 is located on E_s function. Hence, from (A.36) we have

$$E_0(\rho_1, Q_1, W) - E_{s,1}(\rho, P_U, \gamma^\star) = E_0(\rho_2, Q_2, W) - E_s(\rho, P_U), \tag{2.40}$$

where the right hand side of (2.40) is Gallager exponent given by (1.19).

Briefly, there are many sources and channels that the message-dependent exponent for them equals to their iid exponent. However, there are also examples that for them (2.32) is greater than (1.19). Here, we bring one of them. We consider a discrete memoryless source P_U with alphabet $\mathcal{U} = \{1, 2\}$ where $P_U(1) = 0.028$ and $P_U(2) = 0.972$. We also consider a discrete memoryless channel with $\mathcal{X} = \{1, ..., 6\}$ and $\mathcal{Y} = \{1, ..., 4\}$. The transition probability of this channel, denoted as W where

$$W = \begin{pmatrix} 1-3k_1 & k_1 & k_1 & k_1 \\ k_1 & 1-3k_1 & k_1 & k_1 \\ k_1 & k_1 & 1-3k_1 & k_1 \\ k_1 & k_1 & k_1 & 1-3k_1 \\ 0.5-k_2 & 0.5-k_2 & k_2 & k_2 \\ k_2 & k_2 & 0.5-k_2 & 0.5-k_2 \end{pmatrix}, \tag{2.41}$$

for $k_1 = 0.056$ and $k_2 = 0.01$. Considering two input distributions as

$$Q_1 = [0\ 0\ 0\ 0\ 0.5\ 0.5], \tag{2.42}$$
$$Q_2 = [0.25\ 0.25\ 0.25\ 0.25\ 0\ 0], \tag{2.43}$$

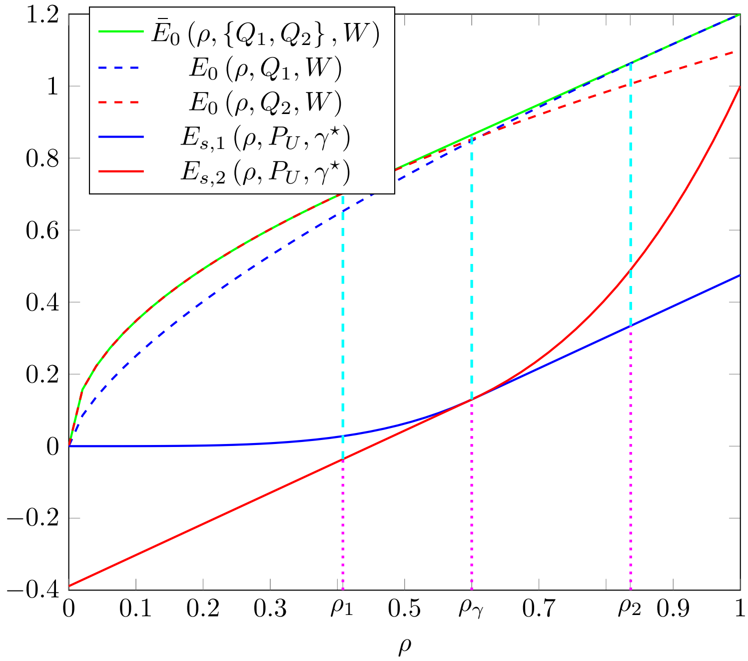

Figure 2.2: Message-dependent random coding ensemble gives larger exponent than iid ensemble.

from (2.32) we find that $\gamma^\star = 0.6541$ and

$$E^{\mathrm{md}} = 0.1734, \tag{2.44}$$

$$E^{\mathrm{iid}} = 0.1721, \tag{2.45}$$

where E^{iid} is given by (1.19) for the best assignment of input distributions. As can be seen, for this example, message-dependent random coding exponent is larger than iid exponent.

2.2.2 iid Random-Coding Exponent

In Section 1.2.1.2, we briefly reviewed the iid random coding exponent. Here, we bring the results in both primal and dual domain. For joint source-channel coding, by drawing the codewords independently of the source messages according to an identical product distribution $Q^n(\boldsymbol{x}) = \prod_{j=1}^n Q(x_j)$, and using random-coding argument, the following exponent is achievable

$$E^{\mathrm{iid}} = \min_{\hat{P}_U \in \mathcal{P}_{\mathcal{U}}} D(\hat{P}_U || P_U) + \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY} || QW) + \Big[D(\hat{P}_{XY} || Q\hat{P}_Y) - H(\hat{P}_U) \Big]^+, \tag{2.46}$$

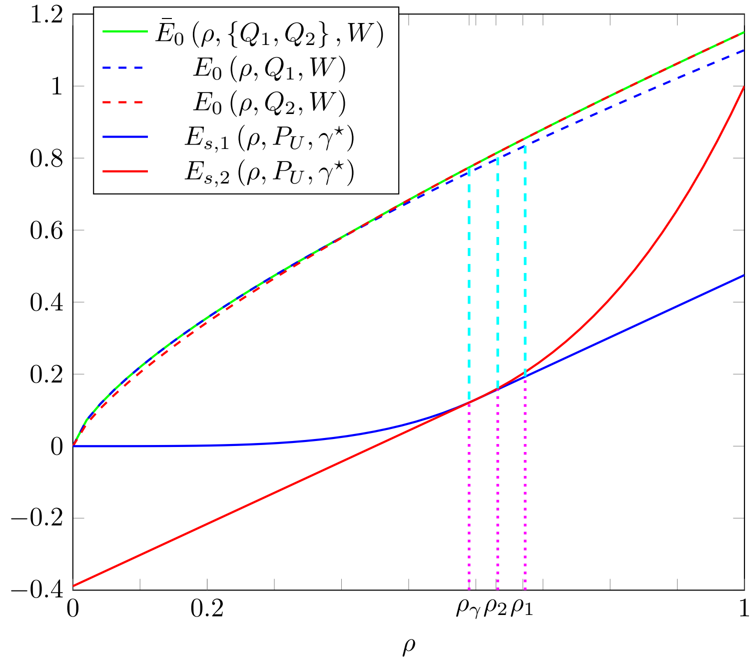

Figure 2.3: Message-dependent random coding ensemble gives the same exponent as iid ensemble.

where by setting $Q(x)$ instead of $Q_{\hat{P}_U}(x)$ in Proposition 2.5, equation (2.46) can be proved. Next,by applying Lagrange duality theory, we find the dual form of (2.46).

Proposition 2.8. *The optimal distributions which minimize the objective function in* (2.46) *are given by*

$$\hat{P}^\star_{XY}(x,y)=\frac{Q(x)W(y|x)^{\frac{1}{1+\rho}}\left(\sum_{\bar{x}}Q(\bar{x})W(y|\bar{x})^{\frac{1}{1+\rho}}\right)^{\rho}}{\sum_{\bar{y}}\left(\sum_{\bar{x}}Q(\bar{x})W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}},\tag{2.47}$$

and

$$\hat{P}^\star_{U}(u)=\frac{P_U(u)^{\frac{1}{1+\rho}}}{\sum_{\bar{u}}P_U(\bar{u})^{\frac{1}{1+\rho}}}.\tag{2.48}$$

In addition, inserting (2.47) *and* (2.48) *into* (2.46)*, an equivalent expression of* (2.46) *is Gallager's exponent [5, Prob. 5.16] and is given by*

$$E^{\text{iid}}=\max_{\rho\in[0,1]}E_0(\rho,Q,W)-E_s(\rho,P_U),\tag{2.49}$$

where $E_s(\rho,P_U)$ and $E_0(\rho,Q,W)$ are given in (1.9) *and* (1.14)*, respectively.*

Proof. The proof of (2.49) was given in many literature, for example [5, Sec. 5.6]. However, here we apply Lagrange duality theory to (2.46) with constraints that $\sum_u \hat{P}_U(u) = 1$ and also $\sum_{x,y} \hat{P}_{XY}(x,y) = 1$. To avoid repeating, we use the the result of Proposition 2.6. Since for the given problem in (2.46), the input distribution Q does not depend on the source type, only one input distribution is considered. As a result, in (2.23), the number of classes is only one. By inserting $L = 1$ in (2.23), (2.49) is proved.

More precisely, since only one input distribution is considered, in the proof of Proposition 2.6 in Section 2.4.5, the two inequality constrains given in (2.144) and (2.145) are inactive and thus their corresponding Lagrangian coefficients (λ_ℓ and $\lambda_{\ell+1}$) are zero. By inserting Q instead of $Q_{\hat{P}_U}$ and applying $\lambda_\ell = \lambda_{\ell+1} = 0$ in (2.152) in view of (2.160), Proposition 2.8 is proved. □

2.2.3 Generalized Constant-Composition Exponent

Constant-composition ensemble was widely studied in [15] and [20]. It has been shown that constant-composition random-coding exponent is larger than iid random-coding exponent [15], [20].

Moreover, as mentioned before, message-dependent random-coding exponent is larger than iid random-coding exponent [16]. Here, we merge the idea of message-dependent and constant-composition ensembles, and we present an achievable exponent for a generalization of the constant composition ensemble.

We consider a case where codebooks are generated according to a type distribution that depends on the composition of the whole source sequence. In other words, for a given source message composition $\hat{P}_U \in \mathcal{P}^n_{\mathcal{U}}$, $Q_{\hat{P}_U} \in \mathcal{P}_{\mathcal{X}}$ is fixed and can be approximated by the type distribution $Q_{n,\hat{P}_U}$, where $Q_{n,\hat{P}_U}(x) \in \left\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\right\}$. The variational distance between $Q_{\hat{P}_U}$ and $Q_{n,\hat{P}_U}$ satisfies $\left|Q_{\hat{P}_U}(x) - Q_{n,\hat{P}_U}(x)\right| < \frac{1}{n}$.

Briefly, sub-codebook $\mathcal{C}(\hat{P}_U) = \left\{\boldsymbol{x}(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{T}^n(\hat{P}_U)\right\}$ is called a constant-composition sub-codebook with input distribution $Q_{\hat{P}_U}$ if $\boldsymbol{x}(\boldsymbol{u}) \in \mathcal{T}^n(Q_{n,\hat{P}_U})$ for all $\boldsymbol{u} \in \mathcal{T}^n(\hat{P}_U)$. And codewords of each sub-codebook are randomly drawn uniformly from the set of sequences with type $Q_{n,\hat{P}_U}$. Additionally, codebook $\mathcal{C}$ is considered as the union of all sub-codebooks, i. e. $\mathcal{C} = \bigcup_{\hat{P}_U \in \mathcal{P}^n_{\mathcal{U}}} \mathcal{C}(\hat{P}_U)$.

By applying constant-composition random coding, it can be proved that the following exponent is achievable

$$E^{\rm gcc} = \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} D(\hat{P}_U || P_U) + \min_{\hat{P}_{XY} \in \mathcal{S}^{\rm gcc}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \left[I\left(X;Y\right) - H(\hat{P}_U)\right]^+, \quad (2.50)$$

where $\hat{P}_Y$ is the marginal distribution of $\hat{P}_{XY}$ and

$$\mathcal{S}^{\rm gcc}(\hat{P}_U) \triangleq \left\{ \hat{P}_{XY} : \hat{P}_{XY} = Q_{\hat{P}_U} \hat{P}_{Y|X}, \hat{P}_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \right\}. \quad (2.51)$$

Note that (2.50) can be derived from (2.19), by considering the fact that for the constant-composition ensemble, we ought to consider only $\hat{P}_{XY}$ that their marginal distribution $\hat{P}_X$ is $Q_{\hat{P}_U}$. By applying the identity $\hat{P}_{XY} = Q_{\hat{P}_U} \hat{P}_{Y|X}$ in $D(\hat{P}_{XY} || Q_{\hat{P}_U} W)$ in (2.19) and noticing that

$$D(Q_{\hat{P}_U} \hat{P}_{Y|X} || Q_{\hat{P}_U} \hat{P}_Y) = I\left(X;Y\right), \quad (2.52)$$

(2.50) is proved.

Like before, the aim of this section is determining an alternative expression for the achievable exponent given by (2.50). By using the fact that $[x]^+ = \max_{0 \le \rho \le 1} \rho x$ and Fan's minimax theorem [22], we obtain

$$E^{\rm gcc} = \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} \max_{\rho \in [0,1]} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) + \min_{\hat{P}_{XY} \in \mathcal{S}^{\rm cc}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho I\left(X;Y\right). \quad (2.53)$$

To determine a dual expression for the achievable exponent given in (2.53), we apply Lagrange duality theory to the two minimizations over $\hat{P}_{XY}$ and $\hat{P}_U$. Firstly, we fix $\hat{P}_U \in \mathcal{P}_\mathcal{U}$ and we consider the inner minimization over $\hat{P}_{XY} \in \mathcal{S}^{\rm gcc}(\hat{P}_U)$, i.e. we focus on the following optimization problem

$$\min_{\hat{P}_{XY} \in \mathcal{S}^{\rm gcc}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho I\left(X;Y\right). \quad (2.54)$$

To apply Lagrange duality theory to the optimization problem given in (2.54), we use the following Lemma.

Lemma 2.2. *For the generalized constant-composition ensemble, we have*

$$\min_{\hat{P}_{XY} \in \mathcal{S}^{\rm gcc}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho I\left(X;Y\right) = E_0^{\rm cc}(\rho, Q_{\hat{P}_U}, W), \quad (2.55)$$

where,

$$E_0^{\rm cc}(\rho, Q, W) = \max_{\bar{\alpha}(x) : \sum_x \bar{\alpha}(x) Q(x) = 0} -\log\left(\sum_y \left(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right). \quad (2.56)$$

Proof. See Section 2.4.7. □

Now, by applying Lemma 2.2 into (2.53), we obtain

$$E^{\text{gcc}} = \min_{\hat{P}_U \in \mathcal{P}_{\mathcal{U}}} \max_{\rho \in [0,1]} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) + E_0^{\text{cc}}(\rho, Q_{\hat{P}_U}, W), \tag{2.57}$$

where the optimization problem over $\hat{P}_U$ in (2.57), is exactly the primal form of (2.24) or if we consider two classes, the primal form of (2.33) and (2.34). Thus, for generalized constant-composition ensemble, we obtain

$$E^{\text{gcc}} = \max_{\gamma \in [0,1]} \min_{i \in \{1,2\}} \max_{\rho \in [0,1]} E_0^{\text{cc}}(\rho, Q_i, W) - E_{s,i}(\rho, P_U, \gamma), \tag{2.58}$$

where for $i = 1, 2$, $E_{s,i}(\rho, P_U)$ is given by (2.33) and (2.34).

As a special case, by considering only one input distribution, (2.58) gives the constant-composition exponent, i.e.

$$E^{\text{cc}} = \max_{\rho \in [0,1]} E_0^{\text{cc}}(\rho, Q, W) - E_s(\rho, P_U). \tag{2.59}$$

To compare (2.58) with (2.59) we use Lemma A.6. By setting $E(\rho, Q_i) = E_0^{\text{cc}}(\rho, Q_i, W)$ for $i = 1, 2$, in Lemma A.6, a simpler expression of (2.58) can be written as

$$E^{\text{gcc}} = \max_{\rho \in [0,1]} \bar{E}_0^{\text{cc}}(\rho, \{Q_1, Q_2\}, W) - E_s(\rho, P_U), \tag{2.60}$$

where since $\bar{E}_0^{\text{cc}}(\cdot) \geq E_0^{\text{cc}}(\cdot)$, we conclude that the generalized constant-composition ensemble leads larger exponent than constant-composition ensemble.

In addition, in Lemma A.9, it is shown that E^{cc} given by (2.60) is greater than E^{iid} in (2.49). From Lemma A.9, let $Q^\star$ be an input distribution maximizing E^{iid}. For the cases that E^{cc} derived by a constant-composition $Q^\star$, we have $E^{\text{cc}} = E^{\text{iid}}$.

2.3 Comparing the Exponents

In this section, by comparing the exponents derived in Sections 2.1 and 2.2, we show that there is no penalty in the error exponent if the messages are assumed to be statistically independent of the codewords.

2.3.1 icd and iid Ensembles

Here, we are going to show that, in terms of error exponent, ensembles generated with a conditional distribution have no advantage over the iid ensembles. To do this, for a given conditional distribution $\bar{Q}$, we define a family of marginal distributions as $\Big\{Q_\rho; \rho \in [0,1]\Big\}$, where

$$Q_\rho(x) = \sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u), \tag{2.61}$$

and

$$P_{\frac{1}{1+\rho}}(u) = \frac{P_U(u)^{\frac{1}{1+\rho}}}{\sum_u P_U(u)^{\frac{1}{1+\rho}}}. \tag{2.62}$$

Considering (2.61) and (2.62), we may note that $Q_\rho(x)\sum_u P_U(u)^{\frac{1}{1+\rho}} = \sum_u P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)$. By replacing $\sum_u P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)$ appeared in (2.10) with $Q_\rho(x)\sum_u P_U(u)^{\frac{1}{1+\rho}}$, we obtain

$$\begin{aligned} E_{0,s}(\rho, P_U, \bar{Q}, W) = \\ -\log\Bigg(\sum_y \Big(\sum_x Q_\rho(x)\sum_u P_U(u)^{\frac{1}{1+\rho}} W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\Bigg) \quad (2.63)\\ = E_0(\rho, Q_\rho, W) - E_s(\rho, P_U), \quad (2.64)\end{aligned}$$

where in (2.64), in view of (1.9) and (1.14), we used the identity that $\log(ab) = \log(a) + \log(b)$.

By comparing (2.64) with (2.49), and in view of (2.9), we conclude that ensembles generated with a conditional distribution attain the same exponent as Gallager's exponent. In other words, for a given conditional input distribution $\bar{Q}$ and ρ, we can always find an iid distribution Q_ρ, such that $E_{0,s}(\rho, P_U, \bar{Q}, W) = E_0(\rho, Q_\rho, W) - E_s(\rho, P_U)$. The proof of this equality in the primal domain is presented in Section 2.4.8, i. e. in Section 2.4.8 we show that

$$D(\hat{P}^*_{UXY}||P_U\bar{Q}W) + \Big[D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U)\Big]^+ = \tag{2.65}$$

$$D(\hat{P}^\star_U||P_U) + D(\hat{P}^\star_{XY}||Q_\rho W) + \Big[D(\hat{P}^\star_{XY}||Q_\rho\hat{P}^\star_Y) - H(\hat{P}^\star_U)\Big]^+, \tag{2.66}$$

where P^*_{UXY}, $\hat{P}^\star_{XY}$ and $\hat{P}^\star_U$ are optimal distributions associated to icd and iid exponents and are given in (2.8), (2.47) and (2.48), respectively.

The same approach can be used to show that message-dependent ensemble with statistical dependency does not give a larger exponent than message-dependent ensemble without any statistical dependency between messages and codewords.

At this point, in the dual domain, we obtain reliable transmission conditions similarly to [5] by evaluating the the partial derivative of the objectives in (2.49) and (2.9) at $\rho = 0$. Since both $E_{0,s}$ and $E_0 - E_s$ are concave functions with respect to ρ, i. e. $\frac{\partial^2 E_{0,s}}{\partial \rho^2} \leq 0$ and $\frac{\partial^2 (E_0 - Es)}{\partial \rho^2} \leq 0$, the maximum of exponents occurs at ρ^* where $\frac{\partial E_{0,s}}{\partial \rho^*} = 0$ and $\frac{\partial (E_0 - Es)}{\partial \rho^*} = 0$, respectively. Now, if the derivative of $E_{0,s}$ and $E_0 - E_s$ at $\rho = 0$ be positive, then $E_{0,s}$ and $E_0 - E_s$ will increase from $\rho = 0$ to ρ^* and then they will start to decrease. Thus, by recalling that $E_{0,s}(0, P_U, \bar{Q}, W) = E_0(0, Q, W) - E_s(0, P_U) = 0$, to make sure $\max_{\rho \in [0,1]} E_{0,s}(\rho, P_U, \bar{Q}, W)$ and $\max_{\rho \in [0,1]} E_0(\rho, Q, W) - E_s(\rho, P_U)$ are always positive, it suffices that their derivatives at $\rho = 0$ be positive.

To compare the transmissible conditions of the iid and icd ensembles, firstly we focus on the partial derivative of the $E_0(\rho, Q_\rho, W) - E_s(\rho, P_U)$ at $\rho = 0$, which yields

$$\frac{\partial E_0(\rho, Q_\rho, W)}{\partial \rho}\Big|_{\rho=0} - \frac{\partial E_s(\rho, \cdot)}{\partial \rho}\Big|_{\rho=0} = \frac{\partial E_0(\rho, Q_\rho, W)}{\partial \rho}\Big|_{\rho=0} - H(P_U) \qquad (2.67)$$

$$= I_{Q_0 W}(X;Y) - \sum_{x,y} W(y|x) \frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0} - H(P_U), \qquad (2.68)$$

where in (2.67), we used the fact that the derivative of $E_s(\rho, P_U)$ with respect to ρ at $\rho = 0$ is the source entropy and in (2.68), we differentiated from $E_0(\rho, Q_\rho, W)$ at $\rho = 0$. Note that Q_0 in (2.68) denotes Q_ρ at $\rho = 0$, and

$$I_{Q_0 W}(X;Y) = \sum_{x,y} Q_0(x) W(y|x) \log \frac{W(y|x)}{\sum_{\bar{x}} Q_0(\bar{x}) W(y|\bar{x})}. \qquad (2.69)$$

To determine the quantity of (2.68), it suffices to compute $\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0}$. Evaluating the derivative of Q_ρ given in (2.61) with respect to ρ, yields

$$\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0} =$$

$$\sum_u \bar{Q}(x|u)\Big(- P_U(u) \log(P_U(u)) + P_U(u) \sum_{\bar{u}} P_U(\bar{u}) \log(P_U(\bar{u}))\Big) \qquad (2.70)$$

$$= -\sum_u \bar{Q}(x|u) P_U(u) \log(P_U(u)) - H(P_U) \sum_u \bar{Q}(x|u) P_U(u), \qquad (2.71)$$

where (2.71) follows from the definition of the source entropy. Putting back $\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0}$ derived in (2.71) into $\sum_{x,y} W(y|x)\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0}$ presented in (2.68), we obtain

$$\sum_{x,y} W(y|x)\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0} = -\sum_{u,x,y} W(y|x)\bar{Q}(x|u)P_U(u)\log\left(P_U(u)\right)$$
$$-H(P_U)\sum_{u,x,y} W(y|x)\bar{Q}(x|u)P_U(u) = 0, \quad (2.72)$$

where in (2.72) we used the fact that $\sum_{u,x,y} P_U(u)\bar{Q}(x|u)W(y|x) = 1$ and also the definition of the source entropy.

By replacing zero instead of $\sum_{x,y} W(y|x)\frac{\partial Q_\rho(x)}{\partial \rho}\Big|_{\rho=0}$ appeared in (2.68), we obtain the reliable transmission condition as

$$\frac{\partial E_0(\rho, Q_\rho, W)}{\partial \rho}\Big|_{\rho=0} - \frac{\partial E_s(\rho, P_U)}{\partial \rho}\Big|_{\rho=0} = I_{Q_0 W}(X;Y) - H(P_U) > 0, \quad (2.73)$$

where as mentioned Q_0 denotes Q_ρ at $\rho = 0$.

Next we focus on the transmissible condition of the source P_U by using icd exponent. Again, by evaluating the partial derivative of the $E_{0,s}(\rho, P_U, \bar{Q}, W)$ at $\rho = 0$, we obtain

$$\frac{\partial E_{0,s}(\rho, P_U, \bar{Q}, W)}{\partial \rho}\Big|_{\rho=0} = I_{\bar{Q}W}(X;Y) - H(P_U) > 0, \quad (2.74)$$

where

$$I_{\bar{Q}W}(X;Y) = \sum_{u,x,y} P_U(u)\bar{Q}(x|u)W(y|x)\log\frac{W(y|x)}{\sum_{\bar{u},\bar{x}} P_U(\bar{u})\bar{Q}(\bar{x}|\bar{u})W(y|\bar{x})}. \quad (2.75)$$

Considering (2.73) and (2.74) and noting to the fact that $\max_{Q_0} I_{Q_0 W}(X;Y) = \max_{\bar{Q}} I_{\bar{Q}W}(X;Y)$, we conclude that to have a reliable transmission over channel W, for both iid and icd ensembles, the source entropy should be lower than channel capacity.

2.3.2 Conditional Constant-Composition Ensemble and Constant-Composition Ensemble

In the section, by comparing (2.17) and (2.59) we show that in terms of error exponent, there is no benefit to use conditional constant-composition ensemble instead of generalized constant-composition ensemble.

Proposition 2.9. *Consider a source with probability distribution P_U, for a given conditional distribution $\bar{Q}(x|u)$, there exists a $Q_\rho(x)$ given in* (2.61)*, such that*

$$E_{0,s}^{\mathrm{ccc}}(\rho, P_U\bar{Q}, W) \leq E_0^{\mathrm{cc}}(\rho, Q_\rho, W) - E_s(\rho, P_U), \tag{2.76}$$

where $E_{0,s}^{\mathrm{ccc}}(\cdot)$, $E_0^{\mathrm{cc}}(\cdot)$ and $E_s(\rho, P_U)$ are given in (2.18), (2.56) *and* (1.9)*, respectively.*

Proof. See Section 2.4.9. □

Considering the obtained results, we may conclude that in terms of error exponent, using ensembles generated with conditional distribution has no advantage over those of generated with marginal distribution. However, combining the results presented in Section 2.2, we have the following relation between achievable exponents

$$E^{\mathrm{gcc}} \geq E^{\mathrm{md}} \geq E^{\mathrm{iid}}. \tag{2.77}$$

2.4 Proofs

2.4.1 Proof of Proposition 2.1

We first bound $\bar{\epsilon}^n$, the average error probability over the ensemble, for a given block length n. Applying the random coding union bound [23] for joint source channel coding, we have

$$\bar{\epsilon}^n \leq \sum_{\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}} P_{\boldsymbol{UXY}}^n(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}) \min\left\{1, \sum_{\boldsymbol{u}'\neq\boldsymbol{u}} \mathbb{P}\left[\frac{P_{\boldsymbol{U}}^n(\boldsymbol{u}')W^n(\boldsymbol{y}|\boldsymbol{X}')}{P_{\boldsymbol{U}}^n(\boldsymbol{u})W^n(\boldsymbol{y}|\boldsymbol{x})} \geq 1\right]\right\}, \tag{2.78}$$

where $\boldsymbol{x}'$ has the same distribution as $\boldsymbol{x}$ but is independent of $\boldsymbol{y}$. Recalling that $\pi(\boldsymbol{u})$ denotes the type of source sequence $\boldsymbol{u}$, and codewords are generated according to a conditional distribution which depends on the $\pi(\boldsymbol{u})$, we bound $\bar{\epsilon}^n$ as

$$\bar{\epsilon}^n \leq \sum_{\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}} P_{\boldsymbol{UXY}}^n(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}) \min\left\{1, \sum_{\substack{\boldsymbol{u}'\neq\boldsymbol{u} \\ \boldsymbol{x}':\frac{P_{\boldsymbol{U}}^n(\boldsymbol{u}')W^n(\boldsymbol{y}|\boldsymbol{x}')}{P_{\boldsymbol{U}}^n(\boldsymbol{u})W^n(\boldsymbol{y}|\boldsymbol{x})}\geq 1}} \bar{Q}_{\pi(\boldsymbol{u}')}^n(\boldsymbol{x}'|\boldsymbol{u}')\right\}. \tag{2.79}$$

Next, we group the outer and inner summations in (2.79) based on the empirical distributions of $(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y})$ and $(\boldsymbol{u}',\boldsymbol{x}')$, respectively, and then sum over all possible empirical distributions, respectively denoted by $\hat{P}_{UXY}$ and

$\tilde{P}_{UXY}$. We note that the summation over $\hat{P}_{UXY}$ runs over the set of all possible empirical distributions, $\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}$, while the summation over $\tilde{P}_{UXY}$ is restricted to the set $\mathcal{K}^n$, defined as

$$\mathcal{K}^n(\hat{P}_{UXY}) \triangleq \Big\{\tilde{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} : \tilde{P}_Y = \hat{P}_Y, \mathbb{E}_{\tilde{P}}[\lambda(U,X,Y)] \geq \mathbb{E}_{\hat{P}}[\lambda(U,X,Y)]\Big\}, \quad (2.80)$$

where $\lambda(U,X,Y) = \log\Big(P_U(U)W(Y|X)\Big)$. As a result, we can write the summations in equation (2.79) respectively as

$$\sum_{\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}} P^n_{\boldsymbol{UXY}}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}) = \sum_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} \sum_{(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y})\in\mathcal{T}^n(\hat{P}_{UXY})} P^n_{\boldsymbol{UXY}}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}), \quad (2.81)$$

and

$$\sum_{\substack{\boldsymbol{u}'\neq\boldsymbol{u} \\ \boldsymbol{x}':\frac{P^n_U(\boldsymbol{u}')W^n(\boldsymbol{y}|\boldsymbol{x}')}{P^n_U(\boldsymbol{u})W^n(\boldsymbol{y}|\boldsymbol{x})}\geq 1}} \bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}') = \sum_{\tilde{P}_{UXY}\in\mathcal{K}^n(\hat{P}_{UXY})} \sum_{(\boldsymbol{u}',\boldsymbol{x}')\in\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})} \bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}'), \quad (2.82)$$

where $\mathcal{T}^n_{\boldsymbol{y}}(\cdot)$ is given by (1.40).

Since the conditional distribution $\bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}')$ has the same value for all $(\boldsymbol{u}',\boldsymbol{x}') \in \mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})$, we have

$$\sum_{(\boldsymbol{u}',\boldsymbol{x}')\in\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})} \bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}') = |\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})|\bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}'). \quad (2.83)$$

Considering (1.40) and the fact that $\tilde{P}_Y = \hat{P}_Y$ in $\mathcal{K}^n(\hat{P}_{UXY})$ in (2.80), we have the following upper bound

$$\Big|\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})\Big| = \frac{\Big|\mathcal{T}^n(\tilde{P}_{UXY})\Big|}{\Big|\mathcal{T}^n(\tilde{P}_Y)\Big|} \leq \frac{e^{nH(\tilde{P}_{UXY})+o(n)}}{e^{nH(\tilde{P}_Y)}}, \quad (2.84)$$

where $o(n)$ is a sequence satisfying $\lim_{n\to\infty}\frac{o(n)}{n} = 0$. In addition, using [20, Eq. (1)] for conditional distributions, for all $(\boldsymbol{u}',\boldsymbol{x}') \in \mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})$, we have the following identity on the conditional probability

$$\bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}') = e^{n\sum_{u,x,y}\tilde{P}_{UXY}(u,x,y)\log\bar{Q}_{\tilde{P}_U}(x|u)}, \quad (2.85)$$

where in (2.85), we used the fact that the type of $\boldsymbol{u}$ is $\tilde{P}_U$. Combining inequality (2.84) and identity (2.85) into (2.83), we obtain the following inequality

$$\sum_{(\boldsymbol{u}',\boldsymbol{x}')\in\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})} \bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}') \leq e^{-n\Big(D(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y)-H(\tilde{P}_U)\Big)+o(n)}. \quad (2.86)$$

Further, upper bounding the right hand side of equation (2.86) by the maximum over the empirical probability distributions $\tilde{P}_{UXY} \in \mathcal{K}^n(\hat{P}_{UXY})$, we have

$$\sum_{(\boldsymbol{u}',\boldsymbol{x}')\in\mathcal{T}^n_{\boldsymbol{y}}(\tilde{P}_{UXY})} \bar{Q}^n_{\pi(\boldsymbol{u}')}(\boldsymbol{x}'|\boldsymbol{u}') \leq \max_{\tilde{P}_{UXY}\in\mathcal{K}^n(\hat{P}_{UXY})} e^{-n\left(D(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y)-H(\tilde{P}_U)\right)+o(n)}. \tag{2.87}$$

Moreover, in view of [20, Eq. (12)], the second summation of the right hand side of (2.81) can be expressed as

$$\sum_{(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y})\in\mathcal{T}^n(\hat{P}_{UXY})} P^n_{\boldsymbol{UXY}}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}) \leq e^{-n\left(D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W\right)}. \tag{2.88}$$

Similarly to (2.87), we may upper bound the right hand side of (2.88) by the maximum over the empirical distributions $\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}$, i. e.

$$\sum_{(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y})\in\mathcal{T}^n(\hat{P}_{UXY})} P^n_{\boldsymbol{UXY}}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{y}) \leq \max_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} e^{-n\left(D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W\right)}. \tag{2.89}$$

Putting back the results obtained in equations (2.89) and (2.87) into the respective inner and outer summations (2.81) and (2.82), we obtain that the average error probability (2.79) can be bounded as

$$\bar{\epsilon}^n \leq \sum_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} \max_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} e^{-n\left(D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W\right)} \min\Bigg\{1, \sum_{\tilde{P}_{UXY}\in\mathcal{K}^n(\hat{P}_{UXY})} \max_{\tilde{P}_{UXY}\in\mathcal{K}^n(\hat{P}_{UXY})} e^{-n\left(D(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y)-H(\tilde{P}_U)\right)+o(n)}\Bigg\}, \tag{2.90}$$

where using the fact that the cardinalities of the sets $\mathcal{K}^n(\hat{P}_{UXY})$ and $\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}$ behave polynomially with the codeword length n, and satisfy

$$\left|\mathcal{K}^n(\hat{P}_{UXY})\right| \leq \left|\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}\right| \leq e^{o(n)}, \tag{2.91}$$

we bound $\bar{\epsilon}^n$ as

$$\bar{\epsilon}^n \leq \max_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} e^{-n\left(D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W\right)+o(n)} \min\Bigg\{1, \max_{\tilde{P}_{UXY}\in\mathcal{K}^n(\hat{P}_{UXY})} e^{-n\left(D(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y)-H(\tilde{P}_U)\right)+o(n)}\Bigg\}. \tag{2.92}$$

Using the identity $\min\{1, e^a\} = e^{[a]^+}$, we may write equation (2.92) as

$$\bar{\epsilon}^n \leq e^{-nE^n + o(n)}, \tag{2.93}$$

where

$$E^n = \min_{\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U \bar{Q}_{\hat{P}_U} W) + \Big[\min_{\tilde{P}_{UXY} \in \mathcal{K}^n(\hat{P}_{UXY})} D(\tilde{P}_{UXY}||\tilde{P}_U \bar{Q}_{\tilde{P}_U} \hat{P}_Y) - H(\tilde{P}_U) \Big]^+. \tag{2.94}$$

Using Lemma A.1, we find that $E^n \geq E^n_o$, where

$$E^n_o = \min_{\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U \bar{Q}_{\hat{P}_U} W) + \Big[D(\hat{P}_{UXY}||\hat{P}_U \bar{Q}_{\hat{P}_U} \hat{P}_Y) - H(\hat{P}_U) \Big]^+. \tag{2.95}$$

As a result, the average error probability is bounded as

$$\bar{\epsilon}^n \leq e^{-nE^n_o + o(n)}. \tag{2.96}$$

Taking logarithm and lim inf from both sides of the equation (2.96) and noting that the inequality

$$\liminf_{n\to\infty} \max\{a_n, b_n\} \geq \max\Big\{ \liminf_{n\to\infty} a_n, \liminf_{n\to\infty} b_n \Big\}, \tag{2.97}$$

implies that

$$\liminf_{n\to\infty} [a_n]^+ \geq \Big[\liminf_{n\to\infty} a_n \Big]^+, \tag{2.98}$$

we obtain

$$\liminf_{n\to\infty} -\frac{1}{n} \log(\bar{\epsilon}^n) \geq \liminf_{n\to\infty} \min_{\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U \bar{Q}_{\hat{P}_U} W) + \Big[\liminf_{n\to\infty} D(\hat{P}_{UXY}||\hat{P}_U \bar{Q}_{\hat{P}_U} \hat{P}_Y) - H(\hat{P}_U) \Big]^+. \tag{2.99}$$

We further note that the set of all empirical distributions is dense in the set of all possible probability distributions, and that the functions involved in (2.99) are uniformly continuous over their arguments. Hence, we may replace the optimization over empirical distributions by an optimization over the set of all possible distributions in (2.99) concluding Proposition 2.1.

2.4.2 Proof of Proposition 2.2

Since the proof of Proposition 2.2 is very similar to the proof of Lemma A.3, we omit some details to avoid repetition.

Using the identity that $\max\{0, a\} = \max_{\rho\in[0,1]} \rho a$ and in view of Fan's minimax theorem [22], stating that $\min_a \sup_b f(a, b) = \sup_b \min_a f(a, b)$ provided that the minimum is over a compact set, $f(\cdot, b)$ is convex in a for all b, and $f(a, \cdot)$ is concave in b for all a, equation (2.7) can be written as

$$E^{\text{icd}} = \max_{\rho\in[0,1]} \min_{\hat{P}_{UXY}\in\mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) \\ +\rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}\hat{P}_Y) - \rho H(\hat{P}_U). \tag{2.100}$$

Setting $\hat{P}_{ZY} = \hat{P}_{UXY}$ and $P_Z = \hat{P}_U\bar{Q}$ in Lemma A.4, we will find that $D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}\hat{P}_Y) = \min_{V_Y} D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y)$. By applying this fact to (2.100), we obtain

$$E^{\text{icd}} = \max_{\rho\in[0,1]} \min_{V_Y} \min_{\hat{P}_{UXY}\in\mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) \\ +\rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U), \tag{2.101}$$

where V_Y is an arbitrary probability assignment over the channel output alphabet $\mathcal{Y}$.

Now, to solve the optimization problem in (2.101), we apply Lagrange duality theory. Firstly we consider the minimization over $\hat{P}_{UXY}$ in (2.101). Since the objective function in (2.101) is convex with respect to $\hat{P}_{UXY}$ and the constraint $\sum_{u,x,y} \hat{P}_{UXY} = 1$ is affine, the strong duality conditions are satisfied and we have

$$\min_{\hat{P}_{UXY}\in\mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U) \tag{2.102}$$

$$= \max_{\theta} \min_{\hat{P}_{UXY}\in\mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} \mathcal{L}^{\text{icd}}(\hat{P}_{UXY}, \theta), \tag{2.103}$$

where $\min_{\hat{P}_{UXY}} \mathcal{L}^{\text{icd}}(\hat{P}_{UXY}, \theta)$ is the Lagrange dual function of (2.102) and is given by

$$\mathcal{L}^{\text{icd}}(\hat{P}_{UXY}, \theta) = D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U) \\ + \theta\Big(1 - \sum_{u,x,y} \hat{P}_{UXY}(u, x, y)\Big), \tag{2.104}$$

where θ is the Lagrange multiplier associated with the well-known constraint $\sum_{u,x,y} \hat{P}_{UXY}(u, x, y) = 1$. We proceed by analyzing the KKT conditions for

$\hat{P}_{UXY}$. Firstly, we simplify (2.104) by using the definitions of the relative entropy and entropy and considering the fact that $\sum_j a_j \log b_j + c \sum_j a_j \log b_j = \sum_j a_j \log b_j^{1+c}$, which leads to

$$\mathcal{L}^{\text{icd}}(\hat{P}_{UXY}, \theta) = \sum_{u,x,y} \hat{P}_{UXY}(u,x,y) \log \frac{\hat{P}_{UXY}(u,x,y)^{1+\rho}}{P_U(u)\bar{Q}(x|u)^{1+\rho}W(y|x)V_Y(y)^{\rho}} + \theta\Big(1 - \sum_{u,x,y} \hat{P}_{UXY}(u,x,y)\Big). \tag{2.105}$$

Next, we apply the KKT conditions. Since strong duality holds, optimal $(\hat{P}_{UXY}, \theta)$ must satisfy KKT conditions, i. e. for optimal $(\hat{P}_{UXY}, \theta)$, we have $\frac{\partial \mathcal{L}^{\text{icd}}(\hat{P}_{UXY},\theta)}{\partial \hat{P}_{UXY}} = 0$ which yields

$$\log \frac{\hat{P}_{UXY}(u,x,y)^{1+\rho}}{P_U(u)\bar{Q}(x|u)^{1+\rho}W(y|x)V_Y(y)^{\rho}} + (1+\rho) - \theta = 0. \tag{2.106}$$

Solving (2.106) with respect to $\hat{P}_{UXY}(u,x,y)$, the optimal value of $\hat{P}_{UXY}(u,x,y)$ is derived as

$$\hat{P}_{UXY}(u,x,y) = e^{\frac{\theta-(1+\rho)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}. \tag{2.107}$$

To apply the constraint $\sum_{u,x,y} \hat{P}_{UXY}(u,x,y) = 1$, we sum both sides of (2.107) over (u,x,y) which gives us the optimal value of θ. Putting back the optimal θ in (2.107), the optimal $\hat{P}_{UXY}(u,x,y)$ is derived as

$$\hat{P}_{UXY}(u,x,y) = \frac{P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{u},\bar{x},\bar{y}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}}. \tag{2.108}$$

By inserting the optimal value of $\hat{P}_{UXY}$ obtained in (2.108) into (2.105), we obtain

$$\max_{\theta} \min_{\hat{P}_{UXY} \in \mathcal{P}_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} \mathcal{L}^{\text{icd}}(\hat{P}_{UXY}, \theta) = -(1+\rho) \log \Big(\sum_{u,x,y} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}} \Big). \tag{2.109}$$

Next, in view of (2.101), (2.102) and (2.109), we have

$$E^{\text{icd}} = \max_{\rho \in [0,1]} -(1+\rho) \log \Big(\max_{V_Y} \sum_{u,x,y} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}} \Big), \tag{2.110}$$

where in (2.110), by using the fact that logarithm is an increasing function, we took the minimization inside the logarithm.

Now, in order to find the optimal value of V_Y, we use Lemma A.2. By defining $f(y) = \sum_{u,x} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}$, in view of Lemma A.2, the optimal value of V_Y which maximizes the objective function inside the logarithm in (2.110) is given by

$$V_Y(y) = \frac{\left(\sum_{u,x} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}}{\sum_{\bar{y}} \left(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}}. \tag{2.111}$$

By inserting the obtained $V_Y(y)$ derived in (2.111) into (2.110) and (2.108), we obtain

$$\hat{P}^{\star}_{UXY}(u,x,y) = \frac{P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(y|\bar{x})^{\frac{1}{1+\rho}}\right)^{\rho}}{\sum_{\bar{y}} \left(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}}, \tag{2.112}$$

and

$$E^{\text{icd}} = \max_{\rho\in[0,1]} -\log\left(\sum_{y}\left(\sum_{u,x} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \tag{2.113}$$

which concludes the proof.

2.4.3 Proof of Proposition 2.4

Again, we apply the Lagrange duality theory. For $\hat{P}_{UXY} \in \mathcal{S}^{\text{ccc}}$, we have $I(U,X;Y) = D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}\hat{P}_Y)$. Considering this fact and setting $Z = UX$ and $P_Z = \hat{P}_U\bar{Q}$ in Lemma A.4, we may conclude that $I(U,X;Y) = D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}\hat{P}_Y) = \min_{V_Y} D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y)$. As a result, in view of Lemma A.4, (2.14) can be expressed as

$$E^{\text{ccc}} = \min_{\hat{P}_{UXY}\in\mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \left[\min_{V_Y} D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - H(\hat{P}_U)\right]^{+}, \tag{2.114}$$

where, V_Y is an arbitrary probability assignment over the channel output alphabet $\mathcal{Y}$. In view of the identity that $\max\{0,a\} = \max_{\rho\in[0,1]} \rho a$ and Fan's minimax theorem [22], E^{ccc} given in (2.114) can be written as

$$E^{\text{ccc}} = \max_{\rho\in[0,1]} \min_{V_Y} \min_{\hat{P}_{UXY}\in\mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) \\ +\rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U). \quad (2.115)$$

To derive an alternative expression for E^{ccc} given in (2.115), firstly we focus on the inner minimization over $\hat{P}_{UXY} \in \mathcal{S}^{\text{ccc}}$ in (2.115). In fact, we consider the following optimization problem as the primal problem, where

$$\min_{\hat{P}_{UXY}\in\mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U). \quad (2.116)$$

Like before, to determine the optimal value of $\hat{P}_{UXY}$ which minimizes (2.116), we apply Lagrange duality theory. We consider the presented constraints in (2.15) which leads $\sum_{u,x,y} \hat{P}_{UXY}(u,x,y) = 1$ and the new constraint $\hat{P}_{UXY}(u,x,y) = \hat{P}_U(u)\bar{Q}(x|u)\hat{P}_{Y|XU}(y|x,u)$, i. e. for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$, we have $\hat{P}_{X|U}(x|u) = \bar{Q}(x|u)$ or

$$\bar{Q}(x|u) = \hat{P}_{X|U}(x|u) = \frac{\hat{P}_{UX}(u,x)}{\hat{P}_U(u)} = \frac{\sum_y \hat{P}_{UXY}(u,x,y)}{\sum_{\bar{x},\bar{y}} \hat{P}_{UXY}(u,\bar{x},\bar{y})}, \quad (2.117)$$

where in (2.117), we used the definition of marginal distribution. By multiplying both sides of (2.117) by $\sum_{x,y} \hat{P}_{UXY}(u,x,y)$, an equivalent constrain of the (2.15) is given by

$$\sum_y \hat{P}_{UXY}(u,x,y) = \bar{Q}(x|u) \sum_{\bar{x},\bar{y}} \hat{P}_{UXY}(u,\bar{x},\bar{y}). \quad (2.118)$$

Now, in view of the constraint presented in (2.118) and the obvious fact that $\sum_{u,x,y} \hat{P}_{UXY}(u,x,y) = 1$, we define $\mathcal{L}^{\text{ccc}}$ as the Lagrangian associated with the optimization problem given in (2.116). Since the objective function is a convex function over a convex set and the two constraints (2.118) and $\sum_{u,x,y} \hat{P}_{UXY}(u,x,y) = 1$ are affine, the strong duality conditions hold and therefore

$$\min_{\hat{P}_{UXY}\in\mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U) \quad (2.119)$$

$$= \max_{\beta(\cdot),\theta} \min_{\hat{P}_{UXY}} \mathcal{L}^{\text{ccc}}(\hat{P}_{UXY},\theta,\beta), \quad (2.120)$$

where

$$\begin{aligned}\mathcal{L}^{\text{ccc}}(\hat{P}_{UXY},\theta,\beta) &= D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U) \\ &\quad + \sum_{u,x}\beta(u,x)\Big(\bar{Q}(x|u)\sum_{y,\bar{x}}\hat{P}_{UXY}(u,\bar{x},y) - \sum_{y}\hat{P}_{UXY}(u,x,y)\Big) \\ &\quad + \theta\left(1 - \sum_{uxy}\hat{P}_{UXY}(u,x,y)\right), \qquad (2.121)\end{aligned}$$

and θ, β associate respectively with the constraints $\sum_{u,x,y} P_{UXY}(u,x,y) = 1$ and (2.118). Using the definition of the relative entropy, entropy and the fact that $\sum_j a_j \log b_j + c\sum_j a_j \log b_j = \sum_j a_j \log b_j^{1+c}$, the Lagrangian is simplified as

$$\begin{aligned}\mathcal{L}^{\text{ccc}}(\hat{P}_{UXY},\theta,\beta) &= \sum_{uxy}\hat{P}_{UXY}(u,x,y)\log\frac{\hat{P}_{UXY}(u,x,y)^{1+\rho}}{P_U(u)\bar{Q}(x|u)^{1+\rho}W(y|x)V_Y(y)^{\rho}} \\ &+ \sum_{u,x}\beta(u,x)\Big(\bar{Q}(x|u)\sum_{y,\bar{x}}\hat{P}_{UXY}(u,\bar{x},y) - \sum_{y}\hat{P}_{UXY}(u,x,y)\Big) \\ &\quad + \theta\left(1 - \sum_{uxy}\hat{P}_{UXY}(u,x,y)\right). \qquad (2.122)\end{aligned}$$

Since the strong duality holds, the Lagrange multipliers satisfies the KKT conditions. By setting $\frac{\partial \mathcal{L}^{\text{ccc}}(\hat{P}_{UXY},\theta,\beta)}{\partial \hat{P}_{UXY}} = 0$, we obtain

$$\begin{aligned}&\log\frac{\hat{P}_{UXY}(u,x,y)^{1+\rho}}{P_U(u)\bar{Q}(x|u)^{1+\rho}W(y|x)V_Y(y)^{\rho}} + (1+\rho) - \theta \\ &\qquad + \sum_{\bar{x}}\beta(u,\bar{x})\bar{Q}(\bar{x}|u) - \beta(u,x) = 0. \qquad (2.123)\end{aligned}$$

Solving (2.123) with respect to $\hat{P}_{UXY}(u,x,y)$ and applying the constraint that $\sum_{u,x,y}\hat{P}_{UXY}(u,x,y) = 1$, the optimal $\hat{P}^{\star}_{UXY}(u,x,y)$ is derived as

$$\begin{aligned}&\hat{P}^{\star}_{UXY}(u,x,y) = \\ &\frac{e^{\frac{\beta(u,x)-\sum_{\bar{x}}\beta(u,\bar{x})\bar{Q}(\bar{x}|u)}{1+\rho}}P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)W(y|x)^{\frac{1}{1+\rho}}V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{u,x,y}e^{\frac{\beta(u,x)-\sum_{\bar{x}}\beta(u,\bar{x})\bar{Q}(\bar{x}|u)}{1+\rho}}P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)W(y|x)^{\frac{1}{1+\rho}}V_Y(y)^{\frac{\rho}{1+\rho}}}. \qquad (2.124)\end{aligned}$$

Next, we apply the second constraint given in (2.118). In view of (2.118), we sum $\hat{P}_{UXY}(u,x,y)$ given in (2.124) over y to obtain $\hat{P}_{UX}(u,x)$. We also sum $\hat{P}_{UXY}(u,x,y)$ over (x,y) to obtain $\hat{P}_U(u)$. By inserting the fact that $\bar{Q}(x|u) = \frac{\hat{P}_{UX}(u,x)}{\hat{P}_U(u)}$, we find that the optimal β satisfies

$$\beta(u,x) - \sum_{\bar{x}} \beta(u,\bar{x})\bar{Q}(\bar{x}|u) =$$
$$-(1+\rho)\log \frac{\sum_y W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{x},\bar{y}} e^{\frac{\beta(u,\bar{x}) - \sum_{\bar{\bar{x}}} \beta(u,\bar{\bar{x}})\bar{Q}(\bar{\bar{x}}|u)}{1+\rho}} \bar{Q}(\bar{x}|u) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}}. \quad (2.125)$$

Inserting the optimum values of $(\hat{P}_{UXY}, \beta)$ derived in (2.124) and (2.125), respectively into (2.122), we obtain

$$\mathcal{L}^{\text{ccc}}(\beta) =$$
$$-(1+\rho)\log\left(\sum_{u,x,y} e^{\frac{\beta(u,x) - \sum_{\bar{x}} \beta(u,\bar{x})\bar{Q}(\bar{x}|u)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} Q(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \quad (2.126)$$

where $\mathcal{L}^{\text{ccc}}(\beta) = \max_\theta \min_{\hat{P}_{UXY}} \mathcal{L}^{\text{ccc}}(\hat{P}_{UXY}, \theta, \beta)$. To simplify $\mathcal{L}^{\text{ccc}}(\beta)$, we define $\bar{\beta}(u,x) \triangleq \beta(u,x) - \sum_{\bar{x}} \beta(u,\bar{x})\bar{Q}(\bar{x}|u)$. Multiplying both sides of $\bar{\beta}(u,x)$ by $\bar{Q}(x|u)$ and summing over x, implies that $\sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0$. Replacing $\beta(u,x) - \sum_{\bar{x}} \beta(u,\bar{x})\bar{Q}(\bar{x}|u)$ appeared in (2.126) with the $\bar{\beta}(u,x)$, $\mathcal{L}^{\text{ccc}}(\beta)$ in (2.126) can be expressed as

$$\mathcal{L}^{\text{ccc}}(\bar{\beta}) = -(1+\rho)\log\left(\sum_{u,x,y} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \quad (2.127)$$

subject to the constraint $\sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0$. Considering (2.126) and (2.127), we conclude that $\max_{\beta(u,x)} \mathcal{L}^{\text{ccc}}(\beta) = \max_{\bar{\beta}(u,x): \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0} \mathcal{L}^{\text{ccc}}(\bar{\beta})$. Considering this fact, in view of (2.127) and (2.120), we note that the left hand of (2.120) is equal with the $\max_{\bar{\beta}(u,x): \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0} \mathcal{L}^{\text{ccc}}(\bar{\beta})$, i. e.

$$\min_{\hat{P}_{UXY} \in \mathcal{S}^{\text{ccc}}} D(\hat{P}_{UXY}||P_U\bar{Q}W) + \rho D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}V_Y) - \rho H(\hat{P}_U) =$$
$$\max_{\bar{\beta}(u,x): \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0} -(1+\rho)\log\left(\sum_{u,x,y} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) \times W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right). \quad (2.128)$$

Now, in view of (2.115) and (2.128), we conclude that

$$E^{\text{ccc}} = \max_{\rho\in[0,1]} \min_{V_Y} \max_{\substack{\bar{\beta}(u,x):\\ \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0}}$$
$$-(1+\rho)\log\left(\sum_{u,x,y} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right). \quad (2.129)$$

Again, by using Fan's minimax theorem, we can swap $\min_{V_Y}$ and $\max_{\bar{\beta}(\cdot)}$ as

$$E^{\text{ccc}} = \max_{\rho\in[0,1]} \max_{\substack{\bar{\beta}(u,x):\\ \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0}}$$
$$-(1+\rho)\log\left(\max_{V_Y} \sum_{u,x,y} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \quad (2.130)$$

where in (2.130) by using the fact that the logarithm is an increasing function, we took the minimization inside the logarithm. Considering the maximization over V_Y in (2.130), we define $e(y) = \sum_{u,x} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}$, which leads to

$$\max_{V_Y} \sum_y e(y) V_Y(y)^{\frac{\rho}{1+\rho}}, \quad (2.131)$$

where the objective function in (2.131) is concave function of V_Y. Using Lemma A.2, the optimal value of $V_Y(y)$ of the objective function in (2.130) is derived as

$$V_Y(y) = \frac{\left(\sum_{u,x} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}}{\sum_{\bar{y}} \left(\sum_{\bar{u},\bar{x}} e^{\frac{\bar{\beta}(\bar{u},\bar{x})}{1+\rho}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\right)^{1+\rho}}. \quad (2.132)$$

Putting back $V_Y(y)$ obtained in (2.132) into (2.130), we obtain

$$E^{\text{ccc}} = \max_{\rho\in[0,1]}$$
$$\max_{\substack{\bar{\beta}(u,x):\\ \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0}} -\log\left(\sum_y \left(\sum_{u,x} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \quad (2.133)$$

where by defining

$$E_{0,s}^{\text{ccc}}(\rho, P_U, Q, W) =$$
$$\max_{\substack{\bar{\beta}(u,x):\\ \sum_x \bar{\beta}(u,x)\bar{Q}(x|u)=0}} -\log\left(\sum_y \left(\sum_{u,x} e^{\frac{\bar{\beta}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \quad (2.134)$$

Proposition 2.4 is proved.

2.4.4 Proof of Lemma 2.1

We recall that $\mathcal{P}^n_{\mathcal{U}}$ denotes the set of all empirical distributions on a vector in $\mathcal{U}^n$ and given $\tilde{P}_U \in \mathcal{P}^n_{\mathcal{U}}$, the type class $\mathcal{T}^n(\tilde{P}_U)$ is the set of all sequences in $\mathcal{U}^n$ with type $\tilde{P}_U$. Thus, $\mathcal{U}^n$ can be partitioned by the source type messages, i. e. $\mathcal{U}^n = \bigcup_{\tilde{P}_U \in \mathcal{P}^n_{\mathcal{U}}} \mathcal{T}^n(\tilde{P}_U)$.

Moreover, in view of [21], we define $\mathcal{D}^n_\ell \triangleq \{\boldsymbol{u} : \gamma^n_{\ell+1} < P^n_{\boldsymbol{U}}(\boldsymbol{u}) \leq \gamma^n_\ell\}$ for $\ell \in \{1, ..., L\}$ where $\gamma_{L+1} = 0$ and $\gamma_1 = 1$. To make sure $\mathcal{D}^n_\ell \neq \emptyset$, we assume $\gamma_L > \min P_U(u)$ and $\gamma_2 \leq \max P_U(u)$. Given $\ell \in \{1, ..., L\}$, from the definition of $\mathcal{D}^n_\ell$, it can be verified that $\mathcal{D}^n_\ell$s are disjoint subsets (referred to as classes) such that $\mathcal{U}^n = \bigcup_{\ell=1}^L \mathcal{D}^n_\ell$. Thus, $\mathcal{U}^n$ can be also partitioned by $\mathcal{D}^n_\ell$s. We refer ℓ as the indexed class.

Noting to the fact that all $\boldsymbol{u} \in \mathcal{T}^n(\tilde{P}_U)$ have the same probability, $\mathcal{D}^n_\ell$s are unions of type classes. We shall prove that $\mathcal{T}^n(\tilde{P}_U) \subseteq \mathcal{D}^n_\ell$ provided that $\log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell$, which means that the intersection of $\mathcal{T}^n(\tilde{P}_U)$ and $\mathcal{D}^n_\ell$ can be expressed as

$$\mathcal{D}^n_\ell \bigcap \mathcal{T}^n(\tilde{P}_U) = \begin{cases} \mathcal{T}(\tilde{P}_U) & \log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell, \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.135}$$

To show (2.135), let $\boldsymbol{u} \in \mathcal{T}^n(\tilde{P}_U)$, it means that the string $\boldsymbol{u}$ contains exactly $n\tilde{P}_U(u)$ occurrences of letter u or the probability of sequence $\boldsymbol{u}$ can be written as $P^n_{\boldsymbol{U}}(\boldsymbol{u}) = \prod_{u \in \mathcal{U}} P_U(u)^{n\tilde{P}_U(u)}$. In addition, since $\mathcal{U}^n$ is partitioned by $\mathcal{D}^n_\ell$s, there exists a unique ℓ such that $\boldsymbol{u} \in \mathcal{D}^n_\ell$ or $\gamma^n_{\ell+1} < P^n_{\boldsymbol{U}}(\boldsymbol{u}) \leq \gamma^n_\ell$. By expressing the probability of $\boldsymbol{u}$ in terms of its type, we obtain $\gamma^n_{\ell+1} < \prod_{u \in \mathcal{U}} P_U(u)^{n\tilde{P}_U(u)} \leq \gamma^n_\ell$. Using the properties that $b^c = e^{c \log(b)}$ and $\log(a_1 a_2) = \log(a_1) + \log(a_2)$, we conclude $e^{n \log \gamma_{\ell+1}} < e^{n \sum_u \tilde{P}_U(u) \log P_U(u)} \leq e^{n \log \gamma_\ell}$ which is equivalent to the expression $\log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell$ in (2.135). In fact, we showed that if $\boldsymbol{u} \in \mathcal{T}^n(\tilde{P}_U) \Rightarrow \boldsymbol{u} \in \mathcal{D}^n_\ell$, i. e. $\mathcal{T}^n(\tilde{P}_U) \subseteq \mathcal{D}^n_\ell$, provided that $\log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell$.

As a result, in view of (2.135), $\mathcal{D}^n_\ell$ can be rewritten as

$$\mathcal{D}^n_\ell = \{\boldsymbol{u} : \gamma^n_{\ell+1} < P^n_{\boldsymbol{U}}(\boldsymbol{u}) \leq \gamma^n_\ell\} \tag{2.136}$$

$$= \{\mathcal{T}^n(\tilde{P}_U) : \log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell\}. \tag{2.137}$$

In other words, let $\tilde{P}_U \in \mathcal{P}^n_{\mathcal{U}}$, there exists a unique $\ell \in \{1, ..., L\}$ such that $\log \gamma_{\ell+1} < \sum_u \tilde{P}_U(u) \log P_U(u) \leq \log \gamma_\ell$. In addition, since the set of all empirical distributions is dense in the set of all possible probability distributions $\mathcal{P}_{\mathcal{U}}$, we conclude Lemma 2.1.

2.4.5 Proof of Proposition 2.6

In order to prove Proposition 2.6, we simplify the achievable exponent in (2.19) as

$$E^{\text{md}} = \min_{\hat{P}_U \in \mathcal{P}_{\mathcal{U}}} \max_{\rho \in [0,1]} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) \\ + \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} \hat{P}_Y), \quad (2.138)$$

where in (2.138) we used the identity that $\max\{0, a\} = \max_{\rho\in[0,1]} \rho a$ and Fan's minimax theorem stating that $\min_a \sup_b f(a,b) = \sup_b \min_a f(a,b)$ provided that the minimum is over a compact set, $f(\cdot, b)$ is convex in a for all b, and $f(a, \cdot)$ is concave in b for all a.

Next, to determine the dual expression for the achievable exponent, we apply Lagrange duality theory to the two minimizations over the $\hat{P}_U$ and $\hat{P}_{XY}$ in (2.138). Firstly, we fix $\hat{P}_U \in \mathcal{P}_{\mathcal{U}}$ and we consider the inner minimization over $\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}$, i. e. we focus on the following optimization problem

$$\min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} \hat{P}_Y). \quad (2.139)$$

Lemma A.3 shows that (2.139) can be expressed as

$$\min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} \hat{P}_Y) = E_0(\rho, Q_{\hat{P}_U}, W), \quad (2.140)$$

where $E_0(\cdot)$ is given by (1.14). Now, combining (2.138) and (2.140), we face the following optimization problem

$$E^{\text{md}} = \min_{\hat{P}_U \in \mathcal{P}_{\mathcal{U}}} \max_{\rho \in [0,1]} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) + E_0(\rho, Q_{\hat{P}_U}, W). \quad (2.141)$$

To determine the optimal value of $\hat{P}_U$, firstly we consider the input distribution $Q_{\hat{P}_U}$ and its dependency on $\hat{P}_U$. As mentioned, in view of Lemma 2.1, we can split the minimization over $\hat{P}_U$ into minimization over disjoint classes given by (2.21). For the $\hat{P}_U$ belonging to the class ℓ, we let $Q_{\hat{P}_U} = Q_\ell$. Now, in view of Lemma 2.1, by splitting the minimization over $\hat{P}_U$ in (2.141) into disjoint classes, we obtain

$$E^{\text{md}} = \min_{\ell \in \{1,\dots,L\}} \max_{\rho \in [0,1]} \min_{\substack{\hat{P}_U \in \mathcal{P}_{\mathcal{U}}: \\ \log \gamma_{\ell+1} < \sum_u \hat{P}_U(u) \log P_U(u) \le \log \gamma_\ell}} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) \\ + E_0(\rho, Q_\ell, W), \quad (2.142)$$

where in (2.142), we changed $Q_{\hat{P}_U}$ into Q_ℓ, we used the fact that $\min_{x\in\mathcal{S}} f(x) = \min_j \min_{x\in\mathcal{S}_j} f(x)$, where the set $\mathcal{S}$ is partitioned by the $\mathcal{S}_j$s, and also in view of Sion's minimax theorem [24], we swap the maximization over ρ with the minimization over $\hat{P}_U$.

At this point, we consider the inner minimization over $\hat{P}_U$ in (2.142) to apply Lagrange duality theory. Since $D(\hat{P}_U||P_U) - \rho H(\hat{P}_U)$ is the only quantity in (2.142) that depends on the $\hat{P}_U$, we consider the following optimization problem as the primal problem

$$\min_{\substack{\hat{P}_U\in\mathcal{P}_\mathcal{U}:\\ \log\gamma_{\ell+1}<\sum_u \hat{P}_U(u)\log P_U(u)\le\log\gamma_\ell}} D(\hat{P}_U||P_U) - \rho H(\hat{P}_U), \tag{2.143}$$

where the objective function $D(\hat{P}_U||P_U) - \rho H(\hat{P}_U)$ and the two inequality constrains

$$\log\gamma_{\ell+1} - \sum_u \hat{P}_U(u)\log P_U(u) < 0 \tag{2.144}$$

$$\sum_u \hat{P}_U(u)\log P_U(u) - \log\gamma_\ell < 0, \tag{2.145}$$

are convex with respect to $\hat{P}_U$ and the equality constraints of $\sum_u \hat{P}_U(u) = 1$ is affine. Thus, the primal problem in (2.143) satisfies the strong duality conditions which leads to

$$\min_{\substack{\hat{P}_U\in\mathcal{P}_\mathcal{U}:\\ \log\gamma_{\ell+1}<\sum_u \hat{P}_U(u)\log P_U(u)\le\log\gamma_\ell}} D(\hat{P}_U||P_U) - \rho H(\hat{P}_U) = \max_{\substack{\lambda_{\ell+1}\ge0,\lambda_\ell\ge0\\ \theta}} \min_{\hat{P}_U} \mathcal{L}(\hat{P}_U,\theta,\lambda_{\ell+1},\lambda_\ell), \tag{2.146}$$

where $\min_{\hat{P}_U} \mathcal{L}(\hat{P}_U,\theta,\lambda_{\ell+1},\lambda_\ell)$ is the Lagrange dual function to the primary problem (2.143) and $\mathcal{L}(\hat{P}_U,\theta,\lambda_{\ell+1},\lambda_\ell)$ is the Lagrangian associated with the optimization problem in (2.143) and is given by

$$\mathcal{L}(\hat{P}_U,\theta,\lambda_{\ell+1},\lambda_\ell) = D(\hat{P}_U||P_U) - \rho H(\hat{P}_U) + \theta\Big(1 - \sum_u \hat{P}_U(u)\Big) + \lambda_{\ell+1}\Big(\log\gamma_{\ell+1} - \sum_u \hat{P}_U(u)\log P_U(u)\Big) + \lambda_\ell\Big(\sum_u \hat{P}_U(u)\log P_U(u) - \log\gamma_\ell\Big). \tag{2.147}$$

Using the definition of relative entropy and entropy, the Lagrangian is simplified as

$$\mathcal{L}(\hat{P}_U,\theta,\lambda_{\ell+1},\lambda_\ell) = \sum_u \hat{P}_U(u)\log\frac{\hat{P}_U(u)^{1+\rho}}{P_U(u)^{1+\lambda_{\ell+1}-\lambda_\ell}} + \theta\Big(1 - \sum_u \hat{P}_U(u)\Big) + \lambda_{\ell+1}\log\gamma_{\ell+1} - \lambda_\ell\log\gamma_\ell. \tag{2.148}$$

Since the strong duality holds, we proceed to simplify (2.148) using the KKT conditions. For the optimum values of $(\hat{P}_U, \theta, \lambda_{\ell+1}, \lambda_\ell)$, the KKT conditions implies that $\frac{\partial \mathcal{L}(\hat{P}_{XY}, \theta, \lambda_{\ell+1}, \lambda_\ell)}{\partial \hat{P}_U} = 0$, which yields

$$\log \frac{\hat{P}_U(u)^{1+\rho}}{P_U(u)^{1+\lambda_{\ell+1}-\lambda_\ell}} + (1+\rho) - \theta = 0. \tag{2.149}$$

By solving (2.149) with respect to $\hat{P}_U$, we find that

$$\hat{P}_U(u) = e^{\frac{\theta-(1+\rho)}{1+\rho}} P_U(u)^{\frac{1+\lambda_{\ell+1}-\lambda_\ell}{1+\rho}}. \tag{2.150}$$

Summing both sides of (2.150) over u and considering the fact that $\sum_u \hat{P}_U(u) = 1$, we obtain

$$1 = e^{\frac{\theta-(1+\rho)}{1+\rho}} \sum_u P_U(u)^{\frac{1+\lambda_{\ell+1}-\lambda_\ell}{1+\rho}}. \tag{2.151}$$

Inserting the value of $e^{\frac{\theta-(1+\rho)}{1+\rho}}$ obtained in (2.151) into (2.150), the optimal value of $\hat{P}_U(u)$ is derived as

$$\hat{P}_U(u) = \frac{P_U(u)^{\frac{1+\lambda_{\ell+1}-\lambda_\ell}{1+\rho}}}{\sum_{\bar{u}} P_U(\bar{u})^{\frac{1+\lambda_{\ell+1}-\lambda_\ell}{1+\rho}}}. \tag{2.152}$$

Putting back the optimum values of $(\hat{P}_U, \theta)$ derived in (2.152) and (2.151) into (2.148), the Lagrangian can be written as

$$\mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = -(1+\rho)\log\left(\sum_u P_U(u)^{\frac{1+\lambda_{\ell+1}-\lambda_\ell}{1+\rho}}\right) + \lambda_{\ell+1}\log\gamma_{\ell+1} - \lambda_\ell \log\gamma_\ell, \tag{2.153}$$

where $\mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = \max_\theta \min_{\hat{P}_U} \mathcal{L}(\hat{P}_U, \theta, \lambda_{\ell+1}, \lambda_\ell)$.

Now, in view of (2.146) and (2.153), we may determine the maximum of $\mathcal{L}(\lambda_{\ell+1}, \lambda_\ell)$ with respect to $\lambda_{\ell+1} \geq 0$ and $\lambda_\ell \geq 0$. Let $\lambda^*_{\ell+1}$ and λ^*_ℓ be the quantities which maximize the Lagrangian in (2.153). The KKT conditions imply that for the cases where the constraints presented in (2.144) and (2.145) are inactive, $\lambda^*_{\ell+1} = \lambda^*_\ell = 0$. While, for the case where the inequality constraint in (2.144) is active, since $\frac{\partial^2 \mathcal{L}}{\partial \lambda^2_{\ell+1}} \leq 0$, the maximum of the Lagrangian occurs at the $\lambda^*_{\ell+1} \geq 0$, where $\frac{\partial \mathcal{L}(\lambda^*_{\ell+1}, \lambda_\ell)}{\partial \lambda^*_{\ell+1}} = 0$ (same for the λ^*_ℓ). At this point, based on the activation or inactivation of (2.144) and (2.145) we consider four cases.

Case I: Assume (2.144), (2.145) are active and inactive, respectively. Considering the facts that (2.145) is inactive and $\mathcal{L}$ is a convex function with respect to $\lambda_{\ell+1}$, we conclude that $\lambda_\ell^* = 0$ and $\frac{\partial \mathcal{L}(\lambda_{\ell+1}^*,0)}{\partial \lambda_{\ell+1}^*} = 0$. By setting $\frac{\partial \mathcal{L}(\lambda_{\ell+1}^*,0)}{\partial \lambda_{\ell+1}^*} = 0$, we obtain

$$\frac{\sum_u P_U(u)^{\frac{1+\lambda_{\ell+1}^*}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1+\lambda_{\ell+1}^*}{1+\rho}}} = \log \gamma_{\ell+1}, \tag{2.154}$$

where (2.154) and its properties are given in Lemma A.5.

In view of Lemma A.5, we recall that (2.154) is the same function as (A.33), if $\frac{1+\lambda_{\ell+1}^*}{1+\rho}$ is defined as

$$\frac{1+\lambda_{\ell+1}^*}{1+\rho} = \frac{1}{1+\rho_{\gamma_{\ell+1}}^*} \Rightarrow \lambda_{\ell+1}^* = \frac{\rho - \rho_{\gamma_{\ell+1}}^*}{1+\rho_{\gamma_{\ell+1}}^*}, \tag{2.155}$$

where noting to (2.154), $\rho_{\gamma_{\ell+1}}^*$ satisfies (A.33).

Now, we come back to the problem of determining the maximum of the $\mathcal{L}(\lambda_{\ell+1}, \lambda_\ell)$ given in (2.153), when (2.144) and (2.145) are active and inactive, respectively. Since (2.145) is inactive, $\lambda_\ell^* = 0$ and due to the fact that (2.144) is active, $\lambda_{\ell+1}^* > 0$ in (2.154) maximizes the Lagrangian in (2.153). Moreover, in view of (2.155), by applying $\lambda_{\ell+1}^* > 0$ we conclude that $\frac{1}{1+\rho_{\gamma_{\ell+1}}^*} > \frac{1}{1+\rho}$. By inserting $\lambda_\ell^* = 0$ in (2.153), substituting $\log \gamma_{\ell+1}$ obtained in (2.154) into (2.153), and noting this fact that $\frac{\partial E_s(\rho, P_U)}{\partial \rho} = \log(\sum_u P_U(u)^{\frac{1}{1+\rho}}) - \frac{1}{1+\rho} \frac{\sum_u P_U(u)^{\frac{1}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1}{1+\rho}}}$, the maximum of the Lagrangian for the first case can be expressed as

$$\max_{\lambda_{\ell+1} \geq 0, \lambda_\ell \geq 0} \mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) =$$
$$-E_s(\rho_{\gamma_{\ell+1}}^*, P_U) - (\rho - \rho_{\gamma_{\ell+1}}^*) E_s'(\rho_{\gamma_{\ell+1}}^*), \qquad \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}^*}. \tag{2.156}$$

Case II: In this case, we assume (2.144) is inactive which leads that $\lambda_{\ell+1}^* = 0$ and (2.145) is active which leads that $\frac{\partial \mathcal{L}(0,\lambda_\ell^*)}{\partial \lambda_\ell^*} = 0$. Similarly, by setting $\frac{\partial \mathcal{L}(0,\lambda_\ell^*)}{\partial \lambda_\ell^*} = 0$, we obtain

$$\frac{\sum_u P_U(u)^{\frac{1-\lambda_\ell^*}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1-\lambda_{i^*}}{1+\rho}}} = \log \gamma_\ell, \tag{2.157}$$

where again we put $\frac{1-\lambda_\ell^*}{1+\rho}$ as

$$\frac{1-\lambda_\ell^*}{1+\rho} = \frac{1}{1+\rho_{\gamma_\ell}^*} \Rightarrow \lambda_\ell^* = \frac{\rho_{\gamma_\ell}^* - \rho}{1+\rho_{\gamma_\ell}^*}, \tag{2.158}$$

Due to the activation of (2.145), $\lambda_\ell^* > 0$ where in view of (2.158) we conclude that $\frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_\ell}^*}$. By inserting $\lambda_{\ell+1}^* = 0$ and $\log \gamma_\ell$ obtained in (2.157) into (2.153), and noting to the relation of $\frac{\partial E_s(\rho, P_U)}{\partial \rho}$, the maximum of the Lagrangian for the second case is derived as

$$\max_{\lambda_{\ell+1}\geq 0, \lambda_\ell \geq 0} \mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = -E_s(\rho_{\gamma_\ell}^*, P_U) - (\rho - \rho_{\gamma_\ell}^*)E_s'(\rho_{\gamma_\ell}^*), \qquad \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_\ell}^*}. \tag{2.159}$$

Case III: In this case both (2.144) and (2.145) are inactive, $\lambda_{\ell+1}^* = \lambda_i^* = 0$. In other words, since $\frac{\partial^2 \mathcal{L}}{\partial \lambda_{\ell+1}^2} \leq 0$ and $\frac{\partial^2 \mathcal{L}}{\partial \lambda_\ell^2} \leq 0$, the maximum of the Lagrangian is derived at the $\lambda_{\ell+1} \leq 0$ and $\lambda_\ell \leq 0$. By applying the non-positive condition to (2.155) and (2.158), we conclude $\frac{1}{1+\rho_{\gamma_{\ell+1}}^*} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_\ell}^*}$. Finally, by inserting $\lambda_{\ell+1}^* = \lambda_\ell^* = 0$ in (2.153), the maximum of the Lagrangian is derived as

$$\max_{\lambda_{\ell+1}\geq 0, \lambda_\ell \geq 0} \mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = -E_s(\rho, P_U), \quad \frac{1}{1+\rho_{\gamma_{\ell+1}}^*} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_\ell}^*}. \tag{2.160}$$

Case IV: Finally, for the case where both (2.144) and (2.145) are active, the maximum of the Lagrangian is derived at the $\lambda_{\ell+1} \geq 0$ and $\lambda_\ell \geq 0$. Considering this fact, in view of (2.155) and (2.158), we have $\frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}^*}$ and $\frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_\ell}^*}$ which means $\frac{1}{1+\rho_{\gamma_{\ell+1}}^*} > \frac{1}{1+\rho_{\gamma_\ell}^*}$ and it is contradiction with the inequality presented in (A.34). As a result, this case does not occur.

Combining (2.156), (2.159) and (2.160), the maximum value of the Lagrangian is derived as

$$\max_{\lambda_{\ell+1}\geq 0, \lambda_\ell \geq 0} \mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = \begin{cases} -E_s(\rho_{\gamma_{\ell+1}}^*, P_U) - (\rho - \rho_{\gamma_{\ell+1}}^*)E_s'(\rho_{\gamma_{\ell+1}}^*), & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}^*}, \\ -E_s(\rho, P_U), & \frac{1}{1+\rho_{\gamma_{\ell+1}}^*} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_\ell}^*}, \\ -E_s(\rho_{\gamma_\ell}^*, P_U) - (\rho - \rho_{\gamma_\ell}^*)E_s'(\rho_{\gamma_\ell}^*), & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_\ell}^*}, \end{cases} \tag{2.161}$$

where $\rho^*_{\gamma_{\ell+1}}$ and $\rho^*_{\gamma_\ell}$ satisfies (A.33). Comparing (2.161) with (2.24), we find that

$$\max_{\lambda_{\ell+1}\geq 0, \lambda_\ell \geq 0} \mathcal{L}(\lambda_{\ell+1}, \lambda_\ell) = -E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell). \tag{2.162}$$

Now, in view of the (2.162), (2.143) and (2.147), the optimal value of the objective function in (2.143) is given by

$$\min_{\substack{\hat{P}_U \in \mathcal{P}_\mathcal{U}: \\ \log \gamma_{\ell+1} < \sum_u \hat{P}_U(u) \log P_U(u) \leq \log \gamma_\ell}} D(\hat{P}_U || P_U) - \rho H(\hat{P}_U) = -E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell). \tag{2.163}$$

Inserting (2.163) into (2.142), by optimizing over thresholds, Proposition 2.6 is proved.

2.4.6 Proof of Proposition 2.7

In order to prove Proposition 2.7, firstly we will show that the left hand side of (2.37) is smaller than the right hand side of it. Then, we will prove that the right hand side of (2.37) is smaller than the left hand side of it.

Consider the scheme where the source-message set is partitioned into three classes, i. e. $0 = \gamma_4 < \gamma_3 \leq \gamma_2 < \gamma_1 = 1$ are four positive ordered numbers such that $\gamma_3 > \min P_U(u)$ and $\gamma_2 \leq \max P_U(u)$. For given input distribution Q_ℓ, where $\ell = 1, 2, 3$, from (2.23), the message-dependent random-coding exponent is derived as

$$\begin{aligned} &E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) = \\ &\qquad \max_{\gamma_2 \in [\gamma_3, 1], \gamma_3 \in [0,1]} \min_{\ell \in \{1,2,3\}} \max_{\rho \in [0,1]} E_0(\rho, Q_i, W) - E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell), \end{aligned} \tag{2.164}$$

where since $\gamma_3 \leq \gamma_2$, the maximization over γ_2 in (2.164), is done over $[\gamma_3, 1]$. Additionally, from (2.24) $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$ is

$$E_{s,1}(\rho, P_U, \gamma_2, 1) = \begin{cases} E_s(\rho_{\gamma_2}, P_U) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_2}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_2}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_2}} \leq \frac{1}{1+\rho}, \end{cases} \tag{2.165}$$

and

$$E_{s,2}(\rho, P_U, \gamma_3, \gamma_2) = \begin{cases} E_s(\rho_{\gamma_3}, P_U) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_3}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_3}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_3}} \leq \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_2}}, \\ E_s(\rho_{\gamma_2}, P_U) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_2}) & \frac{1}{1+\rho} \geq \frac{1}{1+\rho_{\gamma_2}}, \end{cases} \tag{2.166}$$

and

$$E_{s,3}(\rho, P_U, 0, \gamma_3) = \begin{cases} E_s(\rho, P_U) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_3}}, \\ E_s(\rho_{\gamma_3}, P_U) + E_s'(\rho_{\gamma_3})(\rho - \rho_{\gamma_3}) & \frac{1}{1+\rho} \geq \frac{1}{1+\rho_{\gamma_3}}. \end{cases} \tag{2.167}$$

In (2.165), (2.166) and (2.167), the parameter ρ_{γ_ℓ} is the solution of the implicit equation given by (2.25).

For given γ_2 and γ_3, Figure 2.4 shows the functions $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$ and $E_{s,i}(\rho, P_U, \gamma_2)$, $E_{s,i}(\rho, P_U, \gamma_3)$ for $i = 1, 2$. As shown in Figure 2.4, by comparing (2.165) with (2.33), and also by comparing (2.167) with (2.34), we immediately conclude that for given γ_2 and γ_3, we have

$$E_{s,1}(\rho, P_U, \gamma_2, 1) = E_{s,1}(\rho, P_U, \gamma_2), \tag{2.168}$$

$$E_{s,3}(\rho, P_U, 0, \gamma_3) = E_{s,2}(\rho, P_U, \gamma_3). \tag{2.169}$$

Similarly, in view of Figure 2.4, the function $E_{s,2}(\rho, P_U, \gamma_3, \gamma_2)$ given by (2.167), can be expressed as

$$E_{s,2}(\rho, P_U, \gamma_3, \gamma_2) = \min\left\{E_{s,2}(\rho, P_U, \gamma_2), E_{s,1}(\rho, P_U, \gamma_3)\right\}. \tag{2.170}$$

To prove (2.170), we note that for given $\gamma_3 \leq \gamma_2$, from (2.25), we have $\frac{1}{1+\rho_{\gamma_3}} \leq \frac{1}{1+\rho_{\gamma_2}}$. From (2.33) and (2.34), $\min\left\{E_{s,2}(\rho, P_U, \gamma_2), E_{s,1}(\rho, P_U, \gamma_3)\right\}$ can be written as

$$\min\left\{E_{s,2}(\rho, P_U, \gamma_2), E_{s,1}(\rho, P_U, \gamma_3)\right\} = \begin{cases} E_s(\rho_{\gamma_3}, P_U) + E_s'(\rho_{\gamma_3})(\rho - \rho_{\gamma_3}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_3}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_2}} > \frac{1}{1+\rho} \geq \frac{1}{1+\rho_{\gamma_3}}, \\ E_s(\rho_{\gamma_2}, P_U) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_2}) & \frac{1}{1+\rho} \geq \frac{1}{1+\rho_{\gamma_2}}, \end{cases} \tag{2.171}$$

where in (2.171), we used the fact that $\frac{1}{1+\rho_{\gamma_3}} \leq \frac{1}{1+\rho_{\gamma_2}}$. Comparing right hand side of (2.171) with (2.166), we conclude (2.170).

Now, considering $E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W)$ in (2.164), for $\ell = 1, 2, 3$, we replace $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ with the right hand sides of equations (2.168), (2.169) and (2.170), i. e.

$$E(P_U, \left\{Q_1, Q_2, Q_3\right\}, W) = \max_{\substack{\gamma_2 \in [\gamma_3, 1], \\ \gamma_3 \in [0,1]}} \min \left\{ \max_{\rho \in [0,1]} E_0(\rho, Q_1, W) - E_{s,1}(\rho, P_U, \gamma_2), \right.$$

$$\max_{\rho \in [0,1]} E_0(\rho, Q_2, W) - \min\left\{E_{s,2}(\rho, P_U, \gamma_2), E_{s,1}(\rho, P_U, \gamma_3)\right\},$$

$$\left. \max_{\rho \in [0,1]} E_0(\rho, Q_3, W) - E_{s,2}(\rho, P_U, \gamma_3) \right\}. \tag{2.172}$$

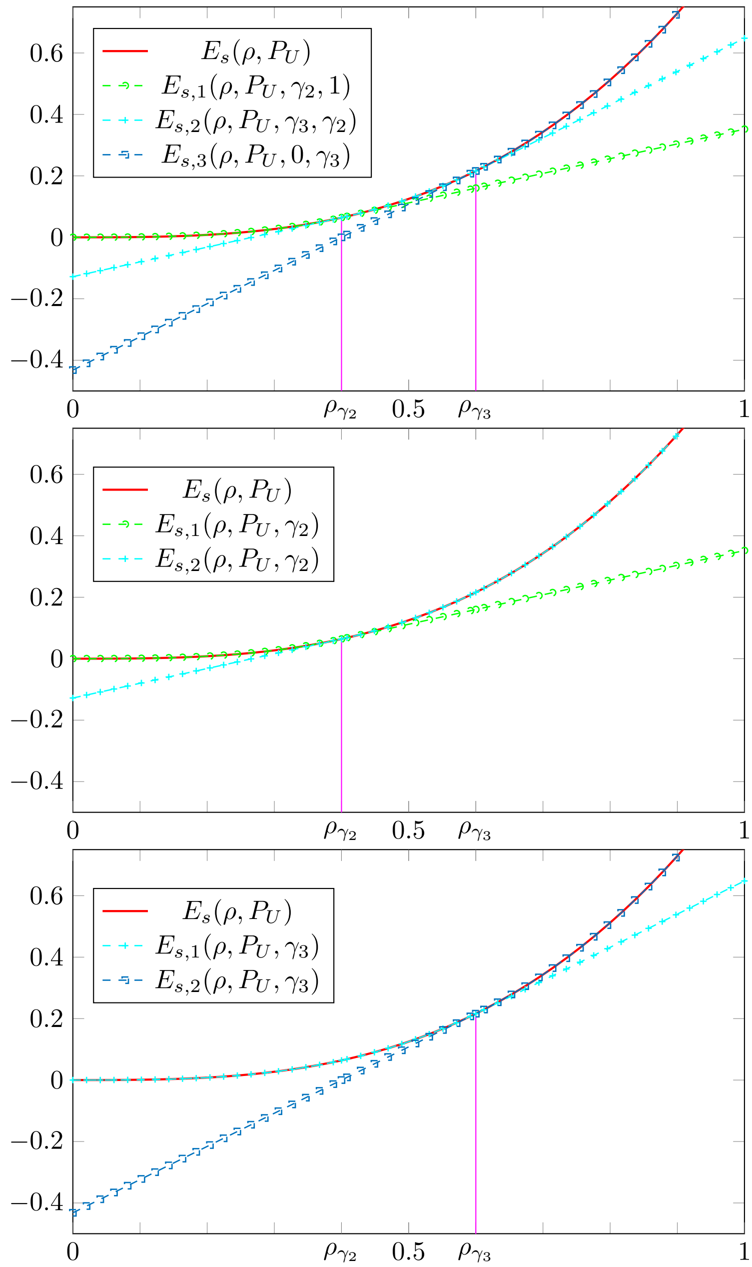

Figure 2.4: $E_{s,\ell}(\rho, P_U, \gamma_{\ell+1}, \gamma_\ell)$ for $\ell = 1, 2, 3$ and $E_{s,i}(\rho, P_U, \gamma_2)$, $E_{s,i}(\rho, P_U, \gamma_3)$ for $i = 1, 2$

Applying $A - \min\{B, C\} = \max\{A - B, A - C\}$, to the second term of

(2.172), we find that

$$E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) = \max_{\gamma_2 \in [\gamma_3, 1], \gamma_3 \in [0,1]} \min \Bigg\{ E_{0s1}(Q_1, \gamma_2), \max\{E_{0s2}(Q_2, \gamma_2), E_{0s1}(Q_2, \gamma_3)\}, E_{0s2}(Q_3, \gamma_3) \Bigg\}, \tag{2.173}$$

where for $i = 1, 2$, $k = 2, 3$ and $j = 1, 2, 3$

$$E_{0si}(Q_j, \gamma_k) = \max_{\rho \in [0,1]} E_0(\rho, Q_j, W) - E_{s,i}(\rho, P_U, \gamma_k). \tag{2.174}$$

Next, we extend the the inner minimization of (2.173). In fact, the minimum of (2.173), chooses the lowest value of three terms which can be modelled as $\min\{a, \max\{c, d\}, b\}$. Using the identity $\min\{a, \max\{c, d\}, b\} = \max\{\min\{a, c, b\}, \min\{a, d, b\}\}$, (2.173) can be rewritten as

$$\begin{aligned} E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) &= \max_{\gamma_2 \in [\gamma_3, 1], \gamma_3 \in [0,1]} \\ \max \Bigg\{ &\min\{E_{0s1}(Q_1, \gamma_2), E_{0s2}(Q_2, \gamma_2), E_{0s2}(Q_3, \gamma_3)\}, \\ &\min\{E_{0s1}(Q_1, \gamma_2), E_{0s1}(Q_2, \gamma_3), E_{0s2}(Q_3, \gamma_3)\} \Bigg\}. \end{aligned} \tag{2.175}$$

Now, in (2.175), by taking the maximizations over γ_3 and γ_2 inside the braces, we find that

$$\begin{aligned} E(P_U, &\{Q_1, Q_2, Q_3\}, W) \le \\ &\max \Bigg\{ \max_{\gamma_2 \in [0,1], \gamma_3 \in [0,1]} \min\{E_{0s1}(Q_1, \gamma_2), E_{0s2}(Q_2, \gamma_2), E_{0s2}(Q_3, \gamma_3)\}, \\ &\max_{\gamma_2 \in [0,1], \gamma_3 \in [0,1]} \min\{E_{0s1}(Q_1, \gamma_2), E_{0s1}(Q_2, \gamma_3), E_{0s2}(Q_3, \gamma_3)\} \Bigg\}, \end{aligned} \tag{2.176}$$

where in (2.176), by taking maximization over $\gamma_2 \in [0, 1]$ rather than the interval of $[\gamma_3, 1]$, we lower bound $E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W)$. As can be seen in (2.176), for every γ_2 and γ_3, $E_{0s1}(Q_1, \gamma_2)$ and $E_{0s2}(Q_2, \gamma_2)$ do not depend on γ_3. Similarly, in the last term of (2.176), $E_{0s1}(Q_2, \gamma_3)$ and $E_{0s2}(Q_3, \gamma_3)$ do not depend on γ_2. Hence, using max-min inequality, we weak (2.176), by swapping the maximization over γ_3 with the minimization inside the second term of (2.176), and also by swapping the maximization over γ_2 with the minimization inside the last term of (2.176), i. e.

$$E(P_U,\{Q_1,Q_2,Q_3\},W)\le \max\left\{\max_{\gamma_2\in[0,1]}\min\left\{E_{0s1}(Q_1,\gamma_2),E_{0s2}(Q_2,\gamma_2),\max_{\gamma_3\in[0,1]}E_{0s2}(Q_3,\gamma_3)\right\},\right.$$
$$\left.\max_{\gamma_3\in[0,1]}\min\left\{\max_{\gamma_2\in[0,1]}E_{0s1}(Q_1,\gamma_2),E_{0s1}(Q_2,\gamma_3),E_{0s2}(Q_3,\gamma_3)\right\}\right\}. \quad (2.177)$$

Now, in view of (2.174), we recall from (2.35) and (2.36), that by moving γ_3 along the interval $[0,1]$, $E_{0s2}(Q_3,\gamma_3)$ decreases from infinity to the function $\max_{\rho\in[0,1]}E_0(\rho,Q_3,W)-E_s(\rho,P_U)$. It means that optimal $\gamma_3=0$ which leads that in (2.177), $\max_{\gamma_3\in[0,1]}E_{0s2}(Q_3,\gamma_3)=+\infty$. Similarly, in the last term of (2.177), $\max_{\gamma_2\in[0,1]}E_{0s1}(Q_1,\gamma_2)=+\infty$. Thus, after taking minimizations between infinity and the remained terms, we obtain

$$E^{\text{md}}(P_U,\{Q_1,Q_2,Q_3\},W)\le\max\left\{\max_{\gamma_2\in[0,1]}\min\{E_{0s1}(Q_1,\gamma_2),E_{0s2}(Q_2,\gamma_2)\},\right.$$
$$\left.\max_{\gamma_3\in[0,1]}\min\{E_{0s1}(Q_2,\gamma_3),E_{0s2}(Q_3,\gamma_3)\}\right\}, \quad (2.178)$$

where we bound the right hand side (2.178) by adding the extra term where

$$E^{\text{md}}(P_U,\{Q_1,Q_2,Q_3\},W)\le\max\left\{\max_{\gamma_2\in[0,1]}\min\{E_{0s1}(Q_1,\gamma_2),E_{0s2}(Q_2,\gamma_2)\},\right.$$
$$\left.\max_{\gamma_3\in[0,1]}\min\{E_{0s1}(Q_2,\gamma_3),E_{0s2}(Q_3,\gamma_3)\},\max_{\gamma\in[0,1]}\min\{E_{0s1}(Q_1,\gamma),E_{0s2}(Q_3,\gamma)\}\right\}. \quad (2.179)$$

Now in view of (2.174), we find that for $i=1,2$, $k=2,3$ and $j,j'=1,2,3$

$$\max_{\gamma_k\in[0,1]}\min\{E_{0si}(Q_j,\gamma_k),E_{0si^c}(Q_{j'},\gamma_k)\}=E^{\text{md}}\left(P_U,\left\{Q_j,Q_{j'}\right\},W\right), \quad (2.180)$$

where i^c denotes the complement index of i over the set $\{1,2\}$. Thus, (2.179) can be rewritten as

$$E^{\text{md}}(P_U,\{Q_1,Q_2,Q_3\},W)\le\max\left\{E^{\text{md}}(P_U,\{Q_1,Q_2\},W),\right.$$
$$\left.E^{\text{md}}(P_U,\{Q_2,Q_3\},W),E^{\text{md}}(P_U,\{Q_1,Q_3\},W)\right\}. \quad (2.181)$$

Up to here we proved that (2.37) is smaller than the right hand side of it. Next, we are going to prove the reverse direction of the inequality in (2.181).

2.4.6.1 Proving the Reverse Direction of (2.181)

Firstly, we consider the definition of $E_{s,2}(\rho, P_U, \gamma_3, \gamma_2)$ in (2.166). As shown in Figure 2.4, we have

$$E_{s,2}(\rho, P_U, \gamma_3, \gamma_2) \le E_{s,1}(\rho, P_U, \gamma_3), \tag{2.182}$$

$$E_{s,2}(\rho, P_U, \gamma_3, \gamma_2) \le E_{s,2}(\rho, P_U, \gamma_2). \tag{2.183}$$

where $E_{s,1}(\rho, P_U, \gamma_3)$ and $E_{s,2}(\rho, P_U, \gamma_2)$ is defined in (2.33) and (2.34). To show (2.182) and (2.183), we recall that since always the tangent line is lower than $E_s(\rho, P_U)$, from definitions of (2.33), (2.34) and (2.166), we immediately conclude (2.182) and (2.183).

Starting from (2.164), the message-dependent exponent derived by considering three class, is bounded as

$$E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) \ge \max_{\gamma_2 \in [\gamma_3, 1], \gamma_3 \in [0,1]} \min \Big\{ E_{0s1}(Q_1, \gamma_2), \\ E_{0s1}(Q_2, \gamma_3), E_{0s2}(Q_3, \gamma_3) \Big\}, \tag{2.184}$$

where in (2.184), in view of (2.174), we respectively applied (2.168), (2.182) and (2.169) into the $E_{s,\ell}(\cdot)$ for $\ell = 1, 2, 3$. Since the first term of (2.184) does not depend on γ_3, the optimal γ_3 is the point that the increasing and decreasing functions of (2.184) with respect to γ_3 are equal to each other (Lemma A.8). Considering (2.35) and (2.36), and noting to the fact that the last two terms of (2.184) does not depend on γ_2, we conclude that by moving γ_2 along the $[\gamma_3, 1]$, the first term of (2.184) increases from $E_{0s1}(Q_1, \gamma_3)$ to infinity. Hence, the optimal γ_2 is the point that the first term of (2.184) be equal with the rest terms. Thus, by removing the first term of (2.184), and in view of (2.32), (2.184) is bounded as

$$E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) \ge \max_{\gamma_3 \in [0,1]} \min \Bigg\{ \max_{\rho \in [0,1]} E_0(\rho, Q_2, W) - E_{s,1}(\rho, P_U, \gamma_3), \\ \max_{\rho \in [0,1]} E_0(\rho, Q_3, W) - E_{s,2}(\rho, P_U, \gamma_3) \Bigg\} = E^{\text{md}}(P_U, \{Q_2, Q_3\}, W). \tag{2.185}$$

In addition, since $\gamma_3 \le \gamma_2$, in view of (2.174), (2.164) can be also written as

$$E^{\text{md}}(P_U, \{Q_1, Q_2, Q_3\}, W) = \max_{\gamma_3 \in [0, \gamma_2], \gamma_2 \in [0,1]} \min \Big\{ E_{0s1}(Q_1, \gamma_2), \\ \max_{\rho \in [0,1]} E_0(\rho, Q_2, W) - E_{s,2}(\rho, P_U, \gamma_3, \gamma_2), E_{0s2}(Q_3, \gamma_3) \Big\}, \tag{2.186}$$

where in (2.186) the maximization over γ_2 is done over $[0,1]$; however, the maximization over γ_3 is done over $[0,\gamma_2]$. In (2.186) we also applied (2.168) and (2.169). Now, by using (2.183) for the second term of (2.186), we find that

$$E^{\mathrm{md}}(P_U,\{Q_1,Q_2,Q_3\},W) \geq \max_{\gamma_3\in[0,\gamma_2],\gamma_2\in[0,1]} \min\Big\{E_{0s1}(Q_1,\gamma_2), E_{0s2}(Q_2,\gamma_2), E_{0s2}(Q_3,\gamma_3)\Big\}. \quad (2.187)$$

Using the same approach as the way (2.185) was derived, we can remove the third term of (2.187) which yields

$$E^{\mathrm{md}}(P_U,\{Q_1,Q_2,Q_3\},W) \geq E^{\mathrm{md}}(P_U,\{Q_1,Q_2\},W). \quad (2.188)$$

Finally by setting $\gamma_2=\gamma_3$, we upper bound (2.164), and we find

$$E^{\mathrm{md}}(P_U,\{Q_1,Q_2,Q_3\},W) \geq E^{\mathrm{md}}(P_U,\{Q_1,Q_3\},W). \quad (2.189)$$

Combining (2.185), (2.188) and (2.189), we conclude the proof.

2.4.7 Proof of Lemma 2.2

In order to prove Lemma 2.2, we recall that for $\hat{P}_{XY}\in\mathcal{S}^{\mathrm{gcc}}(\hat{P}_U)$, we have $\hat{P}_{XY}=Q_{\hat{P}_U}\hat{P}_{Y|X}$ which leads $I\left(X;Y\right)=D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y)$. Next, we use Lemma A.4. By setting $Z=X$ and $P_Z=Q_{\hat{P}_U}$ in Lemma A.4, the quantity $D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y)$ satisfies equation (A.18) and we can conclude that $I\left(X;Y\right)=D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y)=\min_{V_Y}D(\hat{P}_{XY}||Q_{\hat{P}_U}V_Y)$. Applying this fact to (2.54), an equivalent expression of the (2.54) is given by

$$\min_{\hat{P}_{XY}\in\mathcal{S}^{\mathrm{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W)+\rho I\left(X;Y\right)= \quad (2.190)$$

$$\min_{V_Y}\min_{\hat{P}_{XY}\in\mathcal{S}^{\mathrm{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W)+\rho D(\hat{P}_{XY}||Q_{\hat{P}_U}V_Y), \quad (2.191)$$

where, V_Y is an arbitrary probability assignment over the channel output alphabet $\mathcal{Y}$.

To derive an alternative expression for the optimization problem in (2.191), we fix V_Y and we apply Lagrange duality theory to the inner minimization over $\hat{P}_{XY}$ in (2.191), i. e.

$$\min_{\hat{P}_{XY}\in\mathcal{S}^{\mathrm{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W)+\rho D(\hat{P}_{XY}||Q_{\hat{P}_U}V_Y). \quad (2.192)$$

In order to find the optimal value of $\hat{P}_{XY}$ in (2.192), we consider the constraints in (2.51) and the fact that the sum of the probability distribution $\hat{P}_{XY}$ over all possible values of (x, y) is 1. In other words, the following constraints are considered

$$Q_{\hat{P}_U}(x) = \sum_y \hat{P}_{XY}(x, y), \quad \sum_{x,y} \hat{P}_{XY}(x, y) = 1. \tag{2.193}$$

Now, by using Lagrange duality theory, we minimize (2.192) subject to the constraints presented in (2.193). Note that with this particular problem the strong duality conditions are satisfied, because the objective function in (2.192) is a convex function over a convex set and the two constraints given in (2.193) are affine. Thus, we have

$$\min_{\hat{P}_{XY} \in \mathcal{S}^{\text{gcc}}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} V_Y) = \max_{\alpha(\cdot),\theta} \min_{\hat{P}_{XY}} \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha), \tag{2.194}$$

where $\min_{\hat{P}_{XY}} \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha)$ is the Lagrange dual function of (2.192) and the Lagrangian is given by

$$\begin{aligned} \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha) &= D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} V_Y) \\ &+ \theta\Big(1 - \sum_{x,y} \hat{P}_{XY}(x, y)\Big) + \sum_x \alpha(x)\Big(Q_{\hat{P}_U}(x) - \sum_y \hat{P}_{XY}(x, y)\Big), \end{aligned} \tag{2.195}$$

where α and θ correspond to the Lagrange multipliers for the constraints given in (2.193).

Using the definition of the relative entropy and the fact that $\sum_j a_j \log b_j + c \sum_j a_j \log b_j = \sum_j a_j \log b_j^{1+c}$, the Lagrangian is simplified as

$$\begin{aligned} \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha) &= \sum_{x,y} \hat{P}_{XY}(x, y) \log \frac{\hat{P}_{XY}(x, y)^{1+\rho}}{Q_{\hat{P}_U}(x)^{1+\rho} W(y|x) V_Y(y)^\rho} \\ &+ \theta\Big(1 - \sum_{x,y} \hat{P}_{XY}(x, y)\Big) + \sum_x \alpha(x)\Big(Q_{\hat{P}_U}(x) - \sum_y \hat{P}_{XY}(x, y)\Big). \end{aligned} \tag{2.196}$$

To determine the Lagrange dual function, in view of the KKT conditions, the optimal values of $(\hat{P}_{XY}, \theta, \alpha)$ satisfy $\frac{\partial \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha)}{\partial \hat{P}_{XY}(x,y)} = 0$ which implies

$$\log \frac{\hat{P}_{XY}(x, y)^{1+\rho}}{Q_{\hat{P}_U}(x)^{1+\rho} W(y|x) V_Y(y)^\rho} + (1 + \rho) - \theta - \alpha(x) = 0. \tag{2.197}$$

Solving the equation (2.197) with respect to $\hat{P}_{XY}$, yields

$$\hat{P}_{XY}(x, y) = e^{\frac{\theta - (1+\rho)}{1+\rho}} e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}. \tag{2.198}$$

Summing both sides of (2.198) over (x, y) and applying the constraint that $\sum_{x,y} \hat{P}_{XY}(x, y) = 1$, we obtain

$$e^{\frac{\theta-(1+\rho)}{1+\rho}} = \Big(\sum_x e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) \sum_y W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}} \Big)^{-1}. \tag{2.199}$$

Substituting $e^{\frac{\theta-(1+\rho)}{1+\rho}}$ derived in (2.199) into (2.198), $\hat{P}_{XY}(x, y)$ can be expressed as

$$\hat{P}_{XY}(x, y) = \frac{e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{x}} e^{\frac{\alpha(\bar{x})}{1+\rho}} Q_{\hat{P}_U}(\bar{x}) \sum_{\bar{y}} W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}}. \tag{2.200}$$

Now, we derived the optimal value of $\hat{P}_{XY}(x, y)$. Recalling from (2.51), $Q_{\hat{P}_U}(x) = \sum_y \hat{P}_{XY}(x, y)$, by summing both sides of (2.200) over y, we note that

$$Q_{\hat{P}_U}(x) = \sum_y \frac{e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{x}} e^{\frac{\alpha(\bar{x})}{1+\rho}} Q_{\hat{P}_U}(\bar{x}) \sum_{\bar{y}} W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}}, \tag{2.201}$$

where by removing $Q_{\hat{P}_U}(x) \neq 0$ from both sides and taking the logarithm, we obtain

$$\alpha(x) = -(1+\rho) \log \Bigg(\frac{\sum_y W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{x}} e^{\frac{\alpha(\bar{x})}{1+\rho}} Q_{\hat{P}_U}(\bar{x}) \sum_{\bar{y}} W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}} \Bigg). \tag{2.202}$$

Putting back $\hat{P}_{XY}(x, y)$ obtained in (2.200) and $\alpha(x)$ obtained in (2.202) into (2.196), $\max_\theta \min_{\hat{P}_{XY}} \mathcal{L}^{\mathrm{cc}}(\hat{P}_{XY}, \theta, \alpha)$ can be derived as

$$\begin{aligned} &\sum_{\bar{x}} Q_{\hat{P}_U}(\bar{x}) \log \frac{e^{\alpha(\bar{x})}}{\Big(\sum_x e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) \sum_y W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}} \Big)^{1+\rho}} = \\ &-(1+\rho) \log \Big(\sum_x e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) \sum_y W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}} \Big) + \sum_{\bar{x}} Q_{\hat{P}_U}(\bar{x}) \alpha(\bar{x}), \end{aligned} \tag{2.203}$$

where in (2.203) we used the facts that $\log(\frac{a}{b}) = \log a - \log b$ and $\log e^s = s$. Again, in view of $s = \log e^s$, we rewrite $\sum_{\bar{x}} Q_{\hat{P}_U}(\bar{x}) \alpha(\bar{x})$ in (2.203) as

$-(1+\rho)\log e^{-\frac{\sum_{\bar{x}} Q_{\hat{P}_U}(\bar{x})\alpha(\bar{x})}{1+\rho}}$. Finally, by using the identity that $\log c + \log d = \log(c.d)$, (2.203) can be expressed as

$$\mathcal{L}^{\text{cc}}(\alpha) = -(1+\rho)\log\left(e^{-\frac{1}{1+\rho}\sum_{\bar{x}}\alpha(\bar{x})Q_{\hat{P}_U}(\bar{x})}\sum_{x,y} e^{\frac{\alpha(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \tag{2.204}$$

where $\mathcal{L}^{\text{cc}}(\alpha) = \max_\theta \min_{\hat{P}_{XY}} \mathcal{L}^{\text{cc}}(\hat{P}_{XY}, \theta, \alpha)$.

To simplify $\mathcal{L}^{\text{cc}}(\alpha)$, we define $\bar{\alpha}(x) \triangleq \alpha(x) - \sum_{\bar{x}}\alpha(\bar{x})Q_{\hat{P}_U}(\bar{x})$. Multiplying both sides of $\bar{\alpha}(x)$ by $Q_{\hat{P}_U}(x)$ and summing over x implies $\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x) = 0$. Setting $\bar{\alpha}(x) = \alpha(x) - \sum_{\bar{x}}\alpha(\bar{x})Q_{\hat{P}_U}(\bar{x})$, the quantity of $\mathcal{L}^{\text{cc}}(\alpha)$ in (2.204) can be expressed in terms of $\bar{\alpha}(x)$ as

$$\mathcal{L}^{\text{cc}}(\bar{\alpha}) = -(1+\rho)\log\left(\sum_{x,y} e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \tag{2.205}$$

subject to the constraint $\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x) = 0$. Considering (2.204) and (2.205), we may conclude that $\max_{\alpha(x)} \mathcal{L}^{\text{cc}}(\alpha) = \max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x)=0} \mathcal{L}^{\text{cc}}(\bar{\alpha})$. Considering this fact, in view of (2.205), (2.194) and (2.191), we have

$$\min_{\hat{P}_{XY}\in\mathcal{S}^{\text{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W) + \rho D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y) =$$
$$\min_{V_Y} \max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x)=0} -(1+\rho)\log\left(\sum_{x,y} e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right). \tag{2.206}$$

Again, by using Fan's minimax theorem [22], we can swap $\min_{V_Y}$ and $\max_{\bar{\alpha}(\cdot)}$ as

$$\min_{\hat{P}_{XY}\in\mathcal{S}^{\text{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W) + \rho D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y) =$$
$$\max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x)=0} \min_{V_Y} -(1+\rho)\log\left(\sum_{x,y} e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right). \tag{2.207}$$

Now, we focus on the minimization over $V_Y(y)$. Since the logarithm is an increasing function, (2.207) can be expressed as

$$\min_{\hat{P}_{XY}\in\mathcal{S}^{\text{gcc}}(\hat{P}_U)} D(\hat{P}_{XY}||Q_{\hat{P}_U}W) + \rho D(\hat{P}_{XY}||Q_{\hat{P}_U}\hat{P}_Y) =$$
$$\max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q_{\hat{P}_U}(x)=0} -(1+\rho)\log\left(\max_{V_Y}\sum_{x,y} e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \tag{2.208}$$

where by defining $f(y) = \sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}}$ and using Lemma A.2, we obtain

$$\min_{\hat{P}_{XY} \in \mathcal{S}^{\text{gcc}}(\hat{P}_U)} D(\hat{P}_{XY} || Q_{\hat{P}_U} W) + \rho D(\hat{P}_{XY} || Q_{\hat{P}_U} \hat{P}_Y) =$$

$$\max_{\bar{\alpha}(x): \sum_x \bar{\alpha}(x) Q_{\hat{P}_U}(x) = 0} -\log \Big(\sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_{\hat{P}_U}(x) W(y|x)^{\frac{1}{1+\rho}} \Big)^{1+\rho} \Big), \tag{2.209}$$

By renaming the right hand side of (2.209) as $E_0^{\text{cc}}(\rho, Q_{\hat{P}_U}, W)$, Lemma 2.2 is proved.

2.4.8 Proof of (2.66)

In order to prove (2.66), we start by the optimal $\hat{P}^*_{UXY}$ given in (2.8). By inserting the optimal $\hat{P}^*_{UXY}$ given in (2.8) into (2.7), we obtain

$$E^{\text{icd}} = D(\hat{P}^*_{UXY} || P_U \bar{Q} W) + \Big[D(\hat{P}^*_{UXY} || \hat{P}^*_U \bar{Q} \hat{P}^*_Y) - H(\hat{P}^*_U) \Big]^+. \tag{2.210}$$

Firstly, we focus on the second term of (2.210), i. e. $\Big[D(\hat{P}^*_{UXY} || \hat{P}^*_U \bar{Q} \hat{P}^*_Y) - H(\hat{P}^*_U) \Big]^+$. Using the identity that $\hat{P}^*_Y(y) = \sum_{u,x} \hat{P}^*_{UXY}(u,x,y)$, the optimal quantity of $\hat{P}^*_Y(y)$ associated to the primal domain of icd exponent in (2.7) is given by

$$\hat{P}^*_Y(y) = \frac{\Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(y|\bar{x})^{\frac{1}{1+\rho}} \Big)^{1+\rho}}{\sum_{\bar{y}} \Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} \Big)^{1+\rho}}. \tag{2.211}$$

To determine the second term of (2.210), by using the definitions of the relative entropy and entropy and by inserting $\hat{P}^*_{UXY}$ and $\hat{P}^*_Y(y)$ obtained in (2.8) and (2.211), respectively into the second term of (2.210), we obtain

$$D(\hat{P}^*_{UXY} || \hat{P}^*_U \bar{Q} \hat{P}^*_Y) - H(\hat{P}^*_U) =$$

$$\sum_{u,x,y} \hat{P}^*_{UXY}(u,x,y) \log \frac{P_U(u)^{\frac{1}{1+\rho}} W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(y|\bar{x})^{\frac{1}{1+\rho}} \Big)^{\rho}}{\Big(\sum_{\bar{u},\bar{x}} P_U(\bar{u})^{\frac{1}{1+\rho}} \bar{Q}(\bar{x}|\bar{u}) W(y|\bar{x})^{\frac{1}{1+\rho}} \Big)^{1+\rho}}. \tag{2.212}$$

At this point, we use Q_ρ introduced in (2.61). Inserting $\hat{P}^*_{UXY}$ given in (2.8) into (2.212) and then replacing $\sum_u P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u)$ with $Q_\rho(x) \sum_u P_U(u)^{\frac{1}{1+\rho}}$, equation (2.212) can be written as

$$
\begin{aligned}
& D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U) = \\
& \quad \sum_{u,x,y} \frac{P_U(u)^{\frac{1}{1+\rho}} Q_\rho(x) W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^\rho}{\sum_{\bar{u}} P_U(\bar{u})^{\frac{1}{1+\rho}} \sum_{\bar{y}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}} \\
& \quad \times \log \frac{P_U(u)^{\frac{1}{1+\rho}} W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^\rho}{\sum_{\bar{u}} P_U(\bar{u})^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}. \qquad (2.213)
\end{aligned}
$$

Considering the fact that $\log(a.b) = \log(a) + \log(b)$ and using the definition of the entropy for distribution $P_{\frac{1}{1+\rho}}$ given in (2.62), (2.213) is simplified as

$$
\begin{aligned}
& D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U) = \\
& -H(P_{\frac{1}{1+\rho}}) + \sum_{x,y} \frac{Q_\rho(x) W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^\rho}{\sum_{\bar{y}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}} \\
& \qquad \times \log \frac{W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^\rho}{\left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}. \qquad (2.214)
\end{aligned}
$$

In order to express (2.214) in terms of relative entropy and entropy, in view of (2.47), we define $\hat{P}^\star_{XY}(x,y) \triangleq \frac{Q_\rho(x) W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^\rho}{\sum_{\bar{y}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}$. Summing both sides of $\hat{P}^\star_{XY}(x,y)$ over x, we find $\hat{P}^\star_Y(y)$ as

$$
\hat{P}^\star_Y(y) = \frac{\left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}{\sum_{\bar{y}} \left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}. \qquad (2.215)
$$

To simplify (2.214), in view of the defined $\hat{P}^\star_{XY}(x,y)$, by adding and subtracting the quantity $\sum_{x,y} \hat{P}^\star_{XY}(x,y) \log \frac{Q_\rho(x)}{\left(\sum_{\bar{x}} Q_\rho(\bar{x}) W(y|\bar{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}}$ in (2.214) and considering (2.215), we obtain

$$D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U) = -H(P_{\frac{1}{1+\rho}}) + \sum_{x,y}\hat{P}^\star_{XY}(x,y)\log\frac{\hat{P}^\star_{XY}(x,y)}{Q_\rho(x)\hat{P}^\star_Y(y)}. \tag{2.216}$$

Recalling the definition of the relative entropy, from (2.216), we conclude that

$$\Big[D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U)\Big]^+ = \Big[D(\hat{P}^\star_{XY}||Q_\rho\hat{P}^\star_Y) - H(P_{\frac{1}{1+\rho}})\Big]^+. \tag{2.217}$$

Now, we focus on the first part of the (2.210), $D(\hat{P}^*_{UXY}||P_U\bar{Q}W)$. By inserting the optimal $\hat{P}^*_{UXY}$ given in (2.8) into the $D(\hat{P}^*_{UXY}||P_U\bar{Q}W)$, we have

$$\begin{aligned}D(\hat{P}^*_{UXY}||P_U\bar{Q}W) &= \sum_{u,x,y}\hat{P}^*_{UXY}(u,x,y)\\ &\times\log\frac{P_U(u)^{\frac{1}{1+\rho}}W(y|x)^{\frac{1}{1+\rho}}\Big(\sum_{\bar{u},\bar{x}}P_U(\bar{u})^{\frac{1}{1+\rho}}Q(\bar{x}|\bar{u})W(y|\bar{x})^{\frac{1}{1+\rho}}\Big)^\rho}{P_U(u)W(y|x)\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}}P_U(\bar{u})^{\frac{1}{1+\rho}}Q(\bar{x}|\bar{u})W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}}\Big)^{1+\rho}}.\end{aligned} \tag{2.218}$$

Again by putting $\hat{P}^*_{UXY}(u,x,y)$ given in (2.8) into (2.218) and then replacing $\sum_u P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)$ with $Q_\rho(x)\sum_u P_U(u)^{\frac{1}{1+\rho}}$, in view of $P_{\frac{1}{1+\rho}}$ given in (2.62), and the definition of $\hat{P}^\star_{XY}$, after some mathematical manipulations, we obtain

$$\begin{aligned}D(\hat{P}^*_{UXY}||P_U\bar{Q}W) &= D(P_{\frac{1}{1+\rho}}||P_U) + \sum_{x,y}\hat{P}^\star_{XY}(x,y)\log\frac{\hat{P}^\star_{XY}(x,y)}{Q_\rho(x)W(y|x)} && (2.219)\\ &= D(P_{\frac{1}{1+\rho}}||P_U) + D(\hat{P}^*_{UXY}||Q_\rho W), && (2.220)\end{aligned}$$

where in (2.220), we used the definition of the relative entropy.

Combining (2.217) and (2.220), we conclude that

$$\begin{aligned}&D(\hat{P}^*_{UXY}||P_U\bar{Q}W) + \Big[D(\hat{P}^*_{UXY}||\hat{P}^*_U\bar{Q}\hat{P}^*_Y) - H(\hat{P}^*_U)\Big]^+ =\\ &\qquad D(P_{\frac{1}{1+\rho}}||P_U) + \Big[D(\hat{P}^\star_{XY}||Q_\rho\hat{P}^\star_Y) - H(P_{\frac{1}{1+\rho}})\Big]^+,\end{aligned} \tag{2.221}$$

where by considering $P_{\frac{1}{1+\rho}}$ and $P^\star_U$ given in (2.62) and (2.48), respectively (2.66) is proved.

2.4.9 Proof of Proposition 2.9

In order to prove the proposition 2.9, we use the following lemma.

Lemma 2.3. *Let $Q_\rho(x) = \sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)$. It can be proved that*

$$E_0^{\text{cc}}(\rho, \sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u), W) - E_s(\rho, P_U) = \max_{\substack{\tilde{\alpha}(u,x): \\ \sum_{x,u} P_{\frac{1}{1+\rho}}(u)\tilde{\alpha}(u,x)\bar{Q}(x|u)=0}}$$

$$-\log\left(\sum_y \left(\sum_{u,x} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \tag{2.222}$$

where $P_{\frac{1}{1+\rho}}(u)$ is defined in (2.62).

Proof. See Section 2.4.10. □

Next, in order to prove the proposition 2.9, in view of Lemma 2.3, we shall show that the quantity of $E_0^{\text{cc}}(\rho, Q_\rho, W) - E_s(\rho, P_U)$ given in (2.222) is greater than or equal with the $E_{0,s}^{\text{ccc}}(\rho, P_U, Q, W)$. To do this, we compare $E_{0,s}^{\text{ccc}}$ given in (2.18) with the $E_0^{\text{cc}} - E_s$ given in (2.222). Considering (2.18) and (2.222), we may note that in both cases we have a maximization problem with the same objective function and different constraints. Noting to the fact that the maximum value of the objective function with more constraints is lower than that of with less constraints, to prove the proposition 2.9, it suffices to show that $\bar{\beta}(u,x)$ given in (2.18) is more restrictive than $\tilde{\alpha}(u,x)$ given in (2.222), i.e, it suffices to prove that

$$\left\{\bar{\beta}(u,x) : \sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0\right\} \subseteq \left\{\tilde{\alpha}(u,x) : \sum_{u,x} P_{\frac{1}{1+\rho}}(u)\tilde{\alpha}(u,x)\bar{Q}(x|u) = 0\right\}. \tag{2.223}$$

To show (2.223), we note that for all $\bar{\beta}(u,x) \in \left\{\bar{\beta}(u,x) : \sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0\right\}$, we have $\sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0$. Multiplying both sides of the equality by $P_{\frac{1}{1+\rho}}(u)$ and summing over u, we obtain

$$\sum_{u,x} P_{\frac{1}{1+\rho}}(u)\bar{\beta}(u,x)\bar{Q}(x|u) = 0 \Rightarrow$$
$$\bar{\beta}(u,x) \in \left\{\tilde{\alpha}(u,x) : \sum_{u,x} P_{\frac{1}{1+\rho}}(u)\tilde{\alpha}(u,x)\bar{Q}(x|u) = 0\right\}. \tag{2.224}$$

From $\bar{\beta}(u,x) \in \left\{\bar{\beta}(u,x) : \sum_x \bar{\beta}(u,x)\bar{Q}(x|u) = 0\right\}$, we conclude $\bar{\beta}(u,x) \in \left\{\tilde{\alpha}(u,x) : \sum_{u,x} P_{\frac{1}{1+\rho}}(u)\tilde{\alpha}(u,x)\bar{Q}(x|u) = 0\right\}$, i. e. (2.223) is proved which concludes the proof.

2.4.10 Proof of Lemma 2.3

We start with $E_0^{\rm cc}(\rho, Q_\rho, W) - E_s(\rho, P_U)$. In view of (2.56), $E_0^{\rm cc}(\rho, Q_\rho, W) - E_s(\rho, P_U)$ can be expressed as

$$E_0^{\rm cc}(\rho, Q_\rho, W) - E_s(\rho, P_U)$$
$$= \max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q_\rho(x)=0} -\log\Bigg(\sum_y \Big(\sum_u P_U(u)^{\frac{1}{1+\rho}} \sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q_\rho(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\Bigg), \tag{2.225}$$

where in (2.225), we used the definitions of $E_0^{\rm cc}(\rho, Q_\rho, W)$, $E_s(\rho, P_U)$ and the fact that $\log(a) + \log(b) = \log(a.b)$. To express $E_0^{\rm cc}(\rho, Q_\rho, W) - E_s(\rho, P_U)$ in terms of $\bar{Q}(x|u)$, by using the definition of $Q_\rho(x)$ in (2.61), we insert $\sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)$ instead of $Q_\rho(x)$ in (2.225), i.e.

$$E_0^{\rm cc}(\rho, \sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u), W) - E_s(\rho, P_U) = \max_{\bar{\alpha}(x):\sum_{x,u} \bar{\alpha}(x)P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)=0}$$
$$-\log\Bigg(\sum_y \Big(\sum_{u,x} e^{\frac{\bar{\alpha}(x)}{1+\rho}} P_U(u)^{\frac{1}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\Bigg), \tag{2.226}$$

where $P_{\frac{1}{1+\rho}}(u)$ is given in (2.62). Comparing (2.226) with (2.222) and noting the fact that logarithm is an increasing function, we may conclude that in order to prove Lemma 2.3, it suffices to show that

$$\min_{\substack{\bar{\alpha}(x):\\ \sum_{x,u} P_{\frac{1}{1+\rho}}(u)\bar{\alpha}(x)\bar{Q}(x|u)=0}} \sum_y \Big(\sum_{u,x} P_U(u)^{\frac{1}{1+\rho}} e^{\frac{\bar{\alpha}(x)}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho} \tag{2.227}$$

$$= \min_{\substack{\tilde{\alpha}(u,x):\\ \sum_{x,u} P_{\frac{1}{1+\rho}}(u)\tilde{\alpha}(u,x)\bar{Q}(x|u)=0}} \sum_y \Big(\sum_{u,x} P_U(u)^{\frac{1}{1+\rho}} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} \bar{Q}(x|u) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}. \tag{2.228}$$

To show (2.227) equals to (2.228), we define two functions as $k(x,u) = P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)$ and $h(u,x,y) = P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)W(y|x)^{\frac{1}{1+\rho}}$ and we use Lemmas A.10 and A.11. It can be verified that the defined functions $k(x,u)$ and $h(u,x,y)$ satisfy the conditions of the Lemma A.10 and the optimal $\bar{\alpha}^*(x)$ and $\tilde{\alpha}^*(u,x)$ which minimize (2.227) and (2.228) are given in (A.63) and (A.65), respectively. Next, we shall show that for the defined $k(u,x)$ and $h(u,x,y)$ the conditions of Lemma A.11 are also satisfied, i. e. $\frac{h(u,x,y)}{k(u,x)} = \frac{\sum_u h(u,x,y)}{\sum_u k(u,x)}$.

From the definition of $k(x,u)$ and $h(u,x,y)$, we note that the quantity of $\frac{h(u,x,y)}{k(u,x)}$ equals to

$$\frac{P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)W(y|x)^{\frac{1}{1+\rho}}}{P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)} = \frac{\sum_u P_U(u)^{\frac{1}{1+\rho}}\bar{Q}(x|u)W(y|x)^{\frac{1}{1+\rho}}}{\sum_u P_{\frac{1}{1+\rho}}(u)\bar{Q}(x|u)} = \sum_u P_U(u)^{\frac{1}{1+\rho}}W(y|x)^{\frac{1}{1+\rho}}, \tag{2.229}$$

where the last equality in (2.229) follows from the fact that $P_{\frac{1}{1+\rho}}(u) = \frac{P_U(u)^{\frac{1}{1+\rho}}}{\sum_u P_U(u)^{\frac{1}{1+\rho}}}$. From (2.229), we may verify that for the defined functions $h(u,x,y)$ and $k(u,x)$, we have $\frac{h(u,x,y)}{k(u,x)} = \frac{\sum_u h(u,x,y)}{\sum_u k(u,x)}$. As a result, the functions $h(u,x,y)$ and $k(u,x)$ also satisfy the conditions of Lemma A.11. Considering Lemma A.11, $h(u,x,y)$ and $k(u,x)$ we find that $\tilde{\alpha}^*(u,x) = \tilde{\alpha}^*(x) = \bar{\alpha}^*(x)$ and thus (2.227) is equal with (2.228), i. e. Lemma 2.3 is proved.

Chapter 3

The Multiple-Access Channel with Independent Sources

Many works studied the achievable rates and error exponent for a two-user MAC. Here, we just mention some of them. In [25], by considering separate source-channel coding, a universal exponent for the MAC was derived. By universal, we mean that a fixed choice of codewords and decoding set achieves the exponent. In [26], an achievable region is derived for the MAC under mismatched decoding. For the mismatched decoding, the decoding rule is fixed and possibly suboptimal. For more details about mismatched decoding for single user-communication and multiple-access channel see [27]. In [28], it was shown that using structure coding can improve the error exponent of the MAC. Maximum error probability criterion and feedback for the MAC were studied in [29].

By considering separate source-channel coding, in [30] and [31], respectively lower and upper bounds for the error exponent of the MAC were obtained. For the MAC with independent sources, the idea of considering dependency between messages and codewords was studied in [32].

In this chapter, we study the idea of message-dependent ensemble for the MAC. As discussed in Chapter 2, for single-user communication, message-dependent random-coding exponent is larger than iid random-coding exponent. In this Chapter, we show that this result can be generalized to the MAC with independent sources.

After introducing the system setup in Section 3.1, by considering the message-dependent ensemble, and doing the analysis in the dual domain, in Section 3.2, we present an achievable exponent for the MAC with independent sources. In fact, for each user, the source messages are partitioned into two classes and codebooks are generated by drawing codewords from an input distribution depending on the class index of the source message. The partitioning thresholds that maximize the achievable exponent are given by the solution of a system of equations. We also derive both lower and upper bounds for the achievable exponent in terms of Gallager's source and channel functions.

In Sections 3.3, we generalize the obtained results to the case where more than two users are considered, and also the number of classes for each user is arbitrary. Parts of this chapter were presented in [33].

3.1 System Setup

We consider two independent sources characterized by probability distributions P_{U_1}, P_{U_2} on alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$, respectively. We recall again that underlined font represents a pair of quantities for users 1 and 2, such as $\underline{\gamma} = (\gamma_1, \gamma_2)$, $\underline{u} = (u_1, u_2)$ or $W(y|\underline{x}) = W(y|x_1, x_2)$.

Encoder $\nu = 1, 2$, maps a length-n source message $\boldsymbol{u}_\nu$ to the length-n codeword $\boldsymbol{x}_\nu(\boldsymbol{u}_\nu)$ drawn from the codebook $\mathcal{C}^\nu = \{\boldsymbol{x}_\nu(\boldsymbol{u}_\nu) \in \mathcal{X}_\nu^n : \boldsymbol{u}_\nu \in \mathcal{U}_\nu^n\}$. Both terminals send the codewords over a discrete memoryless MAC with transition probability $W(y|x_1, x_2)$, input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and output alphabet $\mathcal{Y}$.

Given the received sequence $\boldsymbol{y}$, the decoder estimates the transmitted pair of messages $\underline{\boldsymbol{u}}$ based on the maximum a posteriori criterion, i. e.

$$\hat{\underline{\boldsymbol{u}}} = \arg\max_{\underline{\boldsymbol{u}} \in \mathcal{U}_1^n \times \mathcal{U}_2^n} P_{\boldsymbol{U}_1}^n(\boldsymbol{u}_1) P_{\boldsymbol{U}_2}^n(\boldsymbol{u}_2) W^n\Big(\boldsymbol{y}|\boldsymbol{x}_1(\boldsymbol{u}_1), \boldsymbol{x}_2(\boldsymbol{u}_2)\Big). \tag{3.1}$$

An error occurs if $\hat{\underline{\boldsymbol{u}}} \neq \underline{\boldsymbol{u}}$. The error probability for a given pair of codebooks $(\mathcal{C}^1, \mathcal{C}^2)$ is given by

$$\epsilon^n(\mathcal{C}^1, \mathcal{C}^2) \triangleq \mathbb{P}\left[(\hat{\boldsymbol{U}}_1, \hat{\boldsymbol{U}}_2) \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)\right], \tag{3.2}$$

and an exponent E is achievable if there exists a sequence of codebooks such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log \epsilon^n(\mathcal{C}_n^1, \mathcal{C}_n^2) \geq E. \tag{3.3}$$

In order to show the existence of such sequences of codebooks, we use random-coding arguments, i. e. we find a sequence of ensembles whose error probability averaged over the ensemble, denoted as $\bar{\epsilon}^n$, tends to zero.

3.2 Message-Dependent Exponent

Message-dependent ensemble is described in Section 2.2.1. In this section, we generalize the idea of message-dependent random coding for the MAC with independent sources.

For user $\nu = 1, 2$, we fix a threshold $0 \le \gamma_\nu \le 1$ to partition the source-message set $\mathcal{U}_\nu^n$ into two classes $\mathcal{A}_\nu^1(\gamma_\nu)$ and $\mathcal{A}_\nu^2(\gamma_\nu)$ defined as

$$\mathcal{A}_\nu^1(\gamma_\nu) = \Big\{ \boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \;:\; P_{\boldsymbol{U}_\nu}^n(\boldsymbol{u}_\nu) \ge \gamma_\nu^n \Big\}, \tag{3.4}$$

$$\mathcal{A}_\nu^2(\gamma_\nu) = \Big\{ \boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \;:\; P_{\boldsymbol{U}_\nu}^n(\boldsymbol{u}_\nu) < \gamma_\nu^n \Big\}. \tag{3.5}$$

For every message $\boldsymbol{u}_\nu \in \mathcal{A}_\nu^1(\gamma_\nu)$, we randomly generate a codeword $\boldsymbol{x}_\nu(\boldsymbol{u}_\nu)$ according to the probability distribution $Q_{\nu,1}(\boldsymbol{x}_\nu) = \prod_{\ell=1}^n Q_{\nu,1}(x_{\nu,\ell})$, and for every message $\boldsymbol{u}_\nu \in \mathcal{A}_\nu^2(\gamma_\nu)$, we randomly generate a codeword $\boldsymbol{x}_\nu(\boldsymbol{u}_\nu)$ according to the probability distribution $Q_{\nu,2}(\boldsymbol{x}_\nu) = \prod_{\ell=1}^n Q_{\nu,2}(x_{\nu,\ell})$. In fact, $Q_{\nu,i}$, for $i = 1, 2$, is a probability distribution that depends on the class of $\boldsymbol{u}_\nu$. Thus, codewords are generated independently according to message-dependent distributions.

As mentioned in (1.43) and (1.44), the symbol $\tau \in \{\{1\}, \{2\}, \{1, 2\}\}$ is used to denote the error event type of the error probability (3.2), i. e. respectively $(\hat{\boldsymbol{u}}_1, \boldsymbol{u}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $(\boldsymbol{u}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$. We denote the complement of τ as τ^c among the subsets of $\{1, 2\}$. We emphasize that throughout this chapter, for error type $\tau = \{1, 2\}$, we have $P_{U_\tau}(u_\tau) = P_{U_1}(u_1)P_{U_2}(u_2)$. Additionally, since only independent sources are considered, the following facts are obvious $P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}}) = P_{\boldsymbol{U}_1}^n(\boldsymbol{u}_1)P_{\boldsymbol{U}_2}^n(\boldsymbol{u}_2)$, and $P_{\underline{U}}(\underline{u}) = P_{U_1}(u_1)P_{U_2}(u_2)$.

Now, by using the introduced random-coding ensemble, we derive an achievable exponent for the MAC with independent sources.

Proposition 3.1. *For the two-user MAC with transition probability W, two independent sources with joint probability distribution $P_{U_1}P_{U_2}$ and class distributions $\{Q_{\nu,1}, Q_{\nu,2}\}$ with user index $\nu = 1, 2$, the following exponent is achievable*

$$E = \max_{\gamma_1,\gamma_2 \in [0,1]} \; \min_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \; \min_{i_1,i_2=1,2} F_{\tau,i_\tau,i_{\tau^c}}\left(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2\right), \tag{3.6}$$

where

$$\begin{aligned} F_{\tau,i_\tau,i_{\tau^c}}\left(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2\right) &= \max_{\rho \in [0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) \\ &\quad - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau) - E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c}). \end{aligned} \tag{3.7}$$

In (3.7), the functions $E_0(\cdot)$ for three different values of τ, is given by (1.28). $E_{s,1}(\cdot)$ and $E_{s,2}(\cdot)$ are respectively given by (2.33) and (2.34). For error type $\tau = \{1, 2\}$, we define $E_{s,i_\tau}(\rho, P_{U_\tau}, \underline{\gamma}) = E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_2)$ and $F_{\{1,2\},i_\tau,i_{\tau^c}}(\cdot) = F_{\{1,2\},i_1,i_2}(\cdot)$.

Proof. See Section 3.4.1. □

We recall that in (3.7), for $\tau = \{1\}$ and $\tau = \{2\}$, $WQ_{\tau^c,i_{\tau^c}}$ denotes a point-to-point channel with input and output alphabets given by $\mathcal{X}_\tau$ and $\mathcal{X}_{\tau^c} \times \mathcal{Y}$, respectively. For $\tau = \{1,2\}$, the input distribution $Q_{\tau,i_\tau} = Q_{1,i_1}Q_{2,i_2}$ is the product distribution $Q_{1,i_1}(x_1)Q_{2,i_2}(x_2)$ over the alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, and $WQ_{\tau^c,i} = W$.

To have a better insight about three types of error in (3.7), we extend it for different values of τ. Equation (3.7) for $\tau = \{1\}$, $\tau = \{2\}$ and $\tau = \{1,2\}$, respectively is given by

$$F_{\{1\},i_1,i_2}(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}, WQ_{2,i_2}) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(0, P_{U_2}, \gamma_2). \tag{3.8}$$

$$F_{\{2\},i_2,i_1}(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2) = \max_{\rho\in[0,1]} E_0(\rho, Q_{2,i_2}, WQ_{1,i_1}) - E_{s,i_1}(0, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2). \tag{3.9}$$

$$F_{\{1,2\},i_1,i_2}(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2). \tag{3.10}$$

In Proposition 3.1, we remark that the optimal assignment of input distributions to source classes is considered in (3.6). Since we considered two source-message classes $\mathcal{A}^1_\nu(\gamma_\nu)$, $\mathcal{A}^2_\nu(\gamma_\nu)$ and two input distributions $Q_{\nu,1}, Q_{\nu,2}$ for each user $\nu = 1, 2$, there are four possible assignments.

Up to here, we found the message-dependent random-coding exponent for the MAC with independent sources in terms of some maximizations and minimizations. We proceed to simplify it by finding the optimal values of γ_1 and γ_2.

3.2.1 Optimal Thresholds

The derived achievable exponent (3.6) contains a maximization over $\underline{\gamma}$, the thresholds that determine how source messages are partitioned into classes. Rearranging the minimizations over τ, i_τ and i_{τ^c}, defining $f_{i_1,i_2}(\underline{\gamma})$ as

$$f_{i_1,i_2}(\gamma_1, \gamma_2) = \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} F_{\tau,i_\tau,i_{\tau^c}}(Q_{1,i_1}, Q_{2,i_2}, \gamma_1, \gamma_2), \tag{3.11}$$

where $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ is given in (3.7), the exponent (3.6) can be written as

$$E = \max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} f_{i_1,i_2}(\gamma_1, \gamma_2). \tag{3.12}$$

We note that regardless the values of i_2, $f_{1,i_2}(\underline{\gamma})$ is non-decreasing with respect to γ_1 and $f_{2,i_2}(\underline{\gamma})$ is non-increasing with respect to γ_1. Similarly, regardless the values of i_1, $f_{i_1,1}(\underline{\gamma})$ is non-decreasing with respect to γ_2 and $f_{i_1,2}(\underline{\gamma})$ is non-increasing with respect to γ_2. As a result, we derive a system of equations to compute the optimal thresholds $\gamma_1^\star$ and $\gamma_2^\star$.

Proposition 3.2. *The optimal $\gamma_1^\star$ and $\gamma_2^\star$ maximizing (3.6) satisfy*

$$\begin{cases} \min_{i_2=1,2} f_{1,i_2}(\gamma_1^\star, \gamma_2^\star) = \min_{i_2=1,2} f_{2,i_2}(\gamma_1^\star, \gamma_2^\star), \\ \min_{i_1=1,2} f_{i_1,1}(\gamma_1^\star, \gamma_2^\star) = \min_{i_1=1,2} f_{i_1,2}(\gamma_1^\star, \gamma_2^\star). \end{cases} \tag{3.13}$$

When (3.13) has no solutions, then $\gamma_\nu^\star \in \{0,1\}$. In particular, if $f_{1,i_2}(0,\gamma_2) > f_{2,i_2}(0,\gamma_2)$ then $\gamma_1^\star = 0$, otherwise $\gamma_1^\star = 1$; and if $f_{i_1,1}(\gamma_1,0) > f_{i_1,2}(\gamma_1,0)$, we have $\gamma_2^\star = 0$, otherwise $\gamma_2^\star = 1$.

Proof. See Section 3.4.2. □

We note that the optimal $\gamma_1^\star$ and $\gamma_2^\star$ are the points where the minimum of all non-decreasing functions with respect to γ_ν is equal with the minimum of all non-increasing functions with respect to γ_ν, for both $\nu = 1,2$.

As shown in Chapter 2, for single user communication, the final expression of message-dependent random coding exponent, is expressed in terms of (2.4). However, for the MAC, finding an unique expression like (2.4), is difficult. In the sequel, we give a graphical intuition of optimal thresholds for the MAC with independent sources.

As shown in (3.13), due to the values of indices $i_1, i_2 = 1,2$, there are four functions in terms of $f_{i_1,i_2}(\underline{\gamma})$. Depending on the fact that for the optimal $\gamma_1^\star$ and $\gamma_2^\star$, which of these four functions would be equal to each other, the achievable exponent given by (3.6) can be decoupled from atleast one optimal threshold. To have a better insight, we can categorize the possible solutions of (3.13). However, since the possible results of (3.13) may be very complicated, to have an intuition, we only focus on the error type $\{1,2\}$. The intuition can be generalized to (3.13).

3.2.1.1 The Intuition of Optimal Thresholds for Error Type $\{1,2\}$

Assume that we have a magic model that only the error type $\tau = \{1,2\}$ occurs. Thus, from (3.6), the achievable exponent is derived as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) \\ - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2). \tag{3.14}$$

Similar to $f_{i_1,i_2}(\underline{\gamma})$ given by (3.11), we define $E_{0,s}^{i_1,i_2}(\underline{\gamma})$ as

$$E_{0,s}^{i_1,i_2}(\gamma_1,\gamma_2) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) \\ - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2), \qquad (3.15)$$

where by using the same reasoning in Proposition 3.2, the optimal $\gamma_1^\star$ and $\gamma_2^\star$ maximizing (3.14), satisfy

$$\min_{i_2=1,2} E_{0,s}^{1,i_2}(\gamma_1^\star,\gamma_2^\star) = \min_{i_2=1,2} E_{0,s}^{2,i_2}(\gamma_1^\star,\gamma_2^\star), \qquad (3.16)$$

$$\min_{i_1=1,2} E_{0,s}^{i_1,1}(\gamma_1^\star,\gamma_2^\star) = \min_{i_1=1,2} E_{0,s}^{i_1,2}(\gamma_1^\star,\gamma_2^\star), \qquad (3.17)$$

and like before, when there is no solution for (3.16) and (3.17), then $\gamma_\nu^\star \in \{0,1\}$ for $\nu = 1,2$. To simplify expressions, we also define $\rho_{i_1,i_2}^\star$ as

$$\rho_{i_1,i_2}^\star = \arg\max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) \\ - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2). \qquad (3.18)$$

As (3.16) and (3.17) suggest, for $\nu = 1,2$, at the points where the minimum of non-decreasing functions with respect to γ_ν is equal with the minimum of non-increasing functions with respect to γ_ν, the optimal $\gamma_1^\star$ and $\gamma_2^\star$ are obtained. Here, we discuss about the possible outcomes for the solution of the system of equations given by (3.16) and (3.17). We start by introducing some functions.

The exponent given by (3.14), is expressed in terms of $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_1)$ for $i_1, i_2 = 1,2$. From (2.33) and (2.34), we recall that $E_{s,i_\nu}(\cdot)$ is $E_s(\cdot)$ for an interval of ρ, while it is $\hat{E}_s(\cdot)$ beyond that interval, where $i_\nu = 1,2$. Similarly, for the MAC with independent sources, the function $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_1)$ for $i_1, i_2 = 1,2$, is one of the following equations

$$E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2}) = \\ (1+\rho)\log\left(\sum_{u_1} P_{U_1}(u_1)^{\frac{1}{1+\rho}}\right) + (1+\rho)\log\left(\sum_{u_2} P_{U_2}(u_2)^{\frac{1}{1+\rho}}\right), \qquad (3.19)$$

$$E_s(\rho, P_{U_1}) + \hat{E}_s(\rho, P_{U_2}, \gamma_2) = \\ E_s(\rho, P_{U_1}) + E_s(\rho_{\gamma_2}, P_{U_2}) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_2}), \qquad (3.20)$$

$$\hat{E}_s(\rho, P_{U_1}, \gamma_1) + E_s(\rho, P_{U_2}) = \\ E_s(\rho_{\gamma_1}, P_{U_1}) + E_s'(\rho_{\gamma_1})(\rho - \rho_{\gamma_1}) + E_s(\rho, P_{U_2}), \qquad (3.21)$$

$$\hat{E}_s(\rho, P_{U_1}, \gamma_1) + \hat{E}_s(\rho, P_{U_2}, \gamma_2) = \\ E_s(\rho_{\gamma_1}, P_{U_1}) + E_s'(\rho_{\gamma_1})(\rho - \rho_{\gamma_1}) + E_s(\rho_{\gamma_2}, P_{U_2}) + E_s'(\rho_{\gamma_2})(\rho - \rho_{\gamma_2}), \qquad (3.22)$$

where $E_s(\cdot)$ and $\hat{E}_s(\cdot)$, respectively given by (1.9) and (2.29). Additionally, from (2.25), ρ_{γ_ν} is the tangent point where $\hat{E}_s(\cdot)$ is tangent to $E_s(\cdot)$.

Figure 3.1 shows equations (3.19), (3.20), (3.21) and (3.22) with respect to ρ. As shown in Figure 3.1, for given γ_1 and γ_2, the straight line (3.22) is tangent to both (3.20) and (3.21) at respectively, ρ_{γ_1} and ρ_{γ_2}. Moreover, both (3.20) and (3.21) are themselves tangent to (3.19) at respectively, ρ_{γ_2} and ρ_{γ_1}.

Figure 3.2 shows $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_1)$ when all four combinations of $i_1, i_2 = 1, 2$ are applied. As can be seen, depending on the region of ρ, the function $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_1)$, is one of the equations given by (3.19), (3.20), (3.21) and (3.22).

In Chapter 2, for point-to-point communication with two given input distributions $Q_1, Q_2 \in \mathcal{P}_\mathcal{X}$, we showed that by moving γ along the $[0,1]$ interval, the straight line tangent to $E_s(\cdot)$ function, i. e. $\hat{E}_s(\cdot)$ is changed. Let $\rho_1^\star$ and $\rho_2^\star$ be the points where respectively $E_0(\rho, Q_1, W) - \hat{E}_s(\rho, P_U, \gamma)$, and $E_0(\rho, Q_2, W) - \hat{E}_s(\rho, P_U, \gamma)$ are maximized with respect to ρ. In Chapter 2, we showed that the optimal threshold is derived at the point where the distances between two $E_0(\cdot)$ functions and $\hat{E}_s(\cdot)$ at $\rho_1^\star$ and $\rho_2^\star$ are equal to each other.

Unlike to single-user communication, for the MAC with two user $\nu = 1, 2$, the optimal $\gamma_\nu^\star$ is obtained at the points where the distances between $E_0(\cdot)$ functions and one of the function given by (3.19), (3.21), (3.20) and (3.22) are equal to each other. To have a better insight, we proceed by studying visually how optimal $\gamma_\nu^\star$ is derived when $\nu = 1, 2$. We start by categorizing the possible outcomes of system of equations given by (3.16) and (3.17).

1. No solutions for both (3.16) and (3.17):
Firstly, we assume that the system of equations given by (3.16) and (3.17), does not have any solution. Thus, the optimal γ_ν for $\nu = 1, 2$ is either zero or one. Considering (2.35) and (2.36), it means that we cannot achieve more than iid random-coding exponent, and for this case, (3.14) is simplified as

$$\max_{i_1=1,2} \max_{i_2=1,2} \max_{\rho\in[0,1]} E_0\left(\rho, Q_{1,i_1} Q_{2,i_2}, W\right) - E_s(\rho, P_{U_1}) - E_s(\rho, P_{U_2}). \tag{3.23}$$

2. No solution for (3.17), while (3.16) has solution:
In this case, we assume that (3.16) has solution which gives $\gamma_1^\star$, while there is no solution for (3.17), and $\gamma_2^\star$ is either zero or one. Hence, from (2.35) and (2.36), we note that if $\gamma_2^\star = 0$, then the solution of (3.16) is derived when $i_2 = 1$; however, if $\gamma_2^\star = 1$, the solution of (3.16) is derived for $i_2 = 2$. Figure 3.3 shows an example where $\gamma_2^\star = 0$, and (3.16) is solved such that

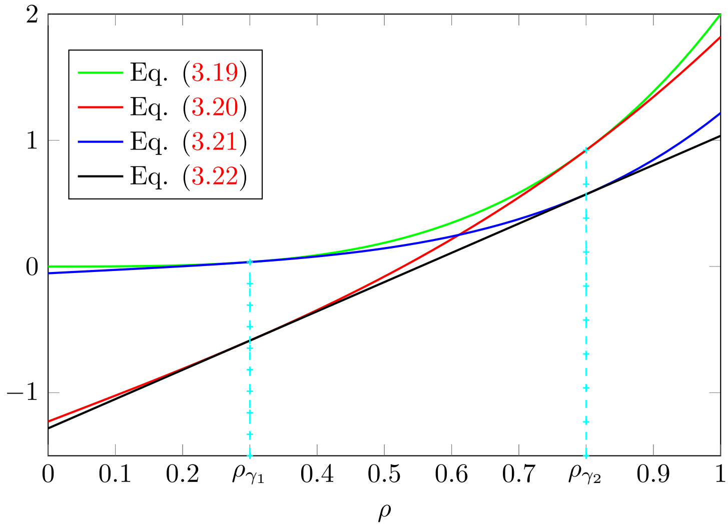

Figure 3.1: Equations (3.19), (3.20), (3.21) and (3.22) with respect to ρ.

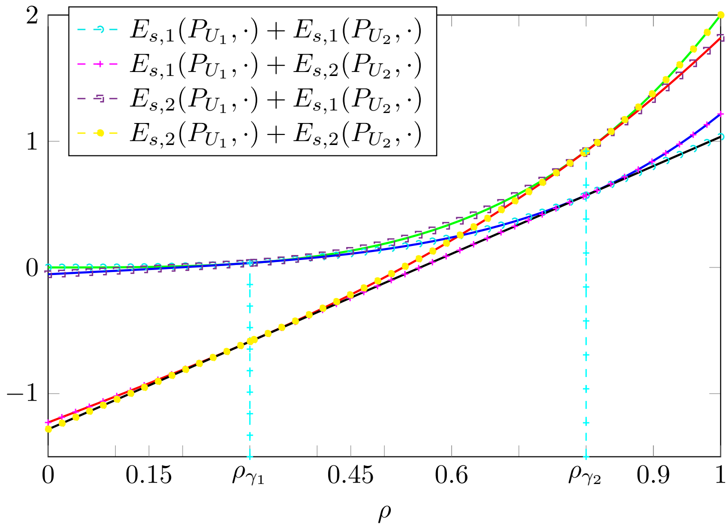

Figure 3.2: $E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_2)$ for $i_1, i_2 = 1, 2$.

$E_{0,s}^{1,1}(\gamma_1^\star, 0) = E_{0,s}^{2,1}(\gamma_1^\star, 0)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1^\star) - E_s(\rho, P_{U_2}) = \\ \max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star) - E_s(\rho, P_{U_2}), \quad (3.24)$$

where as shown in Figure 3.3, the distance between $E_0(\rho, Q_{1,1}Q_{2,1}, W)$ and (3.21) at $\rho_{1,1}^\star$ is equal with the distance between $E_0(\rho, Q_{1,2}Q_{2,1}, W)$ and (3.21)

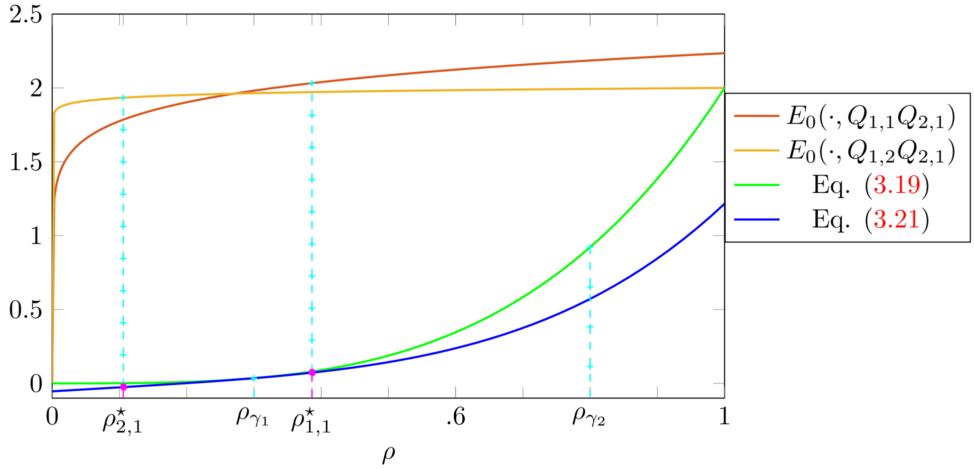

Figure 3.3: The solution of system of equations is derived as $E^{1,1}_{0,s}(\gamma^\star_1, 0) = E^{2,1}_{0,s}(\gamma^\star_1, 0)$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,1}$ are located on $\hat{E}_s(\rho, P_{U_1}, \gamma^\star_1) + E_s(\rho, P_{U_2})$. See Figure 3.2 for more details.

at $\rho^\star_{2,1}$, where $\rho^\star_{1,1}$ and $\rho^\star_{2,1}$ are defined in (3.18).

Since $\gamma^\star_2 = 0$, the $E_{s,1}(\cdot)$ for user 2, is always equal to $E_s(\rho, P_{U_2})$. We define new Gallagher's channel function as $E_0(\cdot) - E_s(\rho, P_{U_2})$. Now, like single-user communication, by moving γ_1, the optimal $\gamma^\star_1$ is obtained at the point where the distances between the new Gallagher's channel functions and $\hat{E}_s(\rho, P_{U_1}, \gamma^\star_1)$ are equal to each other. Figure 3.3 shows the same idea.

Applying Lemma A.6, we can express the achievable exponent by a simpler expression. Setting $E(\rho, Q_1) = E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_s(\rho, P_{U_2})$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_s(\rho, P_{U_2})$ in Lemma A.6, (3.14) for this example can be expressed as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) \\ -E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}_2(\rho, W) - E_s(\rho, P_{U_1}), \tag{3.25}$$

where

$$\bar{E}_2(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{\lambda \max_{i_1=1,2} E_0(\rho_1, Q_{1,i_1}Q_{2,1}, W) - E_s(\rho_1, P_{U_2}) \\ +(1-\lambda) \max_{i_1=1,2} E_0(\rho_2, Q_{1,i_1}Q_{2,1}, W) - E_s(\rho_2, P_{U_2})\Big\}. \tag{3.26}$$

3. No solution for (3.16), while (3.17) has solution:
Similar to the previous case, assume that $\gamma^\star_2$ is derived by solving (3.17),

while (3.16) has no solution, i. e. $\gamma_1^\star = 1$ or $\gamma_1^\star = 0$. Figure 3.4 shows an example where $\gamma_1^\star = 1$, and the solution of (3.17) is obtained such that $E_{0,s}^{2,1}(1,\gamma_2^\star) = E_{0,s}^{2,2}(1,\gamma_2^\star)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_s(\rho, P_{U_1}) - E_{s,1}(\rho, P_{U_2}, \gamma_2^\star) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_s(\rho, P_{U_1}) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star), \quad (3.27)$$

where as shown in Figure 3.4, the distance between $E_0(\rho, Q_{1,2}Q_{2,1}, W)$ and (3.20) at $\rho_{2,1}^\star$ is equal with the distance between $E_0(\rho, Q_{1,2}Q_{2,2}, W)$ and (3.20) at $\rho_{2,2}^\star$, where $\rho_{2,1}^\star$ and $\rho_{2,2}^\star$ are respectively the optimal ρs maximizing the left hand and the right hand sides of (3.27).

The same reasoning applied for the case 2, is also valid for this case, i. e. by moving γ_2, the optimal $\gamma_2^\star$ is derived at the point where the distances between the new Gallagher's channel functions $E_0(\cdot) - E_s(\rho, P_{U_1})$ and $\hat{E}_s(\rho, P_{U_2}, \gamma_2^\star)$ at points $\rho_{2,1}^\star$ and $\rho_{2,2}^\star$ are equal to each other. By setting $E(\rho, Q_1) = E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_s(\rho, P_{U_1})$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_s(\rho, P_{U_1})$ in Lemma A.6, we can express (3.14) as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}_3(\rho, W) - E_s(\rho, P_{U_2}), \quad (3.28)$$

where

$$\bar{E}_3(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda \max_{i_2=1,2} E_0(\rho_1, Q_{1,2}Q_{2,i_2}, W) - E_s(\rho_1, P_{U_1}) + (1-\lambda) \max_{i_2=1,2} E_0(\rho_2, Q_{1,2}Q_{2,i_2}, W) - E_s(\rho_2, P_{U_1}) \Big\}. \quad (3.29)$$

4. Both (3.16) and (3.17) have solutions:
For this case, assume both (3.16) and (3.17) have solutions. Figure 3.5 shows an example of this case where (3.16) is solved such that $E_{0,s}^{1,2}(\gamma_1^\star,\gamma_2^\star) = E_{0,s}^{2,2}(\gamma_1^\star,\gamma_2^\star)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,1}Q_{2,2}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star), \quad (3.30)$$

and (3.17) is solved such that $E_{0,s}^{2,2}(\gamma_1^\star,\gamma_2^\star) = E_{0,s}^{2,1}(\gamma_1^\star,\gamma_2^\star)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star) = \max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,1}(\rho, P_{U_2}, \gamma_2^\star). \quad (3.31)$$

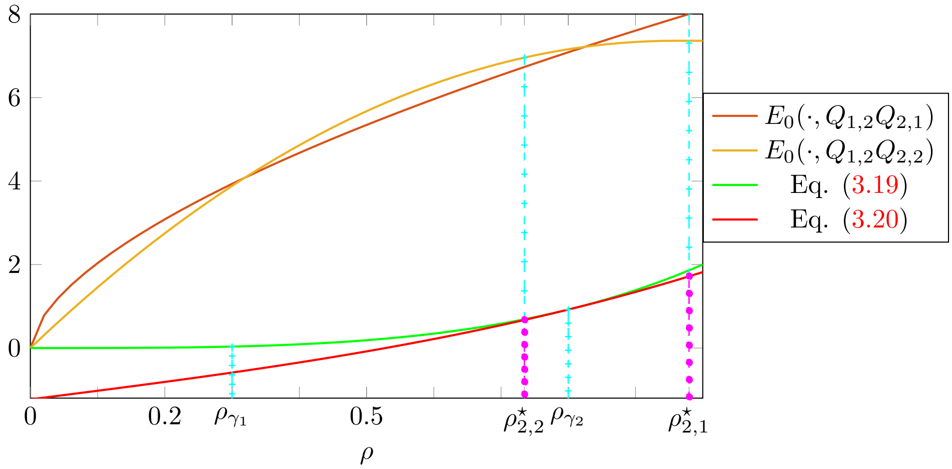

Figure 3.4: The solution of systems of equations is derived as $E^{2,1}_{0,s}(1, \gamma^\star_2) = E^{2,2}_{0,s}(1, \gamma^\star_2)$. Both $\rho^\star_{2,1}$ and $\rho^\star_{2,2}$ are located on $E_s(\rho, P_{U_1}) + \hat{E}_s(\rho, P_{U_2}, \gamma^\star_2)$. See Figure 3.2 for more details.

As shown in Figure 3.5, the distance between $E_0(\rho, Q_{1,1}Q_{2,2}, W)$ and (3.21) at $\rho^\star_{1,2}$ is equal to the distance between $E_0(\rho, Q_{1,2}Q_{2,1}, W)$ and (3.19) at $\rho^\star_{2,1}$, and both are equal to the distance between $E_0(\rho, Q_{1,2}Q_{2,2}, W)$ and (3.22) at $\rho^\star_{2,2}$, where $\rho^\star_{1,2}$, $\rho^\star_{2,1}$ and $\rho^\star_{2,2}$ are defined in (3.18).

Consider (3.30), i. e. $E^{1,2}_{0,s}(\gamma^\star_1, \gamma^\star_2) = E^{2,2}_{0,s}(\gamma^\star_1, \gamma^\star_2)$. We define new Gallager's channel function as $E_0(\cdot) - E_{s,2}(\rho, P_{U_2}, \gamma^\star_2)$. Then, by moving γ_1 along the interval $[0, 1]$, the optimal $\gamma^\star_1$ is derived at the point where the distances between these new Gallager's channel functions and the tangent line $\hat{E}_s(\rho, P_{U_1}, \gamma_1)$ are equal to each other.

Applying Lemma A.6 in (3.30), we set $E(\rho, Q_1) = E_0(\rho, Q_{1,1}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_2}, \gamma^\star_2)$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_2}, \gamma^\star_2)$ in Lemma A.6. Then, (3.14) for this example is obtained as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}^a_4(\rho, W) - E_s(\rho, P_{U_1}), \quad (3.32)$$

where

$$\bar{E}^a_4(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda \max_{i_1=1,2} E_0(\rho_1, Q_{1,i_1}Q_{2,2}, W) - E_{s,2}(\rho_1, P_{U_2}, \gamma^\star_2) + (1-\lambda) \max_{i_1=1,2} E_0(\rho_2, Q_{1,i_1}Q_{2,2}, W) - E_{s,2}(\rho_2, P_{U_2}, \gamma^\star_2) \Big\}. \quad (3.33)$$

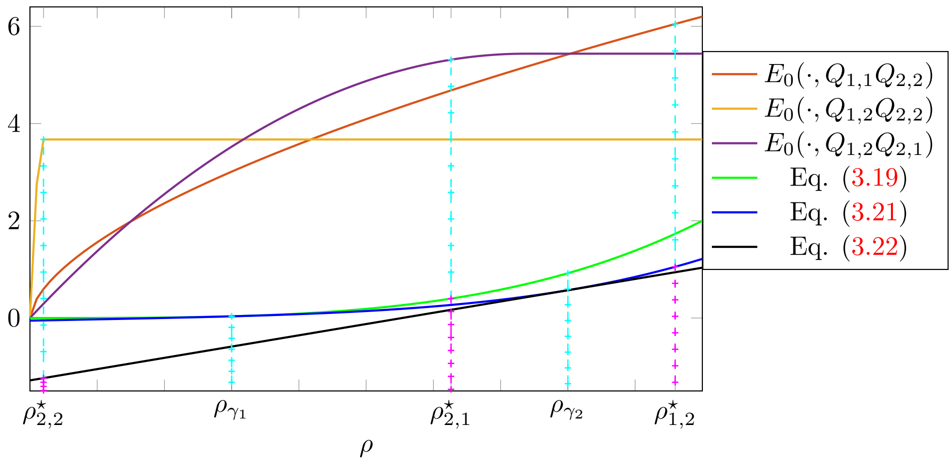

Figure 3.5: The solution of system of equations is derived as $E^{1,2}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,2}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,1}_{0,s}(\gamma_1^\star, \gamma_2^\star)$. $\rho^\star_{1,2}$, $\rho^\star_{2,1}$ and $\rho^\star_{2,2}$, respectively are located on $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + E_s(\rho, P_{U_2})$, $E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2})$ and the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma_1^\star) + \hat{E}_s(\rho, P_{U_1}, \gamma_1^\star)$.

We can also consider (3.31), i. e. $E^{2,2}_{0,s}(\gamma_1^\star, \gamma_2^\star) = E^{2,1}_{0,s}(\gamma_1^\star, \gamma_2^\star)$, and define new Gallager's channel function as $E(\rho, Q_1) = E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star)$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,1}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star)$ in Lemma A.6. Hence, (3.14) is simplified as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W)$$
$$-E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}^b_4(\rho, W) - E_s(\rho, P_{U_2}), \quad (3.34)$$

where

$$\bar{E}^b_4(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{\lambda \max_{i_2=1,2} E_0(\rho_1, Q_{1,2}Q_{2,i_2}, W) - E_{s,2}(\rho_1, P_{U_1}, \gamma_1^\star)$$
$$+ (1-\lambda) \max_{i_1=1,2} E_0(\rho_2, Q_{1,2}Q_{2,i_2}, W) - E_{s,2}(\rho_2, P_{U_1}, \gamma_1^\star)\Big\}. \quad (3.35)$$

5. Both (3.16) and (3.17) give the same answer, and $\rho_{\gamma_1^\star} \neq \rho_{\gamma_2^\star}$:

For this case, again both (3.16) and (3.17) have solutions; however, both of them give the same answer. Figure 3.6 shows an example for this case, where the solution of (3.16) gives $E_{0,s}^{1,1}(\gamma_1^\star,\gamma_2^\star) = E_{0,s}^{2,2}(\gamma_1^\star,\gamma_2^\star)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,1}(\rho, P_{U_2}, \gamma_2^\star) =$$
$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star), \qquad (3.36)$$

and the solution of (3.17) gives again (3.36).

As shown in Figure 3.6, the distance between $E_0(\rho, Q_{1,1}Q_{2,1}, W)$ and (3.22) at $\rho_{1,1}^\star$ is equal to the distance between $E_0(\rho, Q_{1,2}Q_{2,2}, W)$ and (3.22) at $\rho_{2,2}^\star$, where $\rho_{1,1}^\star$ and $\rho_{2,2}^\star$ are defined by (3.18).

Like before, we can define new Gallagher's channel functions as $E(\rho, Q_1) = E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_{s,1}(\rho, P_{U_2}, \gamma_2^\star)$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_2}, \gamma_2^\star)$. Then, by moving $\gamma_1 \in [0,1]$, the optimal $\gamma_1^\star$ is obtained at the point, where the distance between $E(\rho, Q_1)$ and the tangent line $\hat{E}_s(\rho, P_{U_1}, \gamma_1)$ is equal to the distance between $E(\rho, Q_2)$ and the tangent line $\hat{E}_s(\rho, P_{U_1}, \gamma_1)$. Inserting $E(\rho, Q_1)$ and $E(\rho, Q_2)$ in Lemma A.6, we can express (3.14) as

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W)$$
$$-E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}_5^a(\rho, W) - E_s(\rho, P_{U_1}), \quad (3.37)$$

where

$$\bar{E}_5^a(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda \max_{i=1,2} E_0(\rho_1, Q_{1,i}Q_{2,i}, W) - E_{s,1}(\rho_1, P_{U_2}, \gamma_2^\star)$$
$$+ (1-\lambda) \max_{i=1,2} E_0(\rho_2, Q_{1,i}Q_{2,i}, W) - E_{s,2}(\rho_2, P_{U_2}, \gamma_2^\star) \Big\}. \qquad (3.38)$$

Alternatively, we can set $E(\rho, Q_1) = E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1^\star)$ and $E(\rho, Q_2) = E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1^\star)$, which gives the following exponent

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W)$$
$$-E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}_5^b(\rho, W) - E_s(\rho, P_{U_2}), \quad (3.39)$$

where

$$\bar{E}_5^b(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda \max_{i=1,2} E_0(\rho_1, Q_{1,i}Q_{2,i}, W) - E_{s,1}(\rho_1, P_{U_1}, \gamma_1^\star)$$
$$+ (1-\lambda) \max_{i=1,2} E_0(\rho_2, Q_{1,i}Q_{2,i}, W) - E_{s,2}(\rho_2, P_{U_1}, \gamma_1^\star) \Big\}. \qquad (3.40)$$

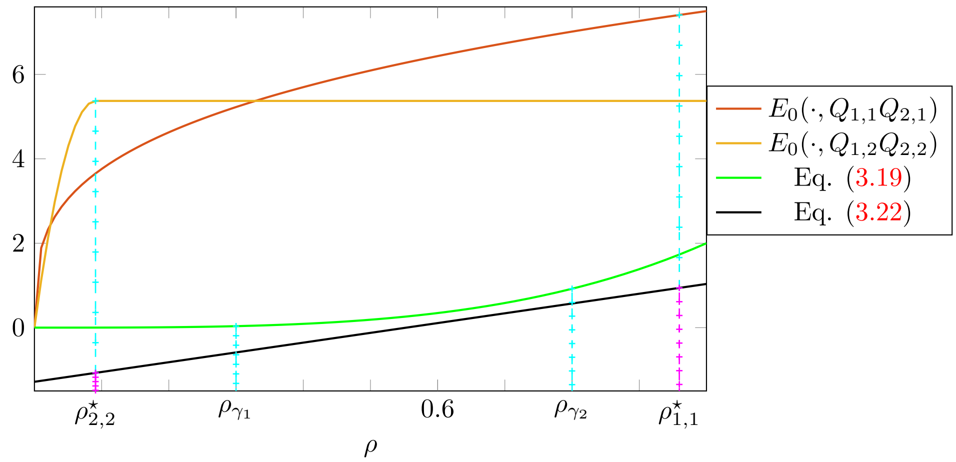

Figure 3.6: $E^{1,1}_{0,s}(\gamma^\star_1, \gamma^\star_2) = E^{2,2}_{0,s}(\gamma^\star_1, \gamma^\star_2)$, while $\rho_{\gamma^\star_1} \neq \rho_{\gamma^\star_2}$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,2}$ are located on the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma^\star_1) + \hat{E}_s(\rho, P_{U_2}, \gamma^\star_2)$.

6. Both (3.16) and (3.17) give the same answer and $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$: Figure 3.7 shows an example of this case where the solutions of both (3.16) and (3.17) gives $E^{1,1}_{0,s}(\gamma^\star_1, \gamma^\star_2) = E^{2,2}_{0,s}(\gamma^\star_1, \gamma^\star_2)$, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,1}Q_{2,1}, W) - E_{s,1}(\rho, P_{U_1}, \gamma^\star_1) - E_{s,1}(\rho, P_{U_2}, \gamma^\star_2) =$$
$$\max_{\rho\in[0,1]} E_0(\rho, Q_{1,2}Q_{2,2}, W) - E_{s,2}(\rho, P_{U_1}, \gamma^\star_1) - E_{s,2}(\rho, P_{U_2}, \gamma^\star_2), \tag{3.41}$$

and also we have $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$, which leads that all (3.20),(3.21) and (3.22) be tangent to (3.19) at $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$. As shown in Figure 3.7, the distance between $E_0(\rho, Q_{1,1}Q_{2,1}, W)$ and (3.22) at $\rho^\star_{1,1}$ is equal to the distance between $E_0(\rho, Q_{1,2}Q_{2,2}, W)$ and (3.22) at $\rho^\star_{2,2}$, where $\rho^\star_{1,1}$ and $\rho^\star_{2,2}$ are given by (3.18).

To simplify the exponent, like before we can use Lemma A.6 two times. However, from Figure 3.7 it can be seen easily that since the distances between two $E_0(\cdot)$ functions and the straight line, $\hat{E}_s(\rho, P_{U_1}, \gamma_1) + \hat{E}_s(\rho, P_{U_2}, \gamma_2)$ are the same, the distance between the parallel line and $\hat{E}_s(\rho, P_{U_1}, \gamma_1) + \hat{E}_s(\rho, P_{U_2}, \gamma_2)$ is equal to the exponent. Moreover, since $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$, (3.22) is tangent to $E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2})$ at $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$. Recalling that, the exponent is the distance between two parallel lines, we focus on the their distance at $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$. Hence, for this example, the following exponent is achievable

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{1,i_1}Q_{2,i_2}, W) - E_{s,i_1}(\rho, P_{U_1}, \gamma_1)$$
$$-E_{s,i_2}(\rho, P_{U_2}, \gamma_2) = \max_{\rho\in[0,1]} \bar{E}_6(\rho, W) - E_s(\rho, P_{U_1}) - E_s(\rho, P_{U_2}), \tag{3.42}$$

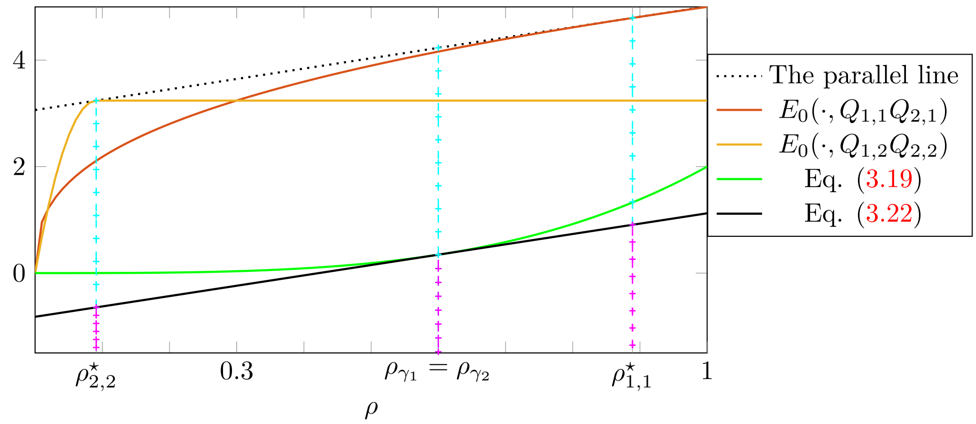

Figure 3.7: $E^{1,1}_{0,s}(\gamma^\star_1, \gamma^\star_2) = E^{2,2}_{0,s}(\gamma^\star_1, \gamma^\star_2)$, and $\rho_{\gamma^\star_1} = \rho_{\gamma^\star_2}$. Both $\rho^\star_{1,1}$ and $\rho^\star_{2,2}$ are located on the straight line $\hat{E}_s(\rho, P_{U_1}, \gamma^\star_1) + \hat{E}_s(\rho, P_{U_2}, \gamma^\star_2)$.

where

$$\bar{E}_6(\rho, W) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \Big\{ \lambda \max_{i=1,2} E_0(\rho_1, Q_{1,i}Q_{2,i}, W) + (1-\lambda) \max_{i=1,2} E_0(\rho_2, Q_{1,i}Q_{2,i}, W) \Big\}. \tag{3.43}$$

All the other possible outcomes of (3.16) and (3.17), can be considered as one of the six cases studied above. We see that, even for $\tau = \{1, 2\}$, finding a unique expression of exponent that does not depend on $\gamma^\star_1$ and $\gamma^\star_2$, is difficult. However, from intuition point of view, these six cases can be easily generalized to (3.13). Since expressing the final exponent atleast is coupled with one of the thresholds, it seems that finding an equation like (2.39) is difficult for the MAC, and maybe unlike single-user communication, the sufficient number of classes is not two.

Even though $\gamma^\star_1$ and $\gamma^\star_2$ can be computed through equation (3.13), the final expression of the achievable exponent (3.6) is still coupled with $\gamma^\star_1$ and $\gamma^\star_2$. In the sequel, we alternatively study both lower and an upper bounds that do not depend on γ_1 and γ_2.

3.2.2 A Lower Bound for the Achievable Exponent

In order to find a lower bound for the achievable exponent presented in (3.6), we use properties (2.35) and (2.36). Firstly, we maximize over $\gamma_\nu \in \{0, 1\}$

rather than $\gamma_\nu \in [0,1]$, to lower bound (3.6). Let $d(\gamma_1,\gamma_2)$ be

$$d(\gamma_1,\gamma_2) = \min_{i_1,i_2} f_{i_1,i_2}(\gamma_1,\gamma_2). \tag{3.44}$$

Then,

$$E = \max_{\gamma_1,\gamma_2\in[0,1]} d(\gamma_1,\gamma_2) \geq \max_{\gamma_1,\gamma_2\in\{0,1\}} d(\gamma_1,\gamma_2). \tag{3.45}$$

On the other hand,

$$\max_{\gamma_1,\gamma_2\in\{0,1\}} d(\gamma_1,\gamma_2) = \max\{d(0,0), d(0,1), d(1,0), d(1,1)\}. \tag{3.46}$$

Taking into account properties (2.35) and (2.36), we note that $f_{i_1,i_2}(\gamma_1,\gamma_2)$, for $\gamma_1,\gamma_2 \in \{0,1\}$, is either infinity, or the Gallager's source-channel exponent, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau}). \tag{3.47}$$

For example, $f_{i_1,i_2}(0,1)$ equals equation (3.47) for $i_1 = 1$ and $i_2 = 2$, and $f_{i_1,i_2}(0,1) = \infty$ for the rest of combinations of i_1 and i_2. Thus, $d(0,1) = \min_\tau \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau})$ for $i_1 = 1$ and $i_2 = 2$. Similarly, $d(1,0) = \min_\tau \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau})$ for $i_1 = 2$ and $i_2 = 1$, and so on. Hence, we obtain the following lower bound

$$E \geq E_{\mathrm{L}}\left(P_{U_1}P_{U_2}, W\right), \tag{3.48}$$

where

$$E_{\mathrm{L}}\left(P_{U_1}P_{U_2}, W\right) = \max_{i_1\in\{1,2\}} \max_{i_2\in\{1,2\}} \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}, \tag{3.49}$$

with

$$F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}} = \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau}). \tag{3.50}$$

We note that for $\tau = \{1\}$ and $\tau = \{2\}$, $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$ in (3.50) is the error exponent of the point-to-point channel $WQ_{\tau^c,i_{\tau^c}}$ for an iid random-coding ensemble with distribution $Q_{\tau,i}$. For $\tau = \{1,2\}$, we have $WQ_{\tau^c,i_{\tau^c}} = W$ and $E_s(\rho, P_{U_\tau}) = E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2})$, so that (3.50) is the error exponent of the point-to-point channel W for an iid random-coding ensemble with distribution $Q_{1,i_1}Q_{2,i_2}$. Hence, the lower bound (3.49) selects the best assignment of input distributions over all four combinations through i_1 and i_2.

3.2.3 An Upper Bound for the Achievable Exponent

Now, we derive an upper bound for (3.6) inspired by the tools used in [16] for single user communication. For the MAC with independent sources, we use the max-min inequality [22] to upper-bound (3.6) by swapping the maximization over γ_1,γ_2 with the minimization over τ. Then, for a given τ, we use Lemma 3.2 in Section 3.4.3 to obtain the following result.

Proposition 3.3. *The achievable exponent* (3.6) *is upper bounded as*

$$E \leq E_{\mathrm{U}}(P_{U_1}P_{U_2}, W), \tag{3.51}$$

where

$$E_{\mathrm{U}}(P_{U_1}P_{U_2}, W) = \min_{\tau \in \{\{1\},\{2\},\{1,2\}\}} F_\tau^{\mathrm{U}}, \tag{3.52}$$

where

$$F_\tau^{\mathrm{U}} = \max_{i_{\tau^c}=1,2} \max_{\rho \in [0,1]} \bar{E}_0(\rho, \{Q_{\tau,1}, Q_{\tau,2}\}, WQ_{\tau^c, i_{\tau^c}}) - E_s(\rho, P_{U_\tau}), \tag{3.53}$$

where $\bar{E}_0(\cdot)$ *is defined by* (2.4). *We recall that for* $\tau = \{1,2\}$, *we have* $\{Q_{\tau,1}, Q_{\tau,2}\} = \{Q_{1,1}, Q_{2,1}, Q_{1,2}, Q_{2,2}\}$ *and* $E_s(\rho, P_{U_\tau}) = E_s(\rho, P_{U_1}) + E_s(\rho, P_{U_2})$.

Proof. See Section 3.4.3. □

From equation (3.52), we observe that the upper bound is the minimum of three terms depending on $\tau \in \{\{1\},\{2\},\{1,2\}\}$. For $\tau \in \{\{1\},\{2\}\}$, we know that the message of user τ^c is decoded correctly so that user τ is virtually sent either over channel $WQ_{\tau^c,1}$ or $WQ_{\tau^c,2}$. Hence, the objective function of (3.52) is the single-user exponent for source P_{U_τ} and point-to-point channel $WQ_{\tau^c,i_{\tau^c}}$ where codewords are generated according to two assigned input distributions $\{Q_{\tau,1}, Q_{\tau,2}\}$ depending on class index of source messages. As a result, we note that the maximization over $i_{\tau^c} = 1, 2$ is equivalent to choose the best channel (either $WQ_{\tau^c,1}$ or $WQ_{\tau^c,2}$) in terms of error exponent.

As pointed in Lemma 3.2, we note that for error type $\tau = \{1\}$ and $\tau = \{2\}$ we have equality in (3.103). In other words, if we had a magic model that only error type $\tau = \{1\}$ or $\tau = \{2\}$ occurred, then the final exponent would be decoupled from both thresholds. Thus, for the mentioned magic model that only error type $\tau = \{1\}$ or $\tau = \{2\}$ occurs, the sufficient number of thresholds for each user is one. More specifically, in view of (3.103), for error type $\tau = \{1\}$, the sufficient number of classes for the first user is two, and it gives the concave-hull term. However, the classes of the second user determine two channels namely $WQ_{2,1}$ and $WQ_{2,2}$.

3.2.4 Numerical Example

Here we provide a numerical example comparing the achievable exponent, the lower bound and the upper bound given in (3.6), (3.49) and (3.52), respectively. We consider two independent discrete memoryless sources with alphabet $\mathcal{U}_\nu = \{1, 2\}$ for $\nu = 1, 2$ where $P_{U_1}(1) = 0.028$ and $P_{U_2}(1) = 0.01155$. We also consider a discrete memoryless multiple-access channel with $\mathcal{X}_1 = \mathcal{X}_2 = \{1, 2, \ldots, 6\}$ and $|\mathcal{Y}| = 4$. The transition probability of this channel, denoted as W, is given by

$$W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{pmatrix}, \tag{3.54}$$

where

$$W_1 = \begin{pmatrix} 1-3k_1 & k_1 & k_1 & k_1 \\ k_1 & 1-3k_1 & k_1 & k_1 \\ k_1 & k_1 & 1-3k_1 & k_1 \\ k_1 & k_1 & k_1 & 1-3k_1 \\ 0.5-k_2 & 0.5-k_2 & k_2 & k_2 \\ k_2 & k_2 & 0.5-k_2 & 0.5-k_2 \end{pmatrix}, \tag{3.55}$$

for $k_1 = 0.056$ and $k_2 = 0.01$. W_2 and W_3 are 6×4 matrices whose rows are all the copy of 5^{th} and 6^{th} row of matrix W_1, respectively. Let the m-th row of matrix W_1 is denoted by $W_1(m)$. W_4, W_5 and W_6 are respectively given by

$$W_4 = \begin{pmatrix} W_1(2) \\ W_1(3) \\ W_1(4) \\ W_1(1) \\ W_1(6) \\ W_1(5) \end{pmatrix} \quad W_5 = \begin{pmatrix} W_1(3) \\ W_1(4) \\ W_1(1) \\ W_1(2) \\ W_1(5) \\ W_1(6) \end{pmatrix} \quad W_6 = \begin{pmatrix} W_1(4) \\ W_1(1) \\ W_1(2) \\ W_1(3) \\ W_1(6) \\ W_1(5) \end{pmatrix}. \tag{3.56}$$

We observe that W is a 36×4 matrix where the transition probability $W(y|x_1, x_2)$ is placed at the row $x_1 + 6(x_2 - 1)$ of matrix W, for $(x_1, x_2) \in \{1, 2, ..., 6\} \times \{1, 2, ..., 6\}$. Recalling that each source has two classes and that four input distributions generate codewords, there are four possible assignments of input distributions to classes. Among all possible permutations, we select the one that gives the highest exponent. For user $\nu = 1, 2$, we consider the set of input distributions $\Big\{[0\ 0\ 0\ 0\ 0.5\ 0.5],\ [0.25\ 0.25\ 0.25\ 0.25\ 0\ 0]\Big\}$. For the channel given in (3.54), the optimal assignment is

$$Q_{\nu,1} = [0\ 0\ 0\ 0\ 0.5\ 0.5], \tag{3.57}$$

$$Q_{\nu,2} = [0.25\ 0.25\ 0.25\ 0.25\ 0\ 0], \tag{3.58}$$

for both $\nu = 1, 2$. Since we consider two input distributions for each user, the function $\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}})$ is not concave in ρ [16]. For this example, from (3.13), we numerically compute the optimal $\gamma_1^\star$ and $\gamma_2^\star$ maximizing (3.6) leading to $\gamma_1^\star = 0.8159$ and $\gamma_2^\star = 0.7057$.

Tables 3.1, 3.2 and 3.3 respectively show the objective functions $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$, $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$, and F^{U}_{τ} given in (3.7), (3.50) and (3.53), involved in the derivation of the achievable exponent (3.6), lower bound (3.49) and upper bound (3.52). The shaded elements in Tables 3.1 and 3.3 respectively are the exponent and the upper bound. Additionally, the shaded elements in Table 3.2 are the iid exponent for different input distributions assignments. Solving equations (3.6), (3.49), (3.52) using the partial optimization in Tables 3.1, 3.2 and 3.3, we respectively obtain

$$E\left(P_{U_1}P_{U_2}, \{Q_{1,1}, Q_{1,2}\}, \{Q_{2,1}, Q_{2,2}\}, W\right) = 0.1057, \tag{3.59}$$

and

$$E_{\mathrm{L}}(P_{U_1}P_{U_2}, W) = 0.0989, \tag{3.60}$$

$$E_{\mathrm{U}}(P_{U_1}P_{U_2}, W) = 0.1073. \tag{3.61}$$

We observe that the percentage difference between the achievable exponent $E(P_{U_1}P_{U_2}, W)$ and the lower bound $E_{\mathrm{L}}(P_{U_1}P_{U_2}, W)$ is 6.875%. For a given set of two distributions for each user, the lower bound $E_{\mathrm{L}}(P_{U_1}P_{U_2}, W)$ corresponds to the iid random-coding error exponent when each user uses only one input distribution. In Chapter 2, a similar comparison is made for point-to-point communication. In view of (2.44) and (2.45), for single-user communication, the exponent achieved by an ensemble with two distributions is 0.75% higher than the one achieved by the iid ensemble. Hence, our example illustrates that using message-dependent random coding with two class distributions may lead to higher error exponent gain in the MAC than in point-to-point communication, compared to iid random coding.

3.3 Generalizing to Multiple-Classes

In this section, we generalize the main results given in Section 3.2 to the K-user MAC. Moreover, in view of [21], we consider multiple classes for each

Table 3.1: Values of $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ in (3.7) with optimal thresholds $\gamma_1^\star = 0.8159$ $\gamma_2^\star = 0.7057$, for types of error τ, and user classes i_τ and i_{τ^c}.

	(i_1, i_2)			
	(1,1)	(2,1)	(1,2)	(2,2)
$\tau = \{1\}$	0.2566	0.1721	0.1057	0.1103
$\tau = \{2\}$	0.2597	0.1057	0.2526	0.2087
$\tau = \{1, 2\}$	0.1057	0.1073	0.1127	0.1180

Table 3.2: Values of $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$ in (3.50) for types of error τ, and input distribution Q_{1,i_1}, Q_{2,i_2}.

	$Q_{1,1}$,$Q_{2,1}$	$Q_{1,2}$,$Q_{2,1}$	$Q_{1,1}$,$Q_{2,2}$	$Q_{1,2}$,$Q_{2,2}$
$\tau = \{1\}$	0.1723	0.1721	0.0251	0.0342
$\tau = \{2\}$	0.2526	0.0989	0.2526	0.2019
$\tau = \{1, 2\}$	0.0900	0.1073	0.0900	0.0984

Table 3.3: Values of F^{U}_{τ} in (3.53) for types of error τ.

$\tau = \{1\}$	$\tau = \{2\}$	$\tau = \{1, 2\}$
0.1734	0.2526	0.1073

user. For user $\nu = 1, ..., K$, let $0 = \gamma_{\nu,L_\nu+1} \leq \gamma_{\nu,L_\nu} \leq ... \leq \gamma_{\nu,2} < \gamma_{\nu,1} = 1$ be $L_\nu + 1$ positive ordered numbers such that $\gamma_{\nu,L_\nu} > \min P_{U_\nu}(u_\nu)$ and $\gamma_{\nu,2} \leq \max P_{U_\nu}(u_\nu)$. The source-message set $\mathcal{U}_\nu^n$ is partitioned into L_ν classes where the class $\ell_\nu \in \mathcal{L}_\nu = \{1, ..., L_\nu\}$ is defined as

$$\mathcal{D}_\nu^{\ell_\nu} = \Big\{ \boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \;:\; \gamma_{\nu,\ell_\nu+1}^n < P_{\boldsymbol{U}_\nu}^n(\boldsymbol{u}_\nu) \leq \gamma_{\nu,\ell_\nu}^n \Big\}. \tag{3.62}$$

When $L_\nu = 2$ by setting $\gamma_{\nu,2} = \gamma_\nu$, we have $\mathcal{D}_\nu^{\ell_\nu} = \mathcal{A}_\nu^{\ell_\nu}(\gamma_\nu)$ where $\mathcal{A}_\nu^{\ell_\nu}(\gamma_\nu)$ is given by (3.4) and (3.5). For the messages belonging to $\mathcal{D}_\nu^{\ell_\nu}$, input distribution Q_{ν,ℓ_ν} is assigned to generate codewords.

Throughout this section, the underlined font denotes an ordered tuple of quantities for K users, i. e. $\underline{\boldsymbol{U}} = (\boldsymbol{U}_1, .., \boldsymbol{U}_K)$ and $P_{\underline{U}}(\underline{u}) = \prod_{l=1}^K P_{U_l}(u_l)$.

Proposition 3.4. *For the K-user MAC with transition probability W and source probability distributions $P_{U_1}, ..., P_{U_K}$, the following exponent is achiev-*

able

$$E(P_{\underline{U}}, \{Q_{1,1}, ..., Q_{1,L_1}\}, ..., \{Q_{K,1}, ..., Q_{K,L_K}\}, W) = \max_{\gamma_{1,2},...,\gamma_{1,L_1}} \cdots \max_{\gamma_{K,2},...,\gamma_{K,L_K}} \min_{\tau \in \{\{1\},...,\{1,...,K\}\}} \min_{\ell_1 \in \mathcal{L}_1,...,\ell_L \in \mathcal{L}_K} F_{\tau,\ell_\tau,\ell_{\tau^c}}(Q_{1,\ell_1}, ..., Q_{K,\ell_K}, \gamma_{1,1}, ..., \gamma_{K,L_K+1}), \tag{3.63}$$

where for all $\nu = 1, ..., K$, $\gamma_{\nu,1} = 1$ *and* $\gamma_{\nu,L_\nu+1} = 0$. *In addition,*

$$F_{\tau,\ell_\tau,\ell_{\tau^c}}(Q_{1,\ell_1}, ..., Q_{K,\ell_K}, \gamma_{1,1}, ..., \gamma_{K,L_K+1}) = \max_{\rho \in [0,1]} E_0(\rho, Q_{\tau,\ell_\tau}, WQ_{\tau^c,\ell_{\tau^c}}) - E_{s,\ell_\tau}(\rho, P_{U_\tau}, \gamma_{\tau,\ell_\tau+1}, \gamma_{\tau,\ell_\tau}) - E_{s,\ell_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c,\ell_{\tau^c}+1}, \gamma_{\tau^c,\ell_{\tau^c}}), \tag{3.64}$$

where $E_{s,\ell_\nu}(\cdot)$ *is defined by* (2.24).

Proof. See Section 3.4.4. □

Remark 3.1. *As a special case, we consider* K*-user MAC where* $\mathcal{L}_\nu = \{1, 2\}$ *for all* $\nu = 1, ..., K$. *When* $L = 2$, *we have* $\gamma_{\nu,1} = 1$ *and* $\gamma_{\nu,3} = 0$ *leading to* $\rho_{\gamma_{\nu,1}} = -1_+$ *and* $\rho_{\gamma_{\nu,3}} = -1_-$. *Thus, from* (2.24), *we conclude that for two-class source,* $E_{s,1}(\rho, P_{U_\nu}\gamma_{\nu,2}, \gamma_{\nu,1}) = E_{s,1}(\rho, P_{U_\nu}, \gamma_{\nu,2})$ *and* $E_{s,2}(\rho, P_{U_\nu}, \gamma_{\nu,3}, \gamma_{\nu,2}) = E_{s,2}(\rho, P_{U_\nu}, \gamma_{\nu,2})$ *where the functions* $E_{s,1}(\rho, P_{U_\nu}, \gamma_{\nu,2})$ *and* $E_{s,2}(\rho, P_{U_\nu}, \gamma_{\nu,2})$ *are defined by* (2.33) *and* (2.34), *respectively. Thus,* (3.63) *is simplified as*

$$E(P_{\underline{U}}, \{Q_{1,1}, Q_{1,2}\}, ..., \{Q_{K,1}, Q_{K,2}\}, W) = \max_{\gamma_{1,2} \in [0,1]} \cdots \max_{\gamma_{K,2} \in [0,1]} \min_{\tau} \min_{\ell_\tau, \ell_{\tau^c} = 1,2} \max_{\rho \in [0,1]} E_0(\rho, Q_{\tau,\ell_\tau}, WQ_{\tau^c,\ell_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_{\tau,2}) - E_{s,i_{\tau^c}}(\rho, P_{U_{\tau^c}}, \gamma_{\tau^c,2}). \tag{3.65}$$

Remark 3.2. *We recall that for the* K*-user MAC, the error probability for a given ordered tuple of codebooks* $(\mathcal{C}^1, ..., \mathcal{C}^K)$, *is given by* $\mathbb{P}\left[\hat{\underline{\boldsymbol{U}}} \neq \underline{\boldsymbol{U}}\right]$. *We use the symbol* $\tau \in \Big\{\{1\}, ..., \{1, 2, ..., K\}\Big\}$ *to denote the error event type of the error probability. There are* $2^K - 1$ *error event types including the events that the message of one source is decoded wrongly, the events that the messages of two sources are decoded wrongly and so on. For* K*-user MAC, the complement of* τ *is denoted by* τ^c *among the subsets of* $\{1, 2, ..., K\}$.

For example, let $K = 6$ *and* $\tau = \{1, 3, 6\}$. *Then,* $\ell_\tau = (\ell_1, \ell_3, \ell_6)$, $\mathcal{L}_\tau = \mathcal{L}_1 \times \mathcal{L}_3 \times \mathcal{L}_6$ *and therefore* $\ell_{\tau^c} = (\ell_2, \ell_4, \ell_5)$, $\mathcal{L}_{\tau^c} = \mathcal{L}_2 \times \mathcal{L}_4 \times \mathcal{L}_5$. *We note that* $E_{s,\ell_\tau}(\rho, P_{U_\tau}, \cdot) = E_{s,1}(\rho, P_{U_1}, \cdot) + E_{s,3}(\rho, P_{U_3}, \cdot) + E_{s,6}(\rho, P_{U_6}, \cdot)$ *and* Q_{τ,ℓ_τ} *is the product distribution* $Q_{1,\ell_1}(x_1)Q_{3,\ell_3}(x_3)Q_{6,\ell_6}(x_6)$ *over the alphabet* $\mathcal{X}_1 \times \mathcal{X}_3 \times \mathcal{X}_6$. *Similarly,* $WQ_{\tau^c,i_{\tau^c}}$ *is a multiple access channel with input and output alphabets given by* $\mathcal{X}_1 \times \mathcal{X}_3 \times \mathcal{X}_6$ *and* $\mathcal{X}_2 \times \mathcal{X}_4 \times \mathcal{X}_5 \times \mathcal{Y}$, *respectively.*

Now, we extend proposition 3.2 to K-user MAC with multiple classes. In order to find a way to optimize $\gamma_{\nu,2},...,\gamma_{\nu,L_\nu}$ for $\nu = 1,...,K$, we recall that as mentioned in Proposition 3.4, $\gamma_{\nu,1} = 1$ and $\gamma_{\nu,L_\nu+1} = 0$ for all $\nu = 1,...,K$. Hence, to optimize $\gamma_{\nu,2},...,\gamma_{\nu,L_\nu}$, we express (3.63) as

$$E(P_{\underline{U}}, \{Q_{1,1},...,Q_{1,L_1}\},...,\{Q_{K,1},...,Q_{K,L_K}\}, W) = \max_{\gamma_{1,2},\cdots,\gamma_{1,L_1}} \cdots \max_{\gamma_{K,2},\cdots,\gamma_{K,L_K}} \min_{\ell_\tau\in\mathcal{L}_\tau,\ell_{\tau^c}\in\mathcal{L}_{\tau^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1}), \tag{3.66}$$

where

$$f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1}) = \min_\tau F_{\tau,\ell_\tau,\ell_{\tau^c}}\left(Q_{1,\ell_1},...,Q_{K,\ell_K},\gamma_{1,1},....,\gamma_{K,L_K+1}\right), \tag{3.67}$$

where $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)$ is given by (3.64). Since $E_{s,\ell_\nu}\left(\rho, P_{U_\nu}, \gamma_{\nu,\ell_\nu+1}, \gamma_{\nu,\ell_\nu},\right)$ only depends on γ_{ν,ℓ_ν} and $\gamma_{\nu,\ell_\nu+1}$, for $i_\nu = 2,...,L_\nu$, the function $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)\Big|_{\ell_\nu=i_\nu}$ only depends on γ_{ν,i_ν}, $\gamma_{\nu,i_\nu+1}$ and does not change with the rest of the partitioning thresholds. Hence, to determine the optimal γ_{ν,i_ν}, it suffices to consider the objective function at $\ell_\nu = i_\nu$ and $\ell_\nu = i_\nu - 1$.

For $i_\nu = 2,...,L_\nu$, let $\ell_\nu = i_\nu$ for an arbitrary τ. Then, $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)\Big|_{\gamma_{\nu,\ell_\nu}}$ for $\ell_\nu = i_\nu$ is of the form of $\max_\rho E(\rho) - E_{s,i_\nu}\left(\rho, P_{U_\nu}, \gamma_{\nu,i_\nu+1}, \gamma_{\nu,i_\nu}\right)$ and similarly the function $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)\Big|_{\gamma_{\nu,\ell_\nu}}$ for $\ell_\nu = i_\nu - 1$ is of the form of $\max_\rho E(\rho) - E_{s,i_\nu-1}\left(\rho, P_{U_\nu}, \gamma_{\nu,i_\nu}, \gamma_{\nu,i_\nu-1}\right)$.

By using Lemma A.15, we find that $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)\Big|_{\ell_\nu=i_\nu}$ and $F_{\tau,\ell_\tau,\ell_{\tau^c}}(\cdot)\Big|_{\ell_\nu=i_\nu-1}$ are respectively non-increasing and non-decreasing with respect to γ_{ν,i_ν}. Using the fact that the minimum of monotonic functions is monotonic, we conclude that $f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu}$ is non-increasing with respect to γ_{ν,i_ν} and $f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu-1}$ is non-decreasing with respect to γ_{ν,i_ν}. We recall again, $f_{\ell_1,...,\ell_K}(\cdot)$ for $\ell_\nu = i_\nu, i_\nu - 1$ changes with γ_{ν,i_ν} and is constant with respect to γ_{ν,ℓ_ν} for the rest of ℓ_ν. Table 3.4 shows the increasing, decreasing and constant behaviour of $f_{\ell_1,...,\ell_K}(\cdot)$ when $K = 2$, and the first source-message set is partitioned into three classes while the second one is partitioned into two classes.

In order to find the optimal source-partition thresholds, we adopt the following notation. The complement index of $\nu \in \{1,...,K\}$ is denoted by ν^c and γ^c_{ν,ℓ_ν} is the sequence of $\gamma_{\nu,2},...,\gamma_{\nu,L_\nu}$ without term γ_{ν,ℓ_ν}. We define $\bar{\gamma}_{\nu,\ell_\nu} = \gamma_{\nu^c,m_{\nu^c}}\Big|_{m_{\nu^c}=2,..,L_{\nu^c}}, \gamma^c_{\nu,\ell_\nu}$. For example, for three-user MAC, let $L_1 = 3$, $L_2 = 4$ and $L_3 = 2$. Then, for $\nu = 2$ and $\ell_\nu = 3$, we have $\gamma^c_{\nu,\ell_\nu} = \gamma_{2,2}, \gamma_{2,4}$

Table 3.4: The behaviour of $f_{\ell_1,\ell_2}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$ given by (3.67) for a two-user MAC.

	$\gamma_{1,2}$	$\gamma_{1,3}$	$\gamma_{2,2}$		$\gamma_{1,2}$	$\gamma_{1,3}$	$\gamma_{2,2}$
$f_{1,1}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$\nearrow$	$-$	$\nearrow$	$f_{1,2}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$\nearrow$	$-$	$\searrow$
$f_{2,1}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$\searrow$	$\nearrow$	$\nearrow$	$f_{2,2}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$\searrow$	$\nearrow$	$\searrow$
$f_{3,1}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$-$	$\searrow$	$\nearrow$	$f_{3,2}(\gamma_{1,2},\gamma_{1,3},\gamma_{2,2})$	$-$	$\searrow$	$\searrow$

and $\gamma_{\nu^c,m_{\nu^c}}\Big|_{m_{\nu^c}=2,..,L_{\nu^c}} = \gamma_{1,m_1}\Big|_{m_1=2,3}, \gamma_{3,m_3}\Big|_{m_3=2}$, which leads

$$\bar{\gamma}_{\nu,\ell_\nu}\Big|_{\nu=2,\ell_\nu=3} = \gamma_{1,2},\gamma_{1,3},\gamma_{2,2},\gamma_{2,4},\gamma_{3,2}. \tag{3.68}$$

Now, for any $i_\nu \in \{2,...,L_\nu\}$, we will determine the optimal γ_{ν,i_ν}. Since only $f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\ell_\nu=i_\nu}$ and $f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\ell_\nu=i_\nu-1}$ depend on γ_{ν,i_ν}, we split the minimization over ℓ_ν of (3.66), as

$$\begin{aligned}
&E(P_{\underline{U}},\left\{Q_{1,1},...Q_{1,L_1}\right\},...,\{Q_{K,1},...,Q_{K,L_K}\},W) = \\
&\max_{\bar{\gamma}_{\nu,i_\nu}}\max_{\gamma_{\nu,i_\nu}}\min\Bigg\{\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu-1}, \\
&\qquad\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu}, \\
&\qquad\min_{\ell_\nu\in\{\mathcal{L}_\nu-\{i_\nu,i_\nu-1\}\}}\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Bigg\}.
\end{aligned} \tag{3.69}$$

We note that for a given $\bar{\gamma}_{\nu,i_\nu}$, the optimization problem given by (3.69) satisfies Lemma A.8. Setting $\gamma=\gamma_{\nu,i_\nu}$, $k_1(\gamma) = \min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\ell_\nu=i_\nu-1}$ and $k_2(\gamma) = \min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\ell_\nu=i_\nu}$, in Lemma A.8, and noting to the fact that the third term of (3.69), i. e. $\min_{\ell_\nu\in\{\mathcal{L}_\nu-\{i_\nu,i_\nu-1\}\}}\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\cdot)$ does not change with γ_{ν,i_ν}, by applying Lemma A.8, for $i_\nu\in\{2,...,L_\nu\}$, the optimal γ_{ν,i_ν} satisfies

$$\begin{aligned}
&\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu-1,\gamma^\star_{\nu,i_\nu}} = \\
&\qquad\qquad\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma_{1,1},...,\gamma_{K,L_K+1})\Big|_{\ell_\nu=i_\nu,\gamma^\star_{\nu,i_\nu}},
\end{aligned} \tag{3.70}$$

whenever (3.70) has solution. Otherwise, $\gamma^\star_{\nu,i_\nu}=0$ when $f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu-1\\ \gamma_{\nu,i_\nu}=0}} > f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu\\ \gamma_{\nu,i_\nu}=0}}$ or $\gamma^\star_{\nu,i_\nu}=1$ when $f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu-1\\ \gamma_{\nu,i_\nu}=1}} \le f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu\\ \gamma_{\nu,i_\nu}=1}}$.

Since (3.70) is valid for all ν and ℓ_ν, by repeating the approach for given $\bar{\gamma}^\star_{\nu,\ell_\nu}$ we conclude the following proposition.

Proposition 3.5. *Let $\nu = 1, ..., K$, for $i_\nu = 2, ..., L_\nu$, $\gamma^\star_{\nu,i_\nu}$ maximizing (3.63) satisfies*

$$\min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma^\star_{1,1},...,\gamma^\star_{K,L_K+1})\Big|_{\ell_\nu=i_\nu-1} = \min_{\ell_{\nu^c}\in\mathcal{L}_{\nu^c}} f_{\ell_1,...,\ell_K}(\gamma^\star_{1,1},...,\gamma^\star_{K,L_K+1})|_{\ell_\nu=i_\nu}, \tag{3.71}$$

where $\gamma^\star_{\nu,1} = 1$ and $\gamma^\star_{\nu,L_\nu+1} = 0$. When (3.71) has no solutions, then $\gamma^\star_{\nu,i_\nu} = \{0,1\}$. In particular if $f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu-1\\ \gamma_{\nu,i_\nu}=0}} > f_{\ell_1,...,\ell_K}(\cdot)\Big|_{\substack{\ell_\nu=i_\nu\\ \gamma_{\nu,i_\nu}=0}}$, $\gamma^\star_{\nu,i_\nu} = 0$ otherwise $\gamma^\star_{\nu,i_\nu} = 1$.

Remark 3.3. *As a special case, we consider K-user MAC where $\mathcal{L}_\nu = \{1,2\}$ for all $\nu = 1, ..., K$. Since $L_\nu = 2$, $\gamma^\star_{\nu,1} = 1$, $\gamma^\star_{\nu,3} = 0$, it suffices to find $\gamma^\star_{\nu,2}$ in (3.71). Applying the facts that $i_\nu = 2$, (3.71) for K-user MAC with two classes is simplified as*

$$\min_{\ell_{\nu^c}=1,2} f_{\ell_1,...,\ell_K}(\gamma^\star_{1,1},...,\gamma^\star_{K,3})\Big|_{\ell_\nu=1} = \min_{\ell_{\nu^c}=1,2} f_{\ell_1,...,\ell_K}(\gamma^\star_{1,1},...,\gamma^\star_{K,3})|_{\ell_\nu=2}. \tag{3.72}$$

Like Section 3.2.2, we lower bound the exponent given by (3.63), by maximizing over source-partition thresholds belonging to $\{0,1\}$. Considering (2.24), we recall that if $\gamma_{\nu,\ell_\nu+1} = \gamma_{\nu,\ell_\nu} = 1$ or $\gamma_{\nu,\ell_\nu+1} = \gamma_{\nu,\ell_\nu} = 0$, then $E_{s,\ell}(\rho, P_{U_\nu}, \cdot) = -\infty$ [21]. While, if $\gamma_{\nu,\ell_\nu+1} = 0$ and $\gamma_{\nu,\ell_\nu} = 1$, then we find Gallager's source function as $E_{s,\ell}(\rho, P_{U_\nu}, 0, 1) = E_s(\rho, P_{U_\nu})$. As a result, for the case where source-partition thresholds being only zero or one, $f_{\ell_1,...,\ell_K}(\cdot)$ is either infinity, or the Gallager's source-channel exponent, i. e.

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,\ell_\tau}, WQ_{\tau^c,\ell_{\tau^c}}) - E_s(\rho, P_{U_\tau}). \tag{3.73}$$

In fact, by maximizing over $\{0,1\}$, we choose the the Gallager's source-channel exponent for the best assignment of input, i. e. since the source-partition thresholds are ordered as $0 = \gamma_{\nu,L_\nu+1} \leq \gamma_{\nu,L_\nu} \leq ... \leq \gamma_{\nu,2} < \gamma_{\nu,1} = 1$, for user ν, there is an optimal ℓ_ν where by having $\gamma_{\nu,\ell_\nu+1} = 0$ and $\gamma_{\nu,\ell_\nu} = 1$ the maximum Gallager's source-channel exponent is derived. Hence,

$$E(P_{\underline{U}}, \{Q_{1,1}, ..., Q_{1,L_1}\}, ..., \{Q_{K,1}, ..., Q_{K,L_K}\}, W) \geq \max_{\ell_1\in\mathcal{L}_1} ... \max_{\ell_K\in\mathcal{L}_K} d(Q_{1,\ell_1}, ..., Q_{L,\ell_K}), \tag{3.74}$$

where

$$d(Q_{1,\ell_1}, ..., Q_{L,\ell_K}) = \min_{\tau} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,\ell_\tau}, WQ_{\tau^c,\ell_{\tau^c}}) - E_s(\rho, P_{U_\tau}). \tag{3.75}$$

For a K-user MAC with two classes, we only apply $\mathcal{L}_\nu = \{1,2\}$ for all $\nu = 1, ..., K$ in (3.74).

3.4 Proofs

3.4.1 Proof of Proposition 3.1

To prove Proposition 3.1, we follow similar steps than in [16]. Firstly, we start by bounding $\bar{\epsilon}^n$, the average error probability over the ensemble, for a given block length n. Applying the random-coding union bound [23] for joint source-channel coding, we have

$$\bar{\epsilon}^n \leq \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \min\left\{1, \sum_{\hat{\underline{\boldsymbol{u}}}\neq\underline{\boldsymbol{u}}} \mathbb{P}\left[\frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\underline{\boldsymbol{u}}})W^n(\boldsymbol{y}|\hat{\underline{\boldsymbol{X}}})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\underline{\boldsymbol{x}})} \geq 1\right]\right\}, \tag{3.76}$$

where for independent sources, we have $P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}) = P^n_{\boldsymbol{U}_1}(\boldsymbol{u}_1)P^n_{\boldsymbol{U}_2}(\boldsymbol{u}_2)$, and $\hat{\underline{\boldsymbol{x}}}$ has the same distribution as $\underline{\boldsymbol{x}}$ but is independent of $\boldsymbol{y}$.

The summation over $\hat{\underline{\boldsymbol{u}}} \neq \underline{\boldsymbol{u}}$ can be grouped into three types of events, specifically $(\hat{\boldsymbol{u}}_1, \boldsymbol{u}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $(\boldsymbol{u}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$. These three types of error events are denoted by $\tau \in \{\{1\}, \{2\}, \{1, 2\}\}$, respectively. Using the fact that $\min\{1, a + b\} \leq \min\{1, a\} + \min\{1, b\}$, we further bound $\bar{\epsilon}^n$ as

$$\bar{\epsilon}^n \leq \sum_{\tau} \bar{\epsilon}^n_{\tau}, \tag{3.77}$$

where for $P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}) = P^n_{\boldsymbol{U}_1}(\boldsymbol{u}_1)P^n_{\boldsymbol{U}_2}(\boldsymbol{u}_2)$,

$$\bar{\epsilon}^n_{\tau} \leq \sum_{\underline{\boldsymbol{u}}} P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}) \sum_{\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{x}},\boldsymbol{y}) \min\left\{1, \sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau : \frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau)W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)} \geq 1} Q^n_{\tau,\hat{\boldsymbol{u}}_\tau}(\hat{\boldsymbol{x}}_\tau)\right\}, \tag{3.78}$$

and $Q^n_{\tau,\hat{\boldsymbol{u}}_\tau}$ denotes the channel-input distribution corresponding to the source message $\hat{\boldsymbol{u}}_\tau$.

Next, we break the summation over $\underline{\boldsymbol{u}}$ in (3.78) into the summations over the messages belonging to the classes $\mathcal{A}^1_\nu(\gamma_\nu)$, $\mathcal{A}^2_\nu(\gamma_\nu)$ and then summed over all classes. Moreover, by considering the case where codewords are generated according to distributions that depend on the class index of the sources, the outer summation of (3.78), can be written as

$$\sum_{\underline{\boldsymbol{u}}} P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}) \sum_{\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{x}},\boldsymbol{y}) = \sum_{i_1,i_2=1,2} \sum_{\boldsymbol{u}_1 \in \mathcal{A}^{i_1}_1(\gamma_1)} P^n_{\boldsymbol{U}_1}(\boldsymbol{u}_1) \sum_{\boldsymbol{u}_2 \in \mathcal{A}^{i_2}_2(\gamma_2)} P^n_{\boldsymbol{U}_2}(\boldsymbol{u}_2)$$
$$\times \sum_{\underline{\boldsymbol{x}},\boldsymbol{y}} Q^n_{1,i_1}(\boldsymbol{x}_1)Q^n_{2,i_2}(\boldsymbol{x}_2)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2). \tag{3.79}$$

Similarly, the inner summation of (3.78) can be grouped based on the classes of $\hat{\boldsymbol{u}}_\tau$ and then sum over all classes. Applying this fact and in view of Markov's inequality for $s_{i_\tau j_\tau} \geq 0$, the inner summation of (3.78) is bounded as

$$\sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau : \frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau) W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)} \geq 1} Q^n_{\tau, j_\tau}(\hat{\boldsymbol{x}}_\tau) \leq$$

$$\sum_{j_\tau=1,2} \sum_{\hat{\boldsymbol{u}}_\tau \in \mathcal{A}^{j_\tau}_\tau(\gamma_\tau)} \sum_{\hat{\boldsymbol{x}}_\tau} Q^n_{\tau, j_\tau}(\hat{\boldsymbol{x}}_\tau) \left(\frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau) W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)} \right)^{s_{i_\tau j_\tau}}, \tag{3.80}$$

where for $\tau = \{1, 2\}$, we have $j_\tau = j_1, j_2$, and $\hat{\boldsymbol{u}}_\tau \in \mathcal{A}^{j_\tau}_\tau(\gamma_\tau)$ equals to $\hat{\boldsymbol{u}}_1 \in \mathcal{A}^{j_1}_1(\gamma_1), \hat{\boldsymbol{u}}_2 \in \mathcal{A}^{j_2}_2(\gamma_2)$. Inserting (3.80) into the inner minimization of (3.78) and using the inequality $\min\{1, A + B\} \leq \min_{\rho, \rho' \in [0,1]} A^\rho + B^{\rho'}$ for $A, B \geq 0$, $\rho, \rho' \in [0, 1]$, the inner term of (3.78) is derived as

$$\min \left\{ 1, \sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau : \frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau) W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)} \geq 1} Q^n_{\tau, j_\tau}(\hat{\boldsymbol{x}}_\tau) \right\}$$

$$\leq \sum_{j_\tau=1,2} \min_{\rho_{i_\tau j_\tau} \in [0,1]} \frac{G_{j_\tau}(s_{i_\tau j_\tau}, \boldsymbol{x}_{\tau^c}, \boldsymbol{y})^{\rho_{i_\tau j_\tau}}}{\left(P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \right)^{s_{i_\tau j_\tau} \rho_{i_\tau j_\tau}}}, \tag{3.81}$$

where

$$G_{i_\tau}(s, \boldsymbol{x}_{\tau^c}, \boldsymbol{y}) = \sum_{\boldsymbol{u}_\tau \in \mathcal{A}^{i_\tau}_\tau(\gamma_\tau)} \sum_{\boldsymbol{x}_\tau} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^s Q^n_{\tau, i_\tau}(\boldsymbol{x}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_\tau, \boldsymbol{x}_{\tau^c})^s, \tag{3.82}$$

and $\rho_{i_\tau j_\tau} \in [0, 1]$ and $s_{i_\tau j_\tau} \geq 0$. By putting back (3.79) and (3.81) into the respective outer and inner terms of (3.78), the average error probability is bounded as

$$\bar{\epsilon}^n_\tau \leq \sum_{j_\tau=1,2} \sum_{i_1, i_2=1,2} \min_{\rho_{i_\tau j_\tau} \in [0,1]} \sum_{\boldsymbol{y}, \boldsymbol{x}_{\tau^c}} \sum_{\boldsymbol{u}_{\tau^c} \in \mathcal{A}^{i_{\tau^c}}_{\tau^c}(\gamma_{\tau^c})} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) Q^n_{\tau^c, i_{\tau^c}}(\boldsymbol{x}_{\tau^c})$$

$$G_{i_\tau}(1 - s_{i_\tau j_\tau} \rho_{i_\tau j_\tau}, \boldsymbol{x}_{\tau^c}, \boldsymbol{y}) G_{j_\tau}(s_{i_\tau j_\tau}, \boldsymbol{x}_{\tau^c}, \boldsymbol{y})^{\rho_{i_\tau j_\tau}}. \tag{3.83}$$

Applying Hölder's inequality in the form of

$$\sum_i C_i a_i b_i \leq \left(\sum_i C_i a_i^{\frac{1}{p}} \right)^p \left(\sum_i C_i a_i^{\frac{1}{1-p}} \right)^{1-p}, \tag{3.84}$$

for $p \in [0, 1]$, into (3.83), we obtain

$$\bar{\epsilon}^n_\tau \leq \sum_{j_\tau, i_\tau=1,2} \min_{\rho_{i_\tau j_\tau} \in [0,1]} F^n_{i_\tau}\left(1 - s_{i_\tau j_\tau} \rho_{i_\tau j_\tau}, \frac{1}{p_{i_\tau j_\tau}}\right)^{p_{i_\tau j_\tau}} F^n_{j_\tau}\left(s_{i_\tau j_\tau}, \frac{\rho_{i_\tau j_\tau}}{1 - p_{i_\tau j_\tau}}\right)^{1 - p_{i_\tau j_\tau}}, \tag{3.85}$$

where

$$F_{j_\tau}^n(a,b) = \sum_{i_{\tau^c}=1,2} \sum_{\boldsymbol{u}_{\tau^c} \in \mathcal{A}_{\tau^c}^{i_{\tau^c}}(\gamma_{\tau^c})} \sum_{\boldsymbol{x}_{\tau^c}, \boldsymbol{y}} P_{\boldsymbol{U}_{\tau^c}}^n(\boldsymbol{u}_{\tau^c}) Q_{\tau^c, i_{\tau^c}}^n(\boldsymbol{x}_{\tau^c}) G_{j_\tau}(a, \boldsymbol{x}_{\tau^c}, \boldsymbol{y})^b. \tag{3.86}$$

Now, by setting $s_{i_\tau j_\tau} = \frac{1}{1+\rho_{j_\tau}}$, $\rho_{i_\tau j_\tau} = \frac{\rho_{i_\tau}(1+\rho_{j_\tau})}{1+\rho_{i_\tau}}$ and $p_{i_\tau j_\tau} = \frac{1}{1+\rho_{i_\tau}}$, the average error probability can be written as

$$\bar{\epsilon}_\tau^n \leq \sum_{j_\tau, i_\tau = 1,2} \min_{\rho_{i_\tau}, \rho_{j_\tau} \in [0,1]} F_{i_\tau}^n\left(\frac{1}{1+\rho_{i_\tau}}, 1+\rho_{i_\tau}\right)^{\frac{1}{1+\rho_{i_\tau}}} F_{j_\tau}^n\left(\frac{1}{1+\rho_{j_\tau}}, 1+\rho_{j_\tau}\right)^{\frac{\rho_{i_\tau}}{1+\rho_{i_\tau}}}. \tag{3.87}$$

Since $F_{i_\tau}^n(\cdot), F_{j_\tau}^n(\cdot) \geq 0$ and $\frac{1}{1+\rho_{i_\tau}} + \frac{\rho_{i_\tau}}{1+\rho_{i_\tau}} = 1$, by using weighted arithmetic-geometric inequality, (3.87) is bounded as

$$\bar{\epsilon}_\tau^n \leq \sum_{j_\tau, i_\tau = 1}^{2} \min_{\rho_{i_\tau}, \rho_{j_\tau} \in [0,1]} \frac{1}{1+\rho_{i_\tau}} F_{i_\tau}^n\left(\frac{1}{1+\rho_{i_\tau}}, 1+\rho_{i_\tau}\right) + \frac{\rho_{i_\tau}}{1+\rho_{i_\tau}} F_{j_\tau}^n\left(\frac{1}{1+\rho_{j_\tau}}, 1+\rho_{j_\tau}\right), \tag{3.88}$$

where by rearranging the terms of the sum, we have

$$\bar{\epsilon}_\tau^n \leq \sum_{i_\tau=1,2} \min_{\rho_{i_\tau}, \rho_{j_\tau} \in [0,1]} F_{i_\tau}^n\left(\frac{1}{1+\rho_{i_\tau}}, 1+\rho_{i_\tau}\right) \sum_{j_\tau=1,2} \left(\frac{1}{1+\rho_{i_\tau}} + \frac{\rho_{j_\tau}}{1+\rho_{j_\tau}}\right). \tag{3.89}$$

Next, we may use the following Lemma.

Lemma 3.1. *For a given $\rho \in [0,1]$, and $F_{i_\tau}^n(a,b)$ defined in (3.86), the following inequality holds*

$$-\frac{1}{n}\log\left(F_{i_\tau}^n\left(\frac{1}{1+\rho}, 1+\rho\right)\right) \geq \min_{i_{\tau^c}=1,2} E_0\left(\rho, Q_{\tau, i_\tau}, WQ_{\tau^c, i_{\tau^c}}\right) - E_{s,i_\tau}\left(\rho, P_{U_\tau}, \gamma_\tau\right) - E_{s,i_{\tau^c}}\left(0, P_{U_{\tau^c}}, \gamma_{\tau^c}\right) - \frac{1}{n}\log(2), \tag{3.90}$$

where $E_0(\cdot)$ is given by (1.28) and $E_{s,i}(\cdot)$ for $i=1,2$ is given by (2.33) and (2.34).

Proof. To prove Lemma 3.1, we recall that by inserting $G_{i_\tau}\left(\frac{1}{1+\rho}, \boldsymbol{x}_{\tau^c}, \boldsymbol{y}\right)$ defined in (3.82) into (3.86), $F^n_{i_\tau}\left(\frac{1}{1+\rho}, 1+\rho\right)$ can be written as

$$F^n_{i_\tau}\left(\frac{1}{1+\rho}, 1+\rho\right) = \sum_{i_{\tau^c}=1,2} \sum_{\boldsymbol{u}_{\tau^c} \in \mathcal{A}^{i_{\tau^c}}_{\tau^c}(\gamma_{\tau^c})} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) \left(\sum_{\boldsymbol{u}_\tau \in \mathcal{A}^{i_\tau}_\tau(\gamma_\tau)} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^{\frac{1}{1+\rho}} \right)^{1+\rho}$$
$$\times \sum_{\boldsymbol{x}_{\tau^c}, \boldsymbol{y}} Q^n_{\tau^c, i_{\tau^c}}(\boldsymbol{x}_{\tau^c}) \left(\sum_{\boldsymbol{x}_\tau} Q^n_{\tau, i_\tau}(\boldsymbol{x}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_\tau, \boldsymbol{x}_{\tau^c})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (3.91)$$

Applying the identity $\sum_{u\in\mathcal{A}} f(u) = \sum_u f(u)\mathbb{1}\{u \in \mathcal{A}\}$ to the summation over $\boldsymbol{u}_\nu \in \mathcal{A}^{i_\nu}_\nu(\gamma_\nu), \nu = \tau, \tau^c$ of (3.91), we obtain

$$F^n_{i_\tau}\left(\frac{1}{1+\rho}, 1+\rho\right) = \sum_{i_{\tau^c}=1,2} \sum_{\boldsymbol{u}_{\tau^c}} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) \mathbb{1}\left\{\boldsymbol{u}_{\tau^c} \in \mathcal{A}^{i_{\tau^c}}_{\tau^c}(\gamma_{\tau^c})\right\}$$
$$\times \left(\sum_{\boldsymbol{u}_\tau} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^{\frac{1}{1+\rho}} \mathbb{1}\left\{\boldsymbol{u}_\tau \in \mathcal{A}^{i_\tau}_\tau(\gamma_\tau)\right\} \right)^{1+\rho} e^{-E_0\left(\rho, Q^n_{\tau,i_\tau}, W^n Q_{\tau,i_{\tau^c}}\right)}, \quad (3.92)$$

where in (3.92), in view of (1.28) we applied $\sum_b f_b \cdot \left(\sum_a g_a\right)^c = \sum_b \left(\sum_a g_a \cdot f_b^{1/c}\right)^c$ into the first summation of (3.91) and we expressed it in terms of E_0 function.

Next, we focus on the summations over $\boldsymbol{u}_\tau$ and $\boldsymbol{u}_{\tau^c}$ in (3.92). Let $\nu = \tau, \tau^c$, in view of (3.4) and (3.5), for a given $\boldsymbol{u}_\nu$, we have $\mathbb{1}\left\{\boldsymbol{u}_\nu \in \mathcal{A}^1_\nu(\gamma_\nu)\right\} = \mathbb{1}\left\{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) \geq \gamma^n_\nu\right\}$ and $\mathbb{1}\left\{\boldsymbol{u}_\nu \in \mathcal{A}^{i_2}_\nu(\gamma_\nu)\right\} = \mathbb{1}\left\{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) < \gamma^n_\nu\right\}$. Considering this fact and applying $\mathbb{1}\left\{a \leq b\right\} \leq \left(\frac{b}{a}\right)^\lambda$ for $\lambda \geq 0$ to all indicator functions of (3.92), we find that

$$F^n_{i_\tau}\left(\frac{1}{1+\rho}, 1+\rho\right) \leq \min_{\lambda_\tau, \lambda_{\tau^c} \geq 0} \sum_{i_{\tau^c}=1,2} \sum_{\boldsymbol{u}_{\tau^c}} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) \left(\frac{\gamma^n_{\tau^c}}{P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c})} \right)^{(-1)^{i_{\tau^c}}\lambda_{\tau^c}}$$
$$\times \left(\sum_{\boldsymbol{u}_\tau} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^{\frac{1}{1+\rho}} \left(\frac{\gamma^n_\tau}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)} \right)^{\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}} \right)^{1+\rho} e^{-E_0\left(\rho, Q^n_{\tau,i_\tau}, W^n Q_{\tau,i_{\tau^c}}\right)}, \quad (3.93)$$

where in (3.93) we tightened the bound by minimizing the objective function over $\lambda_\tau, \lambda_{\tau^c} \geq 0$.

Using Lemma A.16 in Appendix A, the first and the second terms of (3.93) can be expressed in terms of the $E_{s,i}(\cdot)$ function at $\rho = 0$ and arbitrary ρ, respectively. Doing so, we obtain that

$$F_{i_\tau}^n\left(\frac{1}{1+\rho},1+\rho\right) \leq \sum_{i_{\tau^c}=1,2} e^{E_{s,i_\tau}\left(\rho,P_{\boldsymbol{U}_\tau}^n,\gamma_\tau^n\right)+E_{s,i_{\tau^c}}\left(0,P_{\boldsymbol{U}_{\tau^c}}^n,\gamma_{\tau^c}^n\right)-E_0\left(\rho,Q_{\tau,i_\tau}^n,W^n Q_{\tau,i_{\tau^c}}\right)}. \tag{3.94}$$

Finally, we bound each term in the summation in (3.94) by the maximum term, use that the sources and the channel are memoryless, and taking logarithms, we obtain to (3.90). □

Next, upper bounding (3.89) by the maximum term over i_τ, further upper bounding by the worst type of error τ, taking logarithms and using (3.90), after some mathematical manipulations we find that the exponential decay of $\bar{\epsilon}^n$ is given by

$$-\frac{1}{n}\log(\bar{\epsilon}^n) \geq \min_{\tau}\min_{i_\tau,i_{\tau^c}}\max_{\rho\in[0,1]} E_0\left(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i_{\tau^c}}\right)$$
$$-E_{s,i_\tau}\left(\rho,P_{U_\tau},\gamma_\tau\right)-E_{s,i_{\tau^c}}\left(0,P_{U_{\tau^c}},\gamma_{\tau^c}\right)-\frac{\log(o(n))}{n}, \tag{3.95}$$

where $o(n)$ is a sequence satisfying $\lim_{n\to\infty}\frac{o(n)}{n}=0$. Using the following properties

$$\liminf_{n\to\infty}(a_n+b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \tag{3.96}$$
$$\liminf_{n\to\infty}\min\{a_n,b_n\} = \min\left\{\liminf_{n\to\infty} a_n, \liminf_{n\to\infty} b_n\right\}, \tag{3.97}$$
$$\liminf_{n\to\infty}\max\{a_n\} \geq \max\left\{\liminf_{n\to\infty} a_n\right\}, \tag{3.98}$$

we further obtain that

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\bar{\epsilon}^n) \geq \min_{\tau}\min_{i_\tau,i_{\tau^c}}\max_{\rho\in[0,1]} E_0\left(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i_{\tau^c}}\right)$$
$$-E_{s,i_\tau}\left(\rho,P_{U_\tau},\gamma_\tau\right)-E_{s,i_{\tau^c}}\left(0,P_{U_{\tau^c}},\gamma_{\tau^c}\right). \tag{3.99}$$

Finally, we optimize (3.99) over γ_ν for $\nu=1,2$. This concludes the proof.

3.4.2 Proof of Proposition 3.2

Now, we focus on $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ given in (3.7). Let $i_1=1$ for an arbitrary τ. Since γ_1 and γ_2 are independent from each other, regardless the value of i_2, the function $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ is of the form $\max_\rho E(\rho)-E_{s,1}(\rho,P_{U_1},\gamma_1)$. Then, using Lemma A.7, we have that $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ is non-decreasing with respect to γ_1. Similarly, when $i_1=2$, we have that $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ is of the form $\max_\rho E(\rho)-$

$E_{s,2}(\rho, P_{U_1}, \gamma_1)$ so that it is non-increasing with respect to γ_1. The same reasoning applies for i_2. That is, $F_{\tau,i_\tau,i_{\tau^c}}(\cdot)$ is non-decreasing with respect to γ_2 for $i_2 = 1$, and non-increasing with respect to γ_2 for $i_2 = 2$, always regardless of the value of i_1.

Using the fact that the minimum of monotonic functions is monotonic, we conclude that $f_{i_1,i_2}(\underline{\gamma})$ given in (3.11) is non-decreasing with respect to γ_1 when $i_1 = 1$, and non-increasing with respect to γ_1 when $i_1 = 2$. Similarly, $f_{i_1,i_2}(\underline{\gamma})$ is non-decreasing (non-increasing) with respect to γ_2 when $i_2 = 1$ ($i_2 = 2$).

Writing equation (3.12) as

$$\max_{\gamma_1} \max_{\gamma_2} \min_{i_2} \min_{i_1} f_{i_1,i_2}(\gamma_1, \gamma_2), \tag{3.100}$$

for a fixed γ_1, the optimization problem $\max_{\gamma_2} \min_{i_2} \min_{i_1} f_{i_1,i_2}(\gamma_1, \gamma_2)$ satisfies Lemma A.8 with $\gamma = \gamma_2$, $i = i_2$, and $k_i(\gamma) = \min_{i_1} f_{i_1,i}(\gamma_1, \gamma)$. Therefore, the optimal $\gamma_2^\star$ satisfies

$$\min_{i_1=1,2} f_{i_1,1}(\gamma_1, \gamma_2^\star) = \min_{i_1=1,2} f_{i_1,2}(\gamma_1, \gamma_2^\star), \tag{3.101}$$

whenever (3.101) has solution. Otherwise, we have $\gamma_2^\star = 0$ when $f_{i_1,1}(\gamma_1, 0) > f_{i_1,2}(\gamma_1, 0)$, or $\gamma_2^\star = 1$ when $f_{i_1,1}(\gamma_1, 0) \leq f_{i_1,2}(\gamma_1, 0)$.

Setting $\gamma_2 = \gamma_2^\star$, the optimization problem $\max_{\gamma_1} \min_{i_1} \min_{i_2} f_{i_1,i_2}(\gamma_1, \gamma_2^\star)$ satisfies Lemma A.8 with $\gamma = \gamma_1$, $i = i_1$, and $k_i(\gamma) = \min_{i_2} f_{i,i_2}(\gamma, \gamma_2^\star)$. Hence, $\gamma_1^\star$ maximizing (3.12) satisfies

$$\min_{i_2=1,2} f_{1,i_2}(\gamma_1^\star, \gamma_2^\star) = \min_{i_2=1,2} f_{2,i_2}(\gamma_1^\star, \gamma_2^\star), \tag{3.102}$$

and in the case (3.102) does not have solution, $\gamma_1^\star = 0$ when $f_{1,i_2}(0, \gamma_2) > f_{2,i_2}(0, \gamma_2)$, or $\gamma_1^\star = 1$ otherwise. Combining (3.101) and (3.102) we obtain (3.13).

3.4.3 Proof of Proposition 3.3

In view of the max-min inequality [22], after upper bounding (3.6) by swapping the maximization over γ_1,γ_2 with the minimization over τ, the upper bound given by (3.52), follows immediately from the following Lemma.

Lemma 3.2. *For a given $\tau = \{\{1\}, \{2\}, \{1, 2\}\}$, we have*

$$\max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_\tau, i_{\tau^c}=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau)$$
$$-E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c}) \leq \max_{i_{\tau^c}=1,2} \max_{\rho\in[0,1]} \bar{E}_0(\rho, \{Q_{\tau,1}, Q_{\tau,2}\}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau}), \tag{3.103}$$

where equality holds for $\tau = \{\{1\}, \{2\}\}$.

Proof. Firstly, we consider $\tau = \{\{1\},\{2\}\}$. In this case, by focusing on the optimization problem given on the left hand side of (3.103), we may note that since $E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$ does not depend on ρ, the maximization over ρ of the left hand side of (3.103) is done independently from $E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$. Additionally, in view of (2.35) and (2.36), we may note that by moving γ_{τ^c} along the $[0,1]$ interval, $E_{s,1}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$ decreases from zero to $-\infty$, while $E_{s,2}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$ increases from $-\infty$ to zero. Hence, the minimum over i_τ and i_{τ^c} of

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau) - E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c}), \tag{3.104}$$

is attained at $\gamma_{\tau^c} = 0$ for $i_{\tau^c} = 1$, or $\gamma_{\tau^c} = 1$ for $i_{\tau^c} = 2$, both leading to $E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c}) = 0$. As a result, it is sufficient to consider $\max_{\gamma_{\tau^c}\in\{0,1\}}$ instead of $\max_{\gamma_{\tau^c}\in[0,1]}$. This implies that the left hand side of (3.103) can be written as

$$\max\Bigg\{ \max_{\gamma_\tau\in[0,1]} \min_{i_\tau} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau)\Big|_{i_{\tau^c}=1,\gamma_{\tau^c}=0},$$
$$\max_{\gamma_\tau\in[0,1]} \min_{i_\tau} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau)\Big|_{i_{\tau^c}=2,\gamma_{\tau^c}=1}\Bigg\}, \tag{3.105}$$

or equivalently

$$\max_{i_{\tau^c}=1,2} \max_{\gamma_\tau\in[0,1]} \min_{i_\tau=1,2} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau). \tag{3.106}$$

Equation (3.106) can be interpreted as an achievable exponent for a point-to-point channel with transition-probability $WQ_{\tau^c,i_{\tau^c}}$, a pair of distributions $\{Q_{\tau,1}, Q_{\tau,2}\}$ and a partition of the source message set into two classes. This problem is well-studied in [16]. In fact, i_{τ^c} in (3.106) is just a parameter selecting either $WQ_{\tau^c,1}$ or $WQ_{\tau^c,2}$. From [16, Theorem 2] or Lemma A.6, equation (3.106) is equal to

$$\max_{i_{\tau^c}=1,2} \max_{\rho\in[0,1]} \bar{E}_0(\rho, \{Q_{\tau,1}, Q_{\tau,2}\}, WQ_{\tau^c,i_{\tau^c}}) - E_s(\rho, P_{U_\tau}), \tag{3.107}$$

which leads (3.103) for type $\tau \in \{\{1\},\{2\}\}$.

For $\tau = \{1,2\}$, in view of the min-max inequality [22], we upper bound the left hand side of (3.103) by swapping the maximization over γ_2 with the minimization over i_1 as

$$\max_{\gamma_1\in[0,1]} \min\Big\{T_1(\gamma_1), T_2(\gamma_1)\Big\}, \tag{3.108}$$

where

$$T_1(\gamma_1) = \max_{\gamma_2 \in [0,1]} \min_{i_2=1,2} \max_{\rho \in [0,1]} E_0(\rho, Q_{1,1}Q_{2,i_2}, W) \\ - E_{s,1}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2), \quad (3.109)$$

and

$$T_2(\gamma_1) = \max_{\gamma_2 \in [0,1]} \min_{i_2=1,2} \max_{\rho \in [0,1]} E_0(\rho, Q_{1,2}Q_{2,i_2}, W) \\ - E_{s,2}(\rho, P_{U_1}, \gamma_1) - E_{s,i_2}(\rho, P_{U_2}, \gamma_2). \quad (3.110)$$

We note that $E_{s,1}(\rho, P_{U_1}, \gamma_1)$ in (3.109) does not change with i_2 and γ_2. Thus, the optimization problem (3.109) can be seen as a refined achievable exponent for a point-to-point channel with a new E_0 function as $E_0(\rho, Q_{1,1}Q_{2,i_2}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1)$ having two input distributions $\{Q_{1,1}Q_{2,1}, Q_{1,1}Q_{2,2}\}$, and a partition of a source message into two classes. Equation (3.109) can be written in terms of the concave hull of $\max_{i_2 \in \{1,2\}} E_0(\rho, Q_{1,1}Q_{2,i_2}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1)$. Since $E_{s,1}(\rho, P_{U_1}, \gamma_1)$ is a convex function with respect to ρ, using Lemma A.13 we upper bound the concave hull of $\max_{i_2 \in \{1,2\}} E_0(\rho, Q_{1,1}Q_{2,i_2}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1)$ by $\bar{E}_0(\rho, \{Q_{1,1}Q_{2,1}, Q_{1,1}Q_{2,2}\}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1)$. Therefore, from applying [16, Theorem 2], $T_1(\gamma_1)$ is upper bounded as

$$T_1(\gamma_1) \leq \max_{\rho \in [0,1]} \bar{E}_0(\rho, \{Q_{1,1}Q_{2,1}, Q_{1,1}Q_{2,2}\}, W) - E_{s,1}(\rho, P_{U_1}, \gamma_1) - E_s(\rho, P_{U_2}). \quad (3.111)$$

Similarly,

$$T_2(\gamma_1) \leq \max_{\rho \in [0,1]} \bar{E}_0(\rho, \{Q_{1,2}Q_{2,1}, Q_{1,2}Q_{2,2}\}, W) - E_{s,2}(\rho, P_{U_1}, \gamma_1) - E_s(\rho, P_{U_2}). \quad (3.112)$$

Inserting the right hand sides of (3.111) and (3.112) into (3.108), we obtain

$$\max_{\gamma_1 \in [0,1]} \min\Big\{T_1(\gamma_1), T_2(\gamma_1)\Big\} \leq \max_{\gamma_1} \min_{i_1 \in \{1,2\}} \max_{\rho \in [0,1]} \bar{E}_0(\rho, \{Q_{1,i_1}Q_{2,1}, Q_{1,i_1}Q_{2,2}\}, W) \\ - E_{s,i_1}(\rho, P_{U_1}, \gamma_1) - E_s(\rho, P_{U_2}). \quad (3.113)$$

Again, the right hand side of (3.113) can be written in terms of the concave hull of the function $\bar{E}_0(\rho, \{Q_{1,i_1}Q_{2,1}, Q_{1,i_1}Q_{2,2}\}, W) - E_s(\rho, P_{U_2})$. Since $E_s(\rho, P_{U_2})$ is convex in ρ, we apply Lemma A.13 again to upper bound the concave hull of $\bar{E}_0(\rho, \{Q_{1,i_1}Q_{2,1}, Q_{1,i_1}Q_{2,2}\}, W) - E_s(\rho, P_{U_2})$ by the function $\bar{E}_0(\rho, \{Q_{1,1}, Q_{1,2}, Q_{2,1}, Q_{2,2}\}, W) - E_s(\rho, P_{U_2})$. Finally using [16, Theorem 2], we obtain that (3.108) is upper bounded by

$$\max_{\rho} \bar{E}_0(\rho, \{Q_{1,1}, Q_{1,2}, Q_{2,1}, Q_{2,2}\}, W) - E_s(\rho, P_{U_1}) - E_s(\rho, P_{U_2}). \quad (3.114)$$

□

3.4.4 Proof of Proposition 3.4

In order to prove Proposition 3.4, we follow the same steps given in Section 3.4.1. Adopting the underlined notation for K users, and applying random-coding union bound [23] for joint source-channel coding, $\bar{\epsilon}^n$ the average error probability over the ensemble is bound as

$$\bar{\epsilon}^n \leq \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \min\left\{1, \sum_{\hat{\underline{\boldsymbol{u}}}\neq\underline{\boldsymbol{u}}} \mathbb{P}\left[\frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\underline{\boldsymbol{u}}})W^n(\boldsymbol{y}|\hat{\underline{\boldsymbol{X}}})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\underline{\boldsymbol{x}})} \geq 1\right]\right\}, \tag{3.115}$$

which is the same as bound given by (3.76). Like before, by using the fact that $\min\{1,a+b\} = \min\{1,a\} + \min\{1,b\}$ we have $\bar{\epsilon}^n \leq \sum_\tau \bar{\epsilon}^n_\tau$ where $\bar{\epsilon}^n_\tau$ is given by (3.78) for $\tau \in \Big\{\{1\},...,\{1,...,K\}\Big\}$.

Following the next step in Section 3.4.1, by considering multiple classes for K-users, we find similar equations to (3.79) and (3.80). Now, instead of summing over i_ν and $\mathcal{A}^\nu_{i_\nu}(\gamma_{i_\nu})$, the summations over $\underline{\boldsymbol{u}}$ and $\hat{\boldsymbol{u}}_\tau$ in (3.79) and (3.80) are done as

$$\begin{aligned}\sum_{\underline{\boldsymbol{u}}} P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}) \sum_{\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{x}},\boldsymbol{y}) &= \sum_{\ell_1\in\mathcal{L}_1} \sum_{\boldsymbol{u}_1\in\mathcal{D}^{\ell_1}_1} P^n_{\boldsymbol{U}_1}(\boldsymbol{u}_1) ... \sum_{\ell_K\in\mathcal{L}_K} \sum_{\boldsymbol{u}_K\in\mathcal{D}^{\ell_K}_K} P^n_{\boldsymbol{U}_K}(\boldsymbol{u}_K) \\ &\times \sum_{\underline{\boldsymbol{x}},\boldsymbol{y}} Q^n_{1,\ell_1}(\boldsymbol{x}_1)...Q^n_{K,\ell_K}(\boldsymbol{x}_K) W^n(\boldsymbol{y}|\boldsymbol{x}_1,...,\boldsymbol{x}_K),\end{aligned} \tag{3.116}$$

and

$$\begin{aligned}&\sum_{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau : \frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau)W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)}\geq 1} Q^n_{\tau,\hat{\boldsymbol{u}}_\tau}(\hat{\boldsymbol{x}}_\tau) \leq \\ &\sum_{j_\tau\in\mathcal{L}_\tau} \sum_{\hat{\boldsymbol{u}}_\tau\in\mathcal{D}^{j_\tau}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau} Q^n_{\tau,j_\tau}(\hat{\boldsymbol{x}}_\tau) \left(\frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau)W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)}\right)^{s_{i_\tau j_\tau}}.\end{aligned} \tag{3.117}$$

Same reasoning given in Section 3.4.1, (3.81) for K users and multiple classes is derived as

$$\begin{aligned}&\min\left\{1, \sum_{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau} \sum_{\hat{\boldsymbol{x}}_\tau : \frac{P^n_{\boldsymbol{U}_\tau}(\hat{\boldsymbol{u}}_\tau)W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)}\geq 1} Q^n_{\tau,j_\tau}(\hat{\boldsymbol{x}}_\tau)\right\} \\ &\qquad\qquad \leq \sum_{j_\tau\in\mathcal{L}_\tau} \min_{\rho_{\ell_\tau j_\tau}\in[0,1]} \frac{G_{j_\tau}(s_{\ell_\tau j_\tau},\boldsymbol{x}_{\tau^c},\boldsymbol{y})^{\rho_{\ell_\tau j_\tau}}}{\Big(P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)\Big)^{s_{\ell_\tau j_\tau}\rho_{\ell_\tau j_\tau}}},\end{aligned} \tag{3.118}$$

where $G_{\ell_\tau}(\cdot)$ for K users and multiple classes is modified as

$$G_{l_\tau}(s, \boldsymbol{x}_{\tau^c}, \boldsymbol{y}) = \sum_{\boldsymbol{u}_\tau \in \mathcal{D}_\tau^{l_\tau}} \sum_{\boldsymbol{x}_\tau} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^s Q^n_{\tau,l_\tau}(\boldsymbol{x}_\tau) W^n(\boldsymbol{y}|\boldsymbol{x}_\tau, \boldsymbol{x}_{\tau^c})^s. \tag{3.119}$$

Next, we find a modified version of (3.83) for multiple classes where the summation over j_τ is done over $\mathcal{L}_\tau$ and instead of summing over i_1, i_2 we have summing over $\sum_{\nu=1}^{K} \sum_{\ell_\nu \in \mathcal{L}_\nu}$. By applying Hölder's inequality, we modify (3.85) for K users and multiple classes as

$$\bar{\epsilon}_\tau^n \leq \sum_{j_\tau, \ell_\tau \in \mathcal{L}_\tau} \min_{\rho_{\ell_\tau j_\tau} \in [0,1]} F^n_{\ell_\tau}\left(1 - s_{\ell_\tau j_\tau}\rho_{\ell_\tau j_\tau}, \frac{1}{p_{\ell_\tau j_\tau}}\right)^{p_{\ell_\tau j_\tau}} F^n_{j_\tau}\left(s_{\ell_\tau j_\tau}, \frac{\rho_{\ell_\tau j_\tau}}{1 - p_{\ell_\tau j_\tau}}\right)^{1-p_{\ell_\tau j_\tau}}, \tag{3.120}$$

where for K-user MAC with multiple classes, we have

$$F^n_{j_\tau}(a,b) = \sum_{\ell_{\tau^c} \in \mathcal{L}_{\tau^c}} \sum_{\boldsymbol{u}_{\tau^c} \in \mathcal{D}_{\tau^c}^{\ell_{\tau^c}}} \sum_{\boldsymbol{x}_{\tau^c}, \boldsymbol{y}} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) Q^n_{\tau^c, \ell_{\tau^c}}(\boldsymbol{x}_{\tau^c}) G_{j_\tau}(a, \boldsymbol{x}_{\tau^c}, \boldsymbol{y})^b. \tag{3.121}$$

Setting $s_{\ell_\tau j_\tau} = \frac{1}{1+\rho_{j_\tau}}$, $\rho_{\ell_\tau j_\tau} = \frac{\rho_{\ell_\tau}(1+\rho_{j_\tau})}{1+\rho_{\ell_\tau}}$ and $p_{\ell_\tau j_\tau} = \frac{1}{1+\rho_{\ell_\tau}}$, a modified version of (3.87) is derived where rather than summation over j_τ, i_τ, we have $j_\tau, \ell_\tau \in \mathcal{L}_\tau$. Applying arithmetic-geometric inequality, we find

$$\bar{\epsilon}_\tau^n \leq \sum_{\ell_\tau \in \mathcal{L}_\tau} \min_{\rho_{\ell_\tau}, \rho_{j_\tau} \in [0,1]} F^n_{\ell_\tau}\left(\frac{1}{1+\rho_{\ell_\tau}}, 1 + \rho_{\ell_\tau}\right) \sum_{j_\tau \in \mathcal{L}_\tau} \left(\frac{1}{1+\rho_{\ell_\tau}} + \frac{\rho_{j_\tau}}{1+\rho_{j_\tau}}\right). \tag{3.122}$$

Next, we bound $F^n_{\ell_\tau}(\cdot)$ defined in (3.121) by using similar steps as Lemma 3.1. By inserting $G_{\ell_\tau}\left(\frac{1}{1+\rho_{\ell_\tau}}, \cdot\right)$ defined in (3.119) into (3.121), we will find an expression similar to (3.91) where instead of summing over $i_{\tau^c} \in \{1,2\}$ and $\boldsymbol{u}_\nu \in \mathcal{A}_\nu^{i_\nu}(\gamma_\nu)$ for $\nu = \tau, \tau^c$, we have summation over $l_{\tau^c} \in \mathcal{L}_{\tau^c}$ and $\boldsymbol{u}_\nu \in \mathcal{D}_\nu^{\ell_\nu}$. Thus, (3.92) for K users and multiple classes is modified as

$$\begin{aligned} F^n_{\ell_\tau}\left(\frac{1}{1+\rho_{\ell_\tau}}, 1+\rho_{\ell_\tau}\right) &= \sum_{\ell_{\tau^c} \in \mathcal{L}_{\tau^c}} \sum_{\boldsymbol{u}_{\tau^c}} P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c}) \mathbb{1}\left(\boldsymbol{u}_{\tau^c} \in \mathcal{D}_{\tau^c}^{\ell_{\tau^c}}\right) \\ &\times \left(\sum_{\boldsymbol{u}_\tau} P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^{\frac{1}{1+\rho_{\ell_\tau}}} \mathbb{1}\left(\boldsymbol{u}_\tau \in \mathcal{D}_\tau^{\ell_\tau}\right)\right)^{1+\rho_{\ell_\tau}} e^{-E_0\left(\rho, Q^n_{\tau,\ell_\tau}, W^n Q^n_{\tau,\ell_{\tau^c}}\right)}. \end{aligned} \tag{3.123}$$

Using the fact $\mathbb{1}\Big\{a \leq b\Big\} \leq \left(\frac{b}{a}\right)^\lambda$ for $\lambda \geq 0$, we tighten the bound of indicator function as

$$\begin{aligned} &\mathbb{1}\left(\gamma^n_{\nu,\ell_\nu+1} < P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) \leq \gamma^n_{\nu,\ell_\nu}\right) \leq \\ &\qquad \min_{\lambda_{\nu,\ell_\nu+1}, \lambda_{\nu,\ell_\nu} \geq 0} \left(\frac{\gamma^n_{\nu,\ell_\nu}}{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu)}\right)^{\lambda_{\nu,\ell_\nu}} \left(\frac{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu)}{\gamma^n_{\nu,\ell_\nu+1}}\right)^{\lambda_{\nu,\ell_\nu+1}}, \end{aligned} \tag{3.124}$$

for $\nu=\tau,\tau^c$. Applying (3.124) into the all indicators function of (3.123) yields

$$F^n_{\ell_\tau}\left(\frac{1}{1+\rho_{\ell_\tau}},1+\rho_{\ell_\tau}\right)\leq\sum_{\ell_{\tau^c}\in\mathcal{L}_{\tau^c}}e^{-E_0\left(\rho_{\ell_\tau},Q^n_{\tau,\ell_\tau},W^nQ^n_{\tau,\ell_{\tau^c}}\right)}$$
$$\times\min_{\lambda_{\tau,\ell_\tau+1},\lambda_{\tau,\ell_\tau}\geq 0}\left(\sum_{\boldsymbol{u}_\tau}P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)^{\frac{1}{1+\rho_{\ell_\tau}}}\left(\frac{\gamma^n_{\nu,\ell_{\tau^c}}}{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu)}\right)^{\frac{\lambda_{\tau,\ell_\tau}}{1+\rho_{\ell_\tau}}}\left(\frac{P^n_{\boldsymbol{U}_\tau}(\boldsymbol{u}_\tau)}{\gamma^n_{\tau,\ell_\tau+1}}\right)^{\frac{\lambda_{\tau,\ell_\tau+1}}{1+\rho_{\ell_\tau}}}\right)^{1+\rho_{\ell_\tau}}$$
$$\times\min_{\lambda_{\tau^c,\ell_{\tau^c}+1},\lambda_{\tau^c,\ell_{\tau^c}}\geq 0}\sum_{\boldsymbol{u}_{\tau^c}}P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c})\left(\frac{\gamma^n_{\nu,\ell_{\tau^c}}}{P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu)}\right)^{\lambda_{\tau^c,\ell_{\tau^c}}}\left(\frac{P^n_{\boldsymbol{U}_{\tau^c}}(\boldsymbol{u}_{\tau^c})}{\gamma^n_{\tau^c,\ell_{\tau^c}+1}}\right)^{\lambda_{\tau^c,\ell_{\tau^c}+1}}. \tag{3.125}$$

Applying Lemma A.12 in Appendix A, into the second and third terms of (3.125) at arbitrary ρ and at $\rho=0$, respectively, we find that

$$\min_{\rho_{\ell_\tau}\in[0,1]}F^n_{\ell_\tau}\left(\frac{1}{1+\rho_{\ell_\tau}},1+\rho_{\ell_\tau}\right)\leq\sum_{\ell_{\tau^c}\in\mathcal{L}_{\tau^c}}\min_{\rho_{\ell_\tau}\in[0,1]}e^{-E_0\left(\rho_{\ell_\tau},Q^n_{\tau,\ell_\tau},W^nQ^n_{\tau,\ell_{\tau^c}}\right)}$$
$$\times e^{E_{s,\ell_\tau}(\rho,P^n_{\boldsymbol{U}_\tau},\gamma_{\tau,\ell_\tau+1},\gamma_{\tau,\ell_\tau})+E_{s,\ell_{\tau^c}}(0,P^n_{\boldsymbol{U}_{\tau^c}},\gamma_{\tau^c,\ell_{\tau^c}+1},\gamma_{\tau^c,\ell_{\tau^c}})}. \tag{3.126}$$

By inserting (3.126) into (3.122), bounding each term in the summations over ℓ_τ and ℓ_{τ^c} by the maximum terms, and using the fact that the sources and the channel are memoryless, we obtain

$$\bar{\epsilon}^n_\tau\leq\max_{\ell_\tau\in\mathcal{L}_\tau}\max_{\ell_{\tau^c}\in\mathcal{L}_{\tau^c}}\min_{\rho\in[0,1]}e^{-nE_0\left(\rho,Q_{\tau,\ell_\tau},WQ_{\tau,\ell_{\tau^c}}\right)}$$
$$\times e^{nE_{s,\ell_\tau}(\rho,P_{U_\tau},\gamma_{\tau,\ell_\tau+1},\gamma_{\tau,\ell_\tau})+nE_{s,\ell_{\tau^c}}(0,P_{U_{\tau^c}},\gamma_{\tau^c,\ell_{\tau^c}+1},\gamma_{\tau^c,\ell_{\tau^c}})}\times o(n), \tag{3.127}$$

where $o(n)$ is a sequence satisfying $\lim_{n\to\infty}\frac{\log o(n)}{n}=0$. Recalling that $\bar{\epsilon}^n\leq\sum_\tau\bar{\epsilon}^n_\tau$, by bounding the average error probability by the worst type of error τ, taking logarithm from both sides of it, using the properties given by (3.96), (3.97) and (3.98), we find that

$$\liminf_{n\to\infty}-\frac{1}{n}\log(\bar{\epsilon}^n)\geq\min_\tau\min_{\ell_\tau\in\mathcal{L}_\tau}\min_{\ell_{\tau^c}\in\mathcal{L}_{\tau^c}}\max_{\rho\in[0,1]}E_0\Big(\rho,Q_{\tau,\ell_\tau},WQ_{\tau^c,\ell_{\tau^c}}\Big)$$
$$-E_{s,\ell_\tau}(\rho,P_{U_\tau},\gamma_{\tau,\ell_\tau+1},\gamma_{\tau,\ell_\tau})-E_{s,\ell_{\tau^c}}(0,P_{U_{\tau^c}},\gamma_{\tau^c,\ell_{\tau^c}+1},\gamma_{\tau^c,\ell_{\tau^c}}). \tag{3.128}$$

Finally, we optimize (3.128) over $\gamma_{\nu,1},...,\gamma_{\nu,L_\nu}$ for $\nu=1,...,K$. This concludes the proof.

Chapter 4

The Multiple-Access Channel with Correlated Sources

As mentioned in Chapter 1, for the MAC with correlated sources, the obtained reliable transmission conditions in [12] are not optimal. In [34], by applying coding techniques, a new set of sufficient conditions were proposed. Moreover, in [35] new sufficient conditions for three-user MAC with correlated sources were studied.

Here, we study the MAC with correlated sources which is described in Section 1.1. After introducing the Gallager's source function for correlated sources, in Section 4.1, an achievable random-coding error exponent for joint source-channel coding over a multiple access channel with correlated sources is obtained in both primal and dual domains. The results are analyzed where either messages and codewords are statistically dependent or independent. From transmissible region, we find that considering statistical dependency between messages and codewords leads larger exponent.

In Section 4.2, we generalize the results to the constant-composition families. Like single-user communication, generally, by fixing the composition of the codewords, we attain more exponent. Parts of this chapter were presented in [36] and [37].

In this chapter, we frequently use (1.24). Thus, firstly we introduce the Gallager's source function for two correlated sources.

Lemma 4.1. *Consider two correlated sources characterized by $P_{\underline{U}}$. For source $\nu = 1, 2$, its outputs can be encoded into 2^{nR_ν} codewords such that the probability of ambiguous encoding P_e is bounded as*

$$-\frac{1}{n}\log P_e \geq \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \max_{\rho\in[0,1]} \rho R_\tau - E_{s,\tau}(\rho, P_{\underline{U}}), \tag{4.1}$$

where for $\tau = \{1, 2\}$, $R_\tau = R_1 + R_2$, and $E_{s,\tau}(\cdot)$ is given by (1.24).

Proof. See Section 4.3.1 □

4.1 Message-Dependent Random-Coding Exponent with Statistical Dependency

As shown in Chapter 2, tuning the random-coding ensemble leads to improved exponents in the point-to-point channel [7] and in the multiple-access channel [38, 39].

Inspired by these fact, we are motivated to consider joint source-channel coding where codewords are generated by a conditional probability distribution of the codeword symbol that depends both on the instantaneous source symbol and on the type of the source sequence. In particular, codebooks are drawn from a multi-letter distribution that is the product of independent conditional distributions that depend on the corresponding single-letter value of the source message.

For user $\nu = 1, 2$, we assign to source probability distribution P_{U_ν} a conditional probability distribution $\bar{Q}_{\nu,P_{U_\nu}}(x|u)$. We represent the set of these distributions by $\{\bar{Q}_{\nu,P_{U_\nu}} : P_{U_\nu} \in \mathcal{P}_{U_\nu}\}$. For every message $\boldsymbol{u}_\nu^n \in \mathcal{U}_\nu^n$, we randomly generate a codeword $\boldsymbol{x}_\nu(\boldsymbol{u}_\nu)$ according to the probability distribution $\bar{Q}^n_{\nu,\pi(\boldsymbol{u}_\nu)}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \prod_{i=1}^n \bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}(x_{\nu,i}|u_{\nu,i})$, where $\bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}$ is a probability distribution that depends on the type of $\boldsymbol{u}_\nu$, denoted by $\pi(\boldsymbol{u}_\nu)$.

Proposition 4.1. *For the two-user MAC with transition probability W, correlated sources $P_{\underline{U}}$ and the set of input distributions $\left\{\bar{Q}_{\nu,P_{U_\nu}}, P_{U_\nu} \in \mathcal{P}_{\mathcal{U}_\nu}\right\}$ for $\nu = 1, 2$, an achievable exponent E_1^{mds} is given by*

$$E_1^{\text{mds}} = \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W)$$
$$+\left[\min_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}_\tau(\hat{P}_{\underline{U}\underline{X}Y})} D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right]^+, \tag{4.2}$$

where

$$\mathcal{L}_\tau(\hat{P}_{\underline{U}\underline{X}Y}) \triangleq \Big\{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}} : \tilde{P}_{U_{\tau^c}X_{\tau^c}Y} = \hat{P}_{U_{\tau^c}X_{\tau^c}Y},$$
$$\mathbb{E}_{\tilde{P}}\lambda(\underline{U},\underline{X},Y) \geq \mathbb{E}_{\hat{P}}\lambda(\underline{U},\underline{X},Y)\Big\}, \tag{4.3}$$

and $\lambda(\underline{U},\underline{X},Y) = \log\left(P_{\underline{U}}(\underline{U})W(Y|\underline{X})\right)$, $[x]^+ = \max\{0, x\}$.

We briefly note that by setting $\hat{P}_{\underline{U}\underline{X}Y} = P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W$ and $\hat{P}_{\underline{U}\underline{X}Y} = \tilde{P}_{\underline{U}\underline{X}Y}$, the exponent in (4.2) can be shown to recover the achievable region obtained by Cover, El Gamal and Salehi [12].

Proof. See Section 4.3.2. □

Next, we show that the achievable exponent given by (4.2) is ensemble tightness. In other words, in Section 4.3.3, we prove that

$$\limsup_{n\to\infty} -\frac{1}{n}\bar{\epsilon} \leq E_1^{\mathrm{mds}}, \tag{4.4}$$

where $\bar{\epsilon}$ denotes the average error probability over the ensemble.

Now, to express the dual form of (4.2), like before we can apply the Lagrange duality theory. However, since the interpretation of the results are complicated, firstly we consider the case where the messages and codewords are statistically independent.

4.1.1 Statistically Independent Input Distributions

By applying the same approach in Section 4.3.2, the achievable exponent of (4.2) for statistically independent messages and codewords is simplified to

$$\begin{aligned} E_1^{\mathrm{md}} = & \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \min_{\hat{P}_{\underline{U}}\in\mathcal{P}_{\underline{\mathcal{U}}}} \min_{\hat{P}_{\underline{X}Y}\in\mathcal{P}_{\underline{\mathcal{X}}\times\mathcal{Y}}} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + D(\hat{P}_{\underline{X}Y}||Q_{1,\hat{P}_{U_1}}Q_{2,\hat{P}_{U_2}}W) \\ & + \left[\min_{\tilde{P}_{\underline{U}}\in\mathcal{K}_{s,\tau}(\hat{P}_{\underline{U}})} \min_{\tilde{P}_{\underline{X}Y}\in\mathcal{K}_{c,\tau}(\hat{P}_{\underline{X}Y})} D(\tilde{P}_{\underline{X}Y}||Q_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau|U_{\tau^c}}) \right]^+, \end{aligned} \tag{4.5}$$

where

$$\mathcal{K}_{s,\tau}(\hat{P}_{\underline{U}}) \triangleq \left\{ \tilde{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \tilde{P}_{U_{\tau^c}} = \hat{P}_{U_{\tau^c}},\ \mathbb{E}_{\tilde{P}}\log\left(P_{\underline{U}}(\underline{U})\right) \geq \mathbb{E}_{\hat{P}}\log\left(P_{\underline{U}}(\underline{U})\right) \right\}, \tag{4.6}$$

and

$$\begin{aligned} \mathcal{K}_{c,\tau}(\hat{P}_{\underline{X}Y}) \triangleq \Big\{ & \tilde{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}}\times\mathcal{Y}} : \tilde{P}_{X_{\tau^c}Y} = \hat{P}_{X_{\tau^c}Y}, \\ & \mathbb{E}_{\tilde{P}}\log\left(W(Y|\underline{X})\right) \geq \mathbb{E}_{\hat{P}}\log(W(Y|\underline{X})) \Big\}, \end{aligned} \tag{4.7}$$

and $[x]^+ = \max\{0, x\}$.

Next, by setting $T = U_\tau$, $Z = U_{\tau^c}$, $U = \underline{U}$, $X = X_\tau$, $W = Q_{\tau^c, \hat{P}_{U_{\tau^c}}} W$ and $Y = X_{\tau^c} Y$ in Lemma A.14 and then using the identity $\max\{0, a\} = \max_{\rho \in [0,1]} \rho a$, (4.5) is simplified as

$$E_1^{\mathrm{md}} \geq \min_{\tau} \min_{\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}} \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) + D(\hat{P}_{\underline{X}Y} || Q_{1, \hat{P}_{U_1}} Q_{2, \hat{P}_{U_2}} W)$$
$$+ \max_{\rho \in [0,1]} \rho D(\hat{P}_{\underline{X}Y} || Q_{\tau, \hat{P}_{U_\tau}} \hat{P}_{X_{\tau^c} Y}) - \rho H(\hat{P}_{U_\tau | U_{\tau^c}}). \quad (4.8)$$

To find the dual form of (4.8), we firstly analyze the source-exponent terms. Recalling from (3.4) and (3.5), in Chapter 3, for each user a fixed threshold was considered to partition the source-message set into two classes. Here, we use the same idea in the primal domain. To express the primal form of (3.4) and (3.5) for correlated sources, we recall that due to distributed source coding [9], the messages of each source are encoded independently from the other user. Considering this fact, in the following Lemma the asymptotic form of (3.4) and (3.5) for correlated sources is given.

Lemma 4.2. *Let $P_{\underline{U}}$ be the probability distribution of two correlated sources and for source $\nu = 1, 2$, P_{U_ν} be the marginal distribution of $P_{\underline{U}}$. Given $\gamma_\nu \in [0, 1]$ as the partitioning threshold, the set $\mathcal{P}_{\underline{\mathcal{U}}}$ can be partitioned into disjoint classes namely as $\mathcal{B}_\nu^1(\gamma_\nu)$ and $\mathcal{B}_\nu^2(\gamma_\nu)$ where*

$$\mathcal{B}_\nu^1(\gamma_\nu) = \left\{ \hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) \geq \log(\gamma_\nu) \right\}, \quad (4.9)$$

$$\mathcal{B}_\nu^2(\gamma_\nu) = \left\{ \hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) < \log(\gamma_\nu) \right\}. \quad (4.10)$$

Proof. See Appendix 4.3.4. □

Let $\nu \in \{1, 2\}$ and ν^c denotes the complement index of ν among the set $\{1, 2\}$. Roughly speaking, $\mathcal{B}_\nu^1(\gamma_\nu)$ in (4.9), can be interpreted as the asymptotic union of joint sequences $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ with joint-type $\hat{P}_{\underline{U}}^n$, where as long as the marginal probability $P_{U_\nu}^n(\boldsymbol{u}_\nu)$ is not less than the threshold γ_ν^n, the empirical distribution of $\boldsymbol{u}_{\nu^c}$ can be arbitrary (similarly for $\mathcal{B}_\nu^2(\gamma_\nu)$ in (4.10)). The following Proposition finds the Gallager source exponent function for the messages belonging to classes $\mathcal{B}_\nu^1(\gamma_\nu)$ and $\mathcal{B}_\nu^2(\gamma_\nu)$.

Proposition 4.2. *Let $\nu \in \{1, 2\}$, and ν^c be the complement index of ν among the set $\{1, 2\}$. For given $\gamma_\nu \in [0, 1]$ and $i_\nu \in \{1, 2\}$, in view of $\mathcal{B}_\nu^{i_\nu}(\gamma_\nu)$ given by (4.9) and (4.10), we have*

$$\min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1) \cap \mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau | U_{\tau^c}}) = -E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \quad (4.11)$$

where

$$E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}) =$$

$$\min_{\lambda_1\geq 0,\lambda_2\geq 0} \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_1}(u_1)}{\gamma_1}\right)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} \left(\frac{P_{U_2}(u_2)}{\gamma_2}\right)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}} \right)^{1+\rho}. \tag{4.12}$$

Proof. See Section 4.3.5. □

In fact, in (4.12), the objective function is a convex function with respect to λ_ν for $\nu = 1, 2$, and hence the optimal λ_ν minimizing (4.12) are the solution of an implicit equation which is obtained by setting the partial derivative of the objective function of (4.12) with respect to λ_ν equal to zero. To be precise, for the cases where both constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1)$ and $\hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)$ are active, λ_1 and λ_2 derived as the solution of the implicit equation, are greater than zero. Otherwise, the solution of the implicit equation is negative and the optimal λ_ν is zero.

Here, we compare the result given by (4.12), with independent sources. In Chapter 3, it has been shown that by partitioning the source into two classes, the obtained exponent is given in terms of $E_{s,i_\nu}(\cdot)$ function where $i_\nu = 1, 2$.

Additionally, from Lemma A.16, for the source $\nu = 1, 2$, with probability distribution P_{U_ν}, the partitioning threshold γ_ν, and $i_\nu = 1, 2$, we have

$$E_{s,i_\nu}(\rho, P_{U_\nu}, \gamma_\nu) = \min_{\lambda_\nu \geq 0} \log \sum_{u_\nu} P_{U_\nu}(u_\nu)^{\frac{1}{1+\rho}} \left(\frac{P_{U_\nu}(u_\nu)}{\gamma_\nu}\right)^{-\frac{(-1)^{i_\nu}\lambda_\nu}{1+\rho}}. \tag{4.13}$$

For independent sources, by applying $P_{\underline{U}}(\underline{u}) = P_{U_1}(u_1)P_{U_2}(u_2)$ in (4.12), and in view of (4.13), the function $E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma})$ is simplified as

$$E_{s,\tau,i_1,i_2}(\rho, P_{U_1}(u_1)P_{U_2}(u_2), \underline{\gamma}) = E_{s,i_\tau}(\rho, P_{U_\tau}, \gamma_\tau) + E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c}), \tag{4.14}$$

where as discussed in (3.7), for $\tau = \{1, 2\}$, $E_{s,i_\tau}(\rho, P_{\underline{U}}, \underline{\gamma}) = E_{s,i_1}(\rho, P_{U_1}, \gamma_1) + E_{s,i_2}(\rho, P_{U_2}, \gamma_2)$. In fact, depending on the tangent points given in (2.25), $E_{s,\{1,2\},i_1,i_2}(\cdot)$ as a function of ρ is either $E_s(\rho, P_{U_\nu}) + E_s(\rho, P_{U_{\nu^c}})$ or $E_s(\rho, P_{U_\nu}) + E_{s,i_{\nu^c}}(\rho, P_{U_{\nu^c}}, \gamma_{\nu^c})$ where ν can be 1 or 2, and ν^c denotes the complement index of ν among the set $\{1, 2\}$.

For error type $\tau \in \{\{1\}, \{2\}\}$ and for the four combinations of $i_1, i_2 \in \{1, 2\}$, Figure 4.1 shows (4.14) for two independent sources with given γ_1, γ_2. As shown in (4.14) and for Figure 4.1, the functions $E_{s,\tau,1,1}(\cdot)$ and

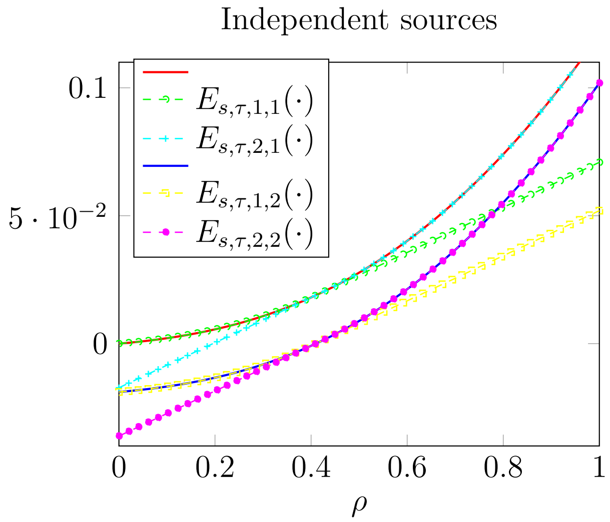

Figure 4.1: The function $E_{s,\tau,i_1,i_2}(\cdot)$ given by (4.14) for two independent sources $P_{U_1}(u_1)P_{U_2}(u_2)$ versus ρ, for the fixed γ_1 and γ_2 where $i_1, i_2 = 1, 2$. For error type $\tau \in \{\{1\}, \{2\}\}$, the solid red and blue curves are respectively $E_s(\rho, P_{U_\tau})$ and $E_s(\rho, P_{U_\tau}) + E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$.

$E_{s,\tau,2,1}(\cdot)$ follow $E_s(\rho, P_{U_\tau})$ given by (1.9), for an interval of ρ, while they are the straight line tangent to Gallager's source function beyond that interval. However, the functions $E_{s,\tau,1,2}(\cdot)$ and $E_{s,\tau,2,2}(\cdot)$ are either the Gallager's source function shifted by $E_{s,i_{\tau^c}}(0, P_{U_{\tau^c}}, \gamma_{\tau^c})$ or the straight line tangent to it.

On the other hand, for correlated sources with four combinations of $i_1, i_2 \in \{1, 2\}$, Figure 4.2 shows (4.12) for two correlated sources with given γ_1, γ_2 and error type τ. It can be seen that for the example of Figure 4.2, the functions $E_{s,\tau,1,1}(\cdot)$ and $E_{s,\tau,2,1}(\cdot)$ are the generalized Gallager's source function (1.24) for an interval of ρ, while they are a curve tangent to $E_{s,\tau}(\cdot)$ beyond that interval. Thus, unlike the independent sources, instead of a straight line tangent to Gallager's source function, for correlated sources, a curve is tangent to $E_{s,\tau}(\cdot)$. The reason for this is explained in the following.

In Figure 4.2, consider $E_{s,\tau,2,1}(\cdot)$ where $i_1 = 2$ and $i_2 = 1$. For the region of ρ where $E_{s,\tau,2,1}(\cdot)$ equals to $E_{s,\tau}(\cdot)$, both constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^2(\gamma_1)$ and $\hat{P}_{\underline{U}} \in \mathcal{B}_2^1(\gamma_2)$ are inactive, while for the region of ρ where $E_{s,\tau,2,1}(\cdot)$ equals to the curve tangent to $E_{s,\tau}(\cdot)$, only one of the constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^2(\gamma_1)$ or $\hat{P}_{\underline{U}} \in \mathcal{B}_2^1(\gamma_2)$ is active (similarly for $E_{s,\tau,1,1}(\cdot)$). For given i_1, i_2, let $\nu \in \{1, 2\}$ correspond to the active constraint. For example, in Figure 4.2, for the region of ρ where $E_{s,\tau,2,1}(\cdot)$ equals to the tangent curve, only the constraint

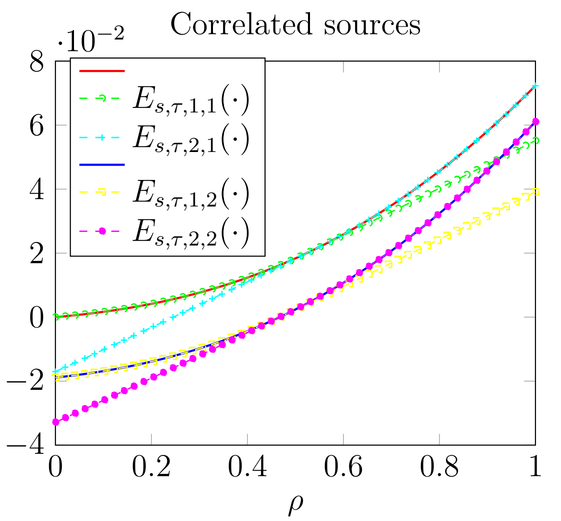

Figure 4.2: The function $E_{s,\tau,i_1,i_2}(\cdot)$ given by (4.12) for two correlated sources versus ρ, for the fixed γ_1 and γ_2 where $i_1, i_2 = 1, 2$. The solid red and blue curves are respectively given by (1.24) and (4.16).

$\hat{P}_{\underline{U}} \in \mathcal{B}_{\nu}^{i_\nu}(\gamma_\nu)$ is active. Then, the primal form of the curve is

$$-\min_{\substack{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \\ \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) = \log(\gamma_\nu)}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau | U_{\tau^c}}), \tag{4.15}$$

that corresponds to the Gallager's source exponent function of messages source ν whose empirical distributions are fixed, i. e. the set $\Big\{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) = \log(\gamma_\nu)\Big\}$.

We note that (4.15), describes the situation that only the type class of one of the sources is fixed. Thus, we have more freedom in the source-type class of another source. This implies that for correlated sources the joint-type class is not fixed, but rather contains the union of joint-type classes whose type class of one of the sources is fixed. Thus, unlike the independent sources, for correlated sources (4.15) is a curve rather than a straight line.

Coming back to Figure 4.2, for an for an interval of ρ, the functions $E_{s,\tau,1,2}(\cdot)$ ($E_{s,\tau,2,2}(\cdot)$) is

$$\min_{\lambda_\nu \geq 0} \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_\nu}(u_\nu)}{\gamma_\nu} \right)^{-\frac{(-1)^{i_\nu} \lambda_\nu}{1+\rho}} \right)^{1+\rho}, \tag{4.16}$$

where $\nu \in \{1, 2\}$ corresponds to the fact that only the constraint $\hat{P}_{\underline{U}} \in \mathcal{B}_{\nu}^{i_\nu}(\gamma_\nu)$ is active. In addition, beyond that interval of ρ, the functions

$E_{s,\tau,1,2}(\cdot)$ $(E_{s,\tau,2,2}(\cdot))$ is (4.12) where both constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1)$ and $\hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)$ are active.

Next, considering (4.8), we note that the optimization problem over $\hat{P}_{\underline{X},Y}$ in (4.8) is coupled with the minimization problem over $\hat{P}_{\underline{U}}$ through $Q_{\nu,\hat{P}_{U_\nu}}$ for $\nu = 1, 2$. In view of classes defined by (4.9) and (4.10), we express the dependency of the input distribution $Q_{\nu,\hat{P}_{U_\nu}}$ on $\hat{P}_{U_\nu}$, through the class index. In other words, for $\hat{P}_{U_\nu} \in \mathcal{B}_\nu^1(\gamma_\nu)$, we let $Q_{\nu,\hat{P}_{U_\nu}} = Q_{\nu,1}$ and similarly for $\hat{P}_{U_\nu} \in \mathcal{B}_\nu^2(\gamma_\nu)$, we let $Q_{\nu,\hat{P}_{U_\nu}} = Q_{\nu,2}$. Applying this to (4.8), and splitting the minimization over $\hat{P}_{\underline{U}}$ into minimization over disjoint classes as $\min_{i_1,i_2=1,2} \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)}$, we find that

$$
\begin{aligned}
E_1^{\mathrm{md}} \geq \min_{\tau} \min_{i_1,i_2=1,2} & \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)} \min_{\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) \\
&+ D(\hat{P}_{\underline{X}Y} || Q_{1,i_1} Q_{2,i_2} W) + \max_{\rho \in [0,1]} \rho D(\hat{P}_{\underline{X}Y} || Q_{\tau,i_\tau} \hat{P}_{X_{\tau^c} Y}) - \rho H(\hat{P}_{U_\tau | U_{\tau^c}}).
\end{aligned}
\tag{4.17}
$$

By using the min-max inequality, we swap the maximization over ρ with the minimizations over $\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}$ and $\hat{P}_{\underline{U}}$ in (4.17), i. e. $E_1^{\mathrm{md}} \geq E^{\mathrm{md}}$ where E^{md} is given by

$$
\begin{aligned}
E^{\mathrm{md}} = \min_{i_1,i_2=1,2} \min_{\tau} \max_{\rho \in [0,1]} & \min_{\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}} D(\hat{P}_{\underline{X}Y} || Q_{1,i_1} Q_{2,i_2} W) \\
&+ \rho D(\hat{P}_{\underline{X}Y} || Q_{\tau,i_\tau} \hat{P}_{X_{\tau^c} Y}) + \min_{\substack{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \\ \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1) \cap \mathcal{B}_2^{i_2}(\gamma_2)}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau | U_{\tau^c}}).
\end{aligned}
\tag{4.18}
$$

In (4.18), the inner minimization problems over $\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}$ and $\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}$, respectively lead to the channel and source exponent functions. The minimization over $\hat{P}_{\underline{U}}$ is discussed in Proposition 4.2, while to find channel exponent function, we use Lemma A.3. By setting $\hat{P}_{XY} = \hat{P}_{\underline{X}Y}$ and $Q = Q_{\tau,i_\tau}$ in Lemma A.3, the minimization over $\hat{P}_{\underline{X}Y}$ in (4.18), is optimized as

$$
\begin{aligned}
\min_{\hat{P}_{\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{X}} \times \mathcal{Y}}} & D(\hat{P}_{\underline{X}Y} || Q_{1,i_1} Q_{2,i_2} W) + \rho D(\hat{P}_{\underline{X}Y} || Q_{\tau,i_\tau} \hat{P}_{X_{\tau^c} Y}) \\
&= E_0(\rho, Q_{\tau,i_\tau}, W Q_{\tau^c, i_\tau^c}),
\end{aligned}
\tag{4.19}
$$

where $E_0(\cdot)$ is given by (1.14).

Now, putting back the results obtained in equations (4.19) and (4.11) into the respective minimization problems over $\hat{P}_{\underline{X}Y}$ and $\hat{P}_{\underline{U}}$ of (4.18), and defining

$$f_{i_1,i_2}(\gamma_1,\gamma_2) = \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau}^c}) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \tag{4.20}$$

an alternative expression for (4.18) is derived as

$$E^{\mathrm{md}} = \max_{\gamma_1,\gamma_2\in[0,1]} \min_{i_1,i_2=1,2} f_{i_1,i_2}(\gamma_1,\gamma_2), \tag{4.21}$$

where in (4.21), we optimized the exponent over γ_ν for $\nu = 1, 2$. We recall that since two source-message classes namely $\mathcal{B}_\nu^1(\gamma_\nu)$, $\mathcal{B}_\nu^2(\gamma_\nu)$ and two input distributions $Q_{\nu,1}, Q_{\nu,2}$ are considered for each user $\nu = 1, 2$, there are four possible assignments where in (4.21) the optimal assignment of input distributions is considered.

In Section 4.3.6, we show that for $\nu = 1, 2$, the function

$$\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau}^c}) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \tag{4.22}$$

is non-decreasing with respect to γ_ν when $i_\nu = 1$ and is non-increasing with respect to γ_ν when $i_\nu = 2$. Considering this fact, to find the optimal $\underline{\gamma}$ maximizing (4.21), we can use the same approach proposed in Proposition 3.2. In other words, the optimal γ_1 and γ_2 are the points where the minimum of all non-decreasing functions with respect to γ_ν is equal with the minimum of all non-increasing functions.

Proposition 4.3. *The optimal $\gamma_1^\star$ and $\gamma_2^\star$ maximizing (4.21) satisfy*

$$\begin{cases} \min\limits_{i_2=1,2} f_{1,i_2}(\gamma_1^\star,\gamma_2^\star) = \min\limits_{i_2=1,2} f_{2,i_2}(\gamma_1^\star,\gamma_2^\star), \\ \min\limits_{i_1=1,2} f_{i_1,1}(\gamma_1^\star,\gamma_2^\star) = \min\limits_{i_1=1,2} f_{i_1,2}(\gamma_1^\star,\gamma_2^\star). \end{cases} \tag{4.23}$$

When (4.23) has no solutions, then $\gamma_\nu^\star \in \{0,1\}$. In particular, if $f_{1,i_2}(0,\gamma_2) > f_{2,i_2}(0,\gamma_2)$ then $\gamma_1^\star = 0$, otherwise $\gamma_1^\star = 1$; and if $f_{i_1,1}(\gamma_1,0) > f_{i_1,2}(\gamma_1,0)$, we have $\gamma_2^\star = 0$, otherwise $\gamma_2^\star = 1$.

Proof. See Section 4.3.6. □

4.1.1.1 iid Random-Coding Exponent

Next, we show that the achievable exponent given by (4.21), is greater than iid random-coding exponent. We recall that for iid ensemble, for each user, only one input distribution generates codewords. For a two-user MAC with two correlated source $P_{\underline{U}}$, transition probability W and given input distributions Q_1 and Q_2, the i.i.d random-coding exponent is given by

$$E^{\text{iid}} = \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \max_{\rho\in[0,1]} E_0(\rho, Q_\tau, WQ_{\tau^c}) - E_{s,\tau}(\rho, P_{\underline{U}}), \tag{4.24}$$

where $E_{s,\tau}(\cdot)$ and $E_0(\cdot)$ are respectively given by (1.24) and (1.14). To prove (4.24), we recall that iid ensemble is a special case of message-dependent ensemble where for each user, only one class is considered. Assume that all the messages of user ν are generated according to $Q_\nu = Q_{\nu,1}$ for $\nu = 1, 2$. Thus, all the messages belong to the first class, i. e. $\gamma_1 = \gamma_2 = 0$ and $i_1 = i_2 = 1$, and hence

$$E_{s,\tau,1,1}(\rho, P_{\underline{U}}, 0, 0) = E_{s,\tau}(\rho, P_U). \tag{4.25}$$

Applying Q_1 and Q_2 in (4.19) as input distributions, by considering (4.25), in view of (4.20), the exponent of (4.21) is simplified as (4.24).

Proposition 4.4. *The achievable exponent given by* (4.21) *is greater than that achieved using only one input distribution for each user, i. e.*

$$E^{\text{md}} \geq \max_{i_1\in\{1,2\}} \max_{i_2\in\{1,2\}} \min_{\tau} F^{\text{L}}_{\tau,i_\tau,i_{\tau^c}}, \tag{4.26}$$

where

$$F^{\text{L}}_{\tau,i_\tau,i_{\tau^c}} = \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_{\tau^c}}) - E_{s,\tau}(\rho, P_{\underline{U}}). \tag{4.27}$$

Like (3.50), the lower bound in (4.26) selects the best iid random-coding exponent among the all four combinations of input distributions through i_1 and i_2.

Proof. See Section 4.3.7. □

4.1.1.2 Numerical Example

In this section, we develop an example showing that using two input distributions for each user, attains larger achievable exponent than the case where each user uses one input distribution. We consider two correlated discrete memoryless sources with alphabet $\mathcal{U}_\nu = \{1, 2\}$ for $\nu = 1, 2$ where

$$P_{\underline{U}} = \begin{pmatrix} 0.0005 & 0.0095 \\ 0.0005 & 0.9895 \end{pmatrix}. \tag{4.28}$$

We also consider a discrete memoryless MAC, given by (3.54). As mentioned before, we observe that W is a 36×4 matrix where the transition

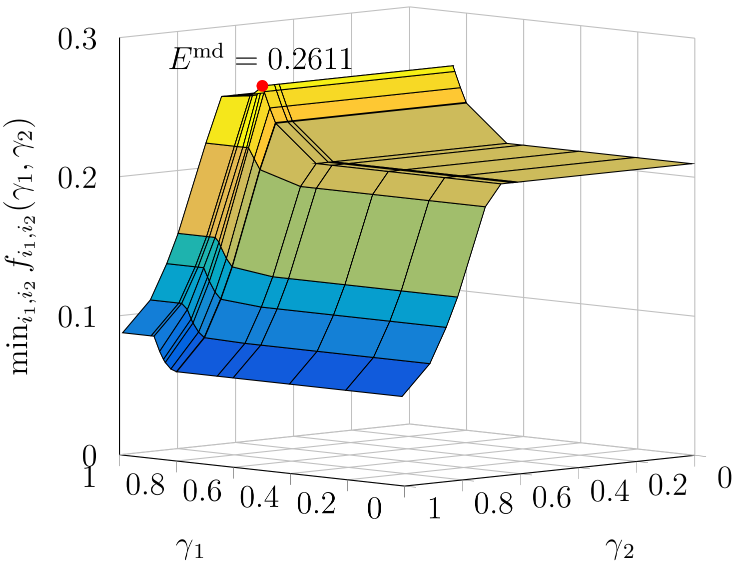

Figure 4.3: $\min_{i_1,i_2} f_{i_1,i_2}(\gamma_1, \gamma_2)$ with respect to γ_1 and γ_2.

probability $W(y|x_1, x_2)$ is located at row $x_1 + 6(x_2 - 1)$ of matrix W, for $(x_1, x_2) \in \{1, 2, ..., 6\} \times \{1, 2, ..., 6\}$. Recalling that each source has two classes and that four input distributions generate codewords, there are four possible assignments of input distributions to classes. Among all possible permutations, we select the one that gives the highest exponent. Here, for user $\nu = 1, 2$, we consider the set of input distributions $\Big\{[0\ 0\ 0\ 0\ 0.5\ 0.5], [0.25\ 0.25\ 0.25\ 0.25\ 0\ 0]\Big\}$. For the channel given in (3.54), the optimal assignment is

$$Q_{\nu,1} = [0\ \ 0\ \ 0\ \ 0\ \ 0.5\ \ 0.5], \tag{4.29}$$
$$Q_{\nu,2} = [0.25\ \ 0.25\ \ 0.25\ \ 0.25\ \ 0\ \ 0], \tag{4.30}$$

for both $\nu = 1, 2$.

For this example, from (4.23), we numerically compute the optimal $\gamma_1^\star$ and $\gamma_2^\star$ maximizing (4.21) leading to $\gamma_1^\star = 0.8469$ and $\gamma_2^\star = 0.6581$. The message-dependent exponent is derived as $E^{\text{md}} = 0.2611$, while iid exponent for the best assignment is derived as 0.2503. Figure 4.3 shows $\min_{i_1,i_2} f_{i_1,i_2}(\underline{\gamma})$ with respect to γ_1 and γ_2. It can be seen that the maximum of $\min_{i_1,i_2} f_{i_1,i_2}(\underline{\gamma})$ is derived at $(0.8469, 0.6581)$; however, the lower bound is obtained at $(1, 0)$.

4.1.1.3 On the Error Type $\tau \in \{\{1\}, \{2\}\}$

In this section, we only focus on the error type $\tau = \{1\}$ or $\tau = \{2\}$. Since for these two error types, the messages of user τ and τ^c are respectively

Table 4.1: Values of (4.22) with optimal thresholds $\gamma_1^\star = 0.8469$ $\gamma_2^\star = 0.6581$, for types of error τ, and user classes i_τ and i_{τ^c}.

	(i_1, i_2)			
	(1,1)	(1,2)	(2,1)	(2,2)
$\tau = \{1\}$	0.3172	0.2735	0.3120	0.2611
$\tau = \{2\}$	0.3986	0.4372	0.2611	0.4119
$\tau = \{1,2\}$	0.2611	0.2972	0.2630	0.2883

Table 4.2: Values of $F^{\mathrm{L}}_{\tau,i_\tau,i_{\tau^c}}$ in (4.27) for types of error τ, and input distribution Q_{1,i_1}, Q_{2,i_2}.

	$Q_{1,1}$,$Q_{2,1}$	$Q_{1,1}$,$Q_{2,2}$	$Q_{1,2}$,$Q_{2,1}$	$Q_{1,2}$,$Q_{2,2}$
$\tau = \{1\}$	0.2682	0.0642	0.3120	0.0879
$\tau = \{2\}$	0.3986	0.3986	0.2503	0.3696
$\tau = \{1,2\}$	0.2097	0.2097	0.2630	0.2360

decoded incorrectly and correctly, one can conclude that for $\tau = \{1\}, \{2\}$, the message of user τ^c is known at the receiver. To be precise, consider Figure 4.4, where the source is characterized by a probability distribution P_{U_τ} on the source alphabet $\mathcal{U}_\tau$. The source message $\boldsymbol{u}_\tau$ with length n is mapped onto codeword $\boldsymbol{x}_\tau(\boldsymbol{u}_\tau)$ which also has length n and is drawn from the codebook $\mathcal{C}^\tau = \{\boldsymbol{x}_\tau(\boldsymbol{u}_\tau), \boldsymbol{u}_\tau \in \mathcal{U}_\tau^n\}$. In addition, the channel state is characterized by a probability distribution $P_{U_{\tau^c}|U_\tau}$.

Like before, we partition the source messages into two classes and we assign two input distributions. In view of (3.4) and (3.5), for $\boldsymbol{u}_\tau \in \mathcal{A}_\tau^1(\gamma_\tau)$, the codewords are generated according to $Q_{\tau,1}$, while for $\boldsymbol{u}_\tau \in \mathcal{A}_\tau^2(\gamma_\tau)$, the input distribution $Q_{\tau,2}$ is applied to generate codewords.

Similarly, we partition the outputs of the channel state $P_{U_{\tau^c}|U_\tau}$ into two classes. According to the class of $\boldsymbol{u}_{\tau^c} \in \mathcal{A}_{\tau^c}^1(\gamma_{\tau^c})$, the encoder sends the codewords over the discrete memoryless channel with transition probability $WQ_{\tau^c,1}$ with input alphabet $\mathcal{X}_\tau$ and output alphabet $\mathcal{Y} \times \mathcal{X}_\tau$ and for $\boldsymbol{u}_{\tau^c} \in \mathcal{A}_{\tau^c}^2(\gamma_{\tau^c})$, the channel $WQ_{\tau^c,2}$ is utilized to transmit the codewords.

For the model described in this section, by using random-coding union bound, for $\tau = \{1\}$ or $\tau = \{2\}$, the achievable exponent is obtained as

$$\max_{\gamma_{\tau^c},\gamma_\tau \in [0,1]} \min_{i_\tau, i_{\tau^c}=1,2} \max_{\rho \in [0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_\tau^c}) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \tag{4.31}$$

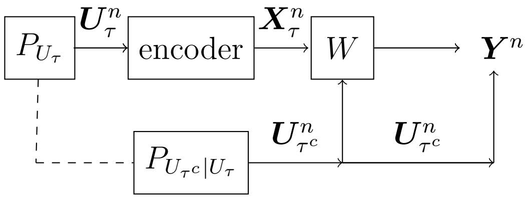

Figure 4.4: Single user transmission in the presence of channel state.

where $E_{s,\tau,i_1,i_2}(\cdot)$ is given by (4.12). We note that since $\tau = \{1\}, \{2\}$ is given, if the minimization over τ in (4.21) is removed, (4.31) will equal to (4.21).

In Section 4.3.8, we show that (4.31) equals to

$$\max\Bigg\{\max_{\gamma_\tau\in[0,1]}\min_{i_\tau=1,2}\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,1}) - E_{s,\tau,i_\tau}(\rho, P_{\underline{U}}, \gamma_\tau),$$
$$\max_{\gamma_\tau\in[0,1]}\min_{i_\tau=1,2}\max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,2}) - E_{s,\tau,i_\tau}(\rho, P_{\underline{U}}, \gamma_\tau)\Bigg\}, \tag{4.32}$$

where

$$E_{s,\tau,i_\tau}(\rho, P_{\underline{U}}, \gamma_\tau) = \min_{\lambda_\tau\geq 0}\log\sum_{u_{\tau^c}}\left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}}\left(\frac{P_{U_\tau}(u_\tau)}{\gamma_\tau}\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}}\right)^{1+\rho}. \tag{4.33}$$

The first term of (4.32) can be interpreted as the message-dependent exponent of channel $WQ_{\tau^c,1}$ where the source P_{U_τ} is correlated to the channel state. Similarly, the second term (4.32) has the similar meaning. Due to (4.32), we can interpret that the final exponent is the exponent of the better channel $WQ_{\tau^c,1}$ or $WQ_{\tau^c,2}$. Applying $P_{\underline{U}} = P_{U_1}P_{U_2}$, we find (3.53) as a special case where the source and channel sate are independent.

4.1.2 Statistical Dependency between Messages and Codewords

Now, we can find the dual form of (4.2). Firstly, by setting $U = U_\tau$, $X = X_\tau$, $W = P_{U_{\tau^c}}Q_{\tau^c,\hat{P}_{U_{\tau^c}}}W$ and $Y = U_{\tau^c}X_{\tau^c}Y$ in Lemma A.1, (4.2) can be bounded as the minimization over only $\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}$. Then, by considering the fact that $P_Z(z) = \sum_t P_{ZT}(z,t)$, (4.9) and (4.10) can be rewritten as

$$\mathcal{B}_\nu^1(\gamma_\nu) = \left\{ \hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}}\underline{\mathcal{X}}\underline{\mathcal{Y}}} : \sum_{\underline{u},\underline{x},y} \hat{P}_{\underline{U}\underline{X}Y}(\underline{u},\underline{x},y) \log P_{U_\nu}(u_\nu) \geq \log(\gamma_\nu) \right\}, \tag{4.34}$$

$$\mathcal{B}_\nu^2(\gamma_\nu) = \left\{ \hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}}\underline{\mathcal{X}}\underline{\mathcal{Y}}} : \sum_{\underline{u},\underline{x},y} \hat{P}_{\underline{U}\underline{X}Y}(\underline{u},\underline{x},y) \log P_{U_\nu}(u_\nu) < \log(\gamma_\nu) \right\}. \tag{4.35}$$

Next, we split the minimization over $\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}}\underline{\mathcal{X}}\mathcal{Y}}$ into minimization over disjoint classes as $\min_{i_1,i_2=1,2} \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)}$, where $\mathcal{B}_\nu^1(\gamma_\nu)$ and $\mathcal{B}_\nu^2(\gamma_\nu)$ are given by (4.34) and (4.35). Finally, by following the same approach presented in Section 4.1.1, the dual form of (4.2) can be written as $E_1^{\text{mds}} \geq E^{\text{mds}}$ where

$$E^{\text{mds}} = \max_{\gamma_1,\gamma_2 \in [0,1]} \min_{i_1,i_2=1,2} \min_{\tau} \max_{\rho \in [0,1]} E_{\tau,i_1,i_2}^{\text{mds}}(\rho, \underline{\gamma}), \tag{4.36}$$

and

$$E_{\tau,i_1,i_2}^{\text{mds}}(\rho,\underline{\gamma}) = \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} -\log \Bigg(\sum_{\substack{y \\ u_{\tau^c}, x_{\tau^c}}} \Bigg(\sum_{u_\tau, x_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \Big(\frac{P_{U_1}(u_1)}{\gamma_1} \Big)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}}$$
$$\times \Big(\frac{P_{U_2}(u_2)}{\gamma_2} \Big)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}} \bar{Q}_{\tau,i_\tau}(x_\tau|u_\tau) \bar{Q}_{\tau^c,i_{\tau^c}}(x_{\tau^c}|u_{\tau^c})^{\frac{1}{1+\rho}} W(y|\underline{x})^{\frac{1}{1+\rho}} \Bigg)^{1+\rho} \Bigg). \tag{4.37}$$

By setting the partial derivative of the objective function of (4.37) with respect to λ_ν, equal to zero, an implicit equation is derived to give λ_ν maximizing (4.37). In fact, if the constraints given by (4.34) and (4.35) are active, the optimal $\lambda_\nu^\star \geq 0$ maximizing (4.37) is obtained as the solution of the implicit equation. Otherwise, $\lambda_\nu^\star = 0$.

Applying the same approach given in the proof of Lemma 4.3, we easily find that for $\nu = 1,2$, $E_{\tau,i_1,i_2}^{\text{mds}}(\cdot)$ is non-decreasing (non-increasing) with respect to γ_ν when $i_\nu = 1$ ($i_\nu = 2$). Thus, like before in view of Lemma A.8, the optimal thresholds obtained at point that the minimum of non-decreasing functions with respect to γ_ν is equal to the minimum of non-increasing functions, i. e. the the optimal γ_ν maximizing (4.45) is obtained at the points where

$$\min_{\tau} \max_{\rho \in [0,1]} E_{\tau,i_1,i_2}^{\text{mds}}(\rho,\underline{\gamma}) \Big|_{i_\nu=1} = \min_{\tau} \max_{\rho \in [0,1]} E_{\tau,i_1,i_2}^{\text{mds}}(\rho,\underline{\gamma}) \Big|_{i_\nu=2}, \tag{4.38}$$

and when (4.38) has no solution, the optimal γ_ν is either zero or one.

Comparing (4.37) with the objective function of (4.20), we find that as long as the messages and codewords are statistically independent

$$E^{\mathrm{mds}}_{\tau,i_1,i_2}(\rho,\underline{\gamma}) = E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i^c_\tau}) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \tag{4.39}$$

where $E^{\mathrm{mds}}_{\tau,i_1,i_2}(\cdot)$, $E_0(\cdot)$ and $E_{s,\tau,i_1,i_2}(\cdot)$ are respectively given by (4.37), (1.14) and (4.12). However, recalling from (1.35), unlike MAC with independent sources, for the MAC with correlated sources, since through $P_{\underline{U}}$ the input distributions of both users depend on each other, the statistical dependency between messages and codewords may affect error exponent.

4.2 Studying Generalized Conditional Constant-Composition Ensemble

In this section, in view of Section 2.1.3, we consider the ensemble defined as follows. For user $\nu = 1, 2$, we assign to every source probability distribution P_{U_ν} a conditional probability distribution $\bar{Q}_{\nu,P_{U_\nu}}(x|u)$. For a given message $\boldsymbol{u}_\nu = (u_{\nu,1}, u_{\nu,2}, ..., u_{\nu,n})$, we consider the sub-sequences of $\boldsymbol{u}$ which have the same symbols. We define $j_{u_\nu}(\boldsymbol{u}_\nu)$ as the set of all positions where the symbol u_ν appears in $\boldsymbol{u}_\nu$, i.e. for all $u_\nu \in \mathcal{U}_\nu$

$$j_{u_\nu}(\boldsymbol{u}_\nu) = \{i \in \{1, 2, ..., n\},\ \text{such that } u_{\nu,i} = u_\nu\}. \tag{4.40}$$

The subsequence can be represented by $\boldsymbol{u}_\nu(j_{u_\nu}(\boldsymbol{u}_\nu))$.

Let $\bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}(x_\nu|u_\nu)$ be a conditional distribution that depends on the type of $\boldsymbol{u}_\nu$, $\pi(\boldsymbol{u}_\nu)$. We approximate the conditional distribution $\bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}$ with a type-p conditional distribution $\bar{Q}_{\nu,p,\pi(\boldsymbol{u}_\nu)}$ that satisfies

$$\bar{Q}_{\nu,p,\pi(\boldsymbol{u}_\nu)}(x_\nu|u_\nu) \in \left\{0, \frac{1}{p}, \frac{2}{p}, ..., 1\right\}, \tag{4.41}$$

for all $x_\nu \in \mathcal{X}_\nu$ and $u_\nu \in \mathcal{U}_\nu$. We set p depends on u_ν and $\boldsymbol{u}_\nu$, $p = |j_{u_\nu}(\boldsymbol{u}_\nu)|$ and choose this distribution such that the variational distance between $\bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}$ and $\bar{Q}_{\nu,p,\pi(\boldsymbol{u}_\nu)}$ satisfies

$$\left|\bar{Q}_{\nu,p,\pi(\boldsymbol{u}_\nu)}(x_\nu|u_\nu) - \bar{Q}_{\nu,\pi(\boldsymbol{u}_\nu)}(x_\nu|u_\nu)\right| < \frac{1}{p}. \tag{4.42}$$

For every $u_\nu \in \mathcal{U}_\nu$, we randomly pick a sequence $\boldsymbol{x}_{\nu,u_\nu}$ of length $|j_{u_\nu}(\boldsymbol{u}_\nu)|$ from the set $\mathcal{T}^p_{u_\nu}(\bar{Q}_{\nu,p_{u_\nu},\pi(\boldsymbol{u}_\nu)})$ and set $\boldsymbol{x}_\nu(j_{u_\nu}(\boldsymbol{u})) = \boldsymbol{x}_{\nu,u_\nu}$. We apply constant composition random coding with the set of distributions $\left\{\bar{Q}_{\nu,p_{u_\nu},\pi(\boldsymbol{u}_\nu)}(x_\nu|u_\nu)\right\}_{u_\nu\in\mathcal{U}_\nu}$ to determine an achievable exponent.

Proposition 4.5. *For generalized conditional constant-composition random coding over a two-user MAC with two correlated sources, E^{gccc} is an achievable exponent where*

$$E^{\text{gccc}} = \min_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{S}^{\text{gccc}}} D(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1,\hat{P}_{U_1}} \bar{Q}_{2,\hat{P}_{U_2}} W) + \left[I\left(X_\tau; Y | U_{\tau^c}, X_{\tau^c}\right) - H(\hat{P}_{U_\tau|U_{\tau^c}}) \right]^+, \tag{4.43}$$

and

$$\mathcal{S}^{\text{gccc}} \triangleq \Big\{ \hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}} : \hat{P}_{\underline{U}\underline{X}Y} = \hat{P}_{\underline{U}} \bar{Q}_{1,\hat{P}_{U_1}} \bar{Q}_{2,\hat{P}_{U_2}} \hat{P}_{Y|\underline{U}\underline{X}}, \hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}, \hat{P}_{Y|\underline{U}\underline{X}} \in \mathcal{P}_{\mathcal{Y}|\underline{\mathcal{U}} \times \underline{\mathcal{X}}} \Big\}. \tag{4.44}$$

Proof. See Section 4.3.9. □

Corollary 4.1. *Two sources with joint distribution $P_{\underline{U}}(\underline{u})$ can be transmitted reliably over a two-user MAC with conditional probability W, if satisfy the achievable region proposed by Cover, El Gamal and Salehi [12].*

As mentioned before, to have a reliable transmission exponents must be strictly positive. The derived exponents in (4.43) is always positive unless $D(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1,\hat{P}_{U_1}} \bar{Q}_{2,\hat{P}_{U_2}} W) = 0$. In this case, by imposing the positivity condition on the second terms of (4.43), Corollary 4.1 will be proved.

Next, by applying Lagrange duality theory, we find the dual form of (4.43). The derivation is exactly the same as the one presented in Section 2.4.3. Applying the same approach as Section 2.4.3, the dual form of (4.43) is obtained as

$$E^{\text{gccc}} = \max_{\gamma_1,\gamma_2 \in [0,1]} \min_{i_1,i_2=1,2} \min_{\tau} \max_{\rho \in [0,1]} E^{\text{gccc}}_{\tau,i_1,i_2}(\rho, \underline{\gamma}), \tag{4.45}$$

where

$$E^{\text{gccc}}_{\tau,i_1,i_2}(\rho, \underline{\gamma}) = \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} \max_{\substack{\bar{\beta}_1(u_1,x_1): \\ \sum_{x_1} \bar{\beta}_1(u_1,x_1)\bar{Q}_{1,i_1}(x_1|u_1)=0}} \max_{\substack{\bar{\beta}_2(u_2,x_2): \\ \sum_{x_2} \bar{\beta}_2(u_2,x_2)\bar{Q}_{2,i_2}(x_2|u_2)=0}} -\log\Bigg(\sum_{\substack{y, \\ u_{\tau^c}, x_{\tau^c}}} \Bigg(\sum_{u_\tau, x_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_1}(u_1)}{\gamma_1}\right)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} \left(\frac{P_{U_2}(u_2)}{\gamma_2}\right)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}} \times e^{\frac{\bar{\beta}_1(u_1,x_1)+\bar{\beta}_2(u_2,x_2)}{1+\rho}} \bar{Q}_{\tau,i_\tau}(x_\tau|u_\tau) \bar{Q}_{\tau^c,i_{\tau^c}}(x_{\tau^c}|u_{\tau^c})^{\frac{1}{1+\rho}} W(y|\underline{x})^{\frac{1}{1+\rho}} \Bigg)^{1+\rho} \Bigg). \tag{4.46}$$

To interpret (4.45), we start by recalling the properties of $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)$. Following the same steps given in the proof of Lemma 4.3, we conclude that for $\nu = 1,2$, $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=1}$ and $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=2}$ are respectively non-decreasing and non-increasing with respect to γ_ν. Thus, in view of Lemma A.8, the optimal γ_ν maximizing (4.45) is obtained at the points where

$$\min_{\tau}\max_{\rho\in[0,1]} E^{\text{gccc}}_{\tau,i_1,i_2}(\rho,\underline{\gamma})\Big|_{i_\nu=1} = \min_{\tau}\max_{\rho\in[0,1]} E^{\text{gccc}}_{\tau,i_1,i_2}(\rho,\underline{\gamma})\Big|_{i_\nu=2}, \tag{4.47}$$

and when (4.47) has no solution, the optimal γ_ν is either zero or one.

In addition, by comparing $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)$ in (4.46) with $E^{\text{mds}}_{\tau,i_1,i_2}(\cdot)$ in (4.37), we find that with respect to $E^{\text{mds}}_{\tau,i_1,i_2}(\cdot)$, the function $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)$ contains two extra constrains shown by $\bar{\beta}_1(u_1,x_1)$ $\bar{\beta}_2(u_2,x_2)$. Roughly speaking, these two constrains guarantee that for user $\nu = 1,2$, codewords of all the messages belonging to the class i_ν have the fixed conditional composition.

We note that by setting $\bar{\beta}_1(u_1,x_1) = \bar{\beta}_2(u_2,x_2u) = 0$ in (4.46), we have $E^{\text{mds}}_{\tau,i_1,i_2}(\cdot) = E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)$. And since $E^{\text{gccc}}_{\tau,i_1,i_2}(\cdot)$ contains maximization over $\bar{\beta}_\nu(u_\nu,x_\nu)$, we have $E^{\text{gccc}}_{\tau,i_1,i_2}(\rho,\underline{\gamma}) \geq E^{\text{mds}}_{\tau,i_1,i_2}(\rho,\underline{\gamma})$ yielding

$$E^{\text{gccc}} \geq E^{\text{mds}}, \tag{4.48}$$

where equality holds when the optimal $\bar{\beta}_1(u_1,x_1)$ and $\bar{\beta}_2(u_2,x_2)$ maximizing (4.46) be zero.

Now, we study the case where codewords and messages are statistically independent. By following exactly the same steps presented in Section 2.4.7, the dual form (4.43) when codewords and messages are statistically independent is obtained as

$$E^{\text{gcc}} = \max_{\gamma_1,\gamma_2\in[0,1]}\min_{i_1,i_2=1,2}\min_{\tau}\max_{\rho\in[0,1]} E^{\text{gcc}}_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau}) - E_{s,\tau,i_1,i_2}(\rho,P_{\underline{U}},\underline{\gamma}), \tag{4.49}$$

where $E_{s,\tau,i_1,i_2}(\cdot)$ is given by (4.12) and

$$E^{\text{gcc}}_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau}) = \max_{\substack{\bar{\alpha}_1(x_1):\\ \sum_{x_1}\bar{\alpha}_1(x_1)Q_1(x_1)=0}}\ \max_{\substack{\bar{\alpha}_2(x_2):\\ \sum_{x_2}\bar{\alpha}_1(x_2)Q_2(x_2)=0}}$$

$$-\log\sum_{x_{\tau^c},y}\left(\sum_{x_\tau} e^{\frac{\bar{\alpha}_1(x_1)}{1+\rho}} e^{\frac{\bar{\alpha}_2(x_2)}{1+\rho}} Q_{\tau,i_\tau}(x_\tau)W(y|\underline{x})^{\frac{1}{1+\rho}} Q_{\tau^c,i_{\tau^c}}(x_{\tau^c})^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{4.50}$$

Comparing (4.50) with $E_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau})$ appeared in (4.20), it can be verified that since input distributions in (4.50) have the fixed composition, $E^{\text{gcc}}_0(\cdot)$ contains two extra maximization over $\alpha_1(x_1)$ and $\alpha_2(x_2)$. Following the steps that are nearly identical to the proof of Lemma A.9, we can conclude that

$$E_0^{\rm gcc}(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_\tau^c}) \geq E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_\tau^c}), \tag{4.51}$$

and hence

$$E^{\rm gcc} \geq E^{\rm md}, \tag{4.52}$$

where $E^{\rm md}$ is given by (4.21).

To summarize the results, for the MAC with correlated source, we obtained $E^{\rm md}$, $E^{\rm mds}$, $E^{\rm gccc}$ and $E^{\rm gc}$, respectively given by (4.21), (4.36), (4.45) and (4.49). Considering (4.48) and (4.52), the final conclusion is

$$E^{\rm gccc} \geq E^{\rm mds}, \tag{4.53}$$

$$E^{\rm gcc} \geq E^{\rm md}. \tag{4.54}$$

However, unlike single-user communication, the comparison between $E^{\rm gcc}$ and $E^{\rm gccc}$ or the relation between $E^{\rm mds}$ and $E^{\rm md}$ are not as easy as the single-user communication.

The results obtained in this chapter are valid for the MAC with independent sources. However, by using the similar input distributions given by (2.61), we can conclude that statistical dependency between messages and codewords has no benefit for independent sources.

4.3 Proofs

4.3.1 Proof of Lemma 4.1

To prove (4.1), we use the idea of random bins [40]. Let $\nu = 1, 2$, for each sequence $\boldsymbol{U}_\nu$, an index is drawn randomly from $\left\{1, 2, .., 2^{nR_\nu}\right\}$. The set of all sequences $\boldsymbol{U}_\nu$ which have the same index are said to form a bin.

To generate codebooks, every $\boldsymbol{u}_\nu \in \mathcal{U}_\nu^n$ is assigned to one of 2^{nR_ν} bins independently according to a uniform distribution on $\left\{1, 2, .., 2^{nR_\nu}\right\}$. Assume that the index of the bin to which $\boldsymbol{u}_\nu$ belongs, is x_ν. Thus, encoder $\phi_\nu : \mathcal{U}_\nu^n \to \left\{1, 2, .., 2^{nR_\nu}\right\}$, sends x_ν, i. e.

$$\phi(\boldsymbol{u}_\nu) = x_\nu. \tag{4.55}$$

Decoder, by receiving the bin indices (x_1, x_2), declares $\hat{\underline{\boldsymbol{u}}}$ as the transmitted message if

$$\hat{\underline{\boldsymbol{u}}} = \arg\max_{(\boldsymbol{u}_1,\boldsymbol{u}_2)\in\mathcal{U}_1^n\times\mathcal{U}_2^n} P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}}) \mathbb{1}\{\phi_1(\boldsymbol{u}_1) = x_1\} \mathbb{1}\{\phi_2(\boldsymbol{u}_2) = x_2\}, \tag{4.56}$$

where $\mathbb{1}\{\cdot\}$ is an indicator function. An error occurs if $\hat{\underline{\boldsymbol{u}}} \neq \underline{\boldsymbol{u}}$, which can be split into three events, namely $(\hat{\boldsymbol{u}}_1, \boldsymbol{u}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $(\boldsymbol{u}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$. We respectively denote these types of error by $\tau \in \{\{1\}, \{2\}, \{1,2\}\}$.

Let P_e denotes the the average error probability. Like before, using the random-coding union bound [23] and the fact that $\min\{1, a+b\} \leq \min\{1, a\} + \min\{1, b\}$, we find that

$$P_e \leq \sum_{\tau \in \{\{1\},\{2\},\{1,2\}\}} P_e^\tau, \tag{4.57}$$

where

$$P_e^\tau \leq \sum_{\underline{\boldsymbol{u}}} P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}}) \sum_{\underline{x}} P_{X_1}(x_1) P_{X_2}(x_2) \min\left\{1, \sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} \mathbb{P}\left[\frac{P_{\underline{\boldsymbol{U}}}^n(\hat{\boldsymbol{u}}_\tau, \boldsymbol{u}_{\tau^c}) \mathbb{1}\{\phi_\tau(\hat{\boldsymbol{u}}_\tau) = x_\tau\}}{P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}})} \geq 1\right]\right\}, \tag{4.58}$$

and P_{X_ν} is uniform over $\left\{1, 2, .., 2^{nR_\nu}\right\}$. We recall that the probability given inside of (4.58), is the probability that $\phi_\tau(\hat{\boldsymbol{u}}_\tau) = x_\tau$ when $P_{\underline{\boldsymbol{U}}}^n(\hat{\boldsymbol{u}}_\tau, \boldsymbol{u}_{\tau^c}) \geq P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}})$.

Applying Markov's inequality to (4.58), we find that

$$P_e^\tau \leq \sum_{\underline{\boldsymbol{u}}} P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}}) \sum_{\underline{x}} P_{X_1}(x_1) P_{X_2}(x_2) \min\left\{1, 2^{-nR_\tau} \sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} \left(\frac{P_{\underline{\boldsymbol{U}}}^n(\hat{\boldsymbol{u}}_\tau, \boldsymbol{u}_{\tau^c})}{P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}})}\right)^s\right\}, \tag{4.59}$$

where in (4.59), we used the fact that $\mathbb{P}\left[\phi(\hat{\boldsymbol{u}}_\tau) = x_\tau\right] = \frac{1}{2^{nR_\tau}}$. Using the inequality $\min\{1, a\} \leq \min_{\rho \in [0,1]} a^\rho$ and inserting $P_{X_\nu}(x_\nu) = 2^{-nR_\nu}$ to (4.59), P_e^τ is bounded as

$$P_e^\tau \leq \min_{\rho \in [0,1]} \sum_{\underline{\boldsymbol{u}}} P_{\underline{\boldsymbol{U}}}^n(\underline{\boldsymbol{u}})^{1-s\rho} \sum_{\underline{x}} 2^{-nR_1} 2^{-nR_2} \left(2^{-nR_\tau} \sum_{\hat{\boldsymbol{u}}_\tau \neq \boldsymbol{u}_\tau} P_{\underline{\boldsymbol{U}}}^n(\hat{\boldsymbol{u}}_\tau, \boldsymbol{u}_{\tau^c})^s\right)^\rho, \tag{4.60}$$

where since P_{X_ν} is uniform over $\left\{1, ..., 2^{nR_\nu}\right\}$, we have $\sum_{x_\nu} 2^{-nR_\nu} = 1$. By applying this fact to (4.60), using the memoryless property of the sources, after some simple mathematical manipulations, we find that

$$-\frac{1}{n} \log(P_e^\tau) \geq \max_{s \geq 0} \max_{\rho \in [0,1]} \rho R_\tau - \log\left(\sum_{\underline{u}} P_{\underline{U}}(\underline{u})^{1-s\rho} \left(\sum_{u_\tau} P_{\underline{U}}(u_\tau, u_{\tau^c})^s\right)^\rho\right), \tag{4.61}$$

where in (4.61), to tight the bound, we maximized over $s \geq 0$. Using Hölder's inequality it can be proved that the optimal $s = \frac{1}{1+\rho}$ maximizes (4.61) [5, Prob. 5.6]. Hence,

$$-\frac{1}{n}\log(P_e^\tau) \geq \max_{\rho\in[0,1]} \rho R_\tau - \log\sum_{u_{\tau^c}}\left(\sum_{u_\tau} P_{\underline{U}}(u_\tau, u_{\tau^c})^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{4.62}$$

Next, in view of (4.57), by upper bounding the summation by the worst type of error, we conclude the proof.

4.3.2 Proof of Proposition 4.1

The proof of Proposition 4.1 is similar to the one presented in Section 2.4.1. Bounding $\bar{\epsilon}^n$, the average error probability over the ensemble, and applying the random coding union bound [23] for joint source channel coding, we obtain

$$\bar{\epsilon}^n \leq \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \min\left\{1, \sum_{\hat{\underline{\boldsymbol{u}}}\neq\underline{\boldsymbol{u}}} \mathbb{P}\left[\frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\underline{\boldsymbol{u}}})W^n(\boldsymbol{y}|\hat{\underline{\boldsymbol{X}}})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\underline{\boldsymbol{x}})} \geq 1\right]\right\}, \tag{4.63}$$

where $\hat{\underline{\boldsymbol{x}}}$ has the same distribution as $\underline{\boldsymbol{x}}$ but is independent of $\boldsymbol{y}$. We group the error events corresponding to the summation over $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$ into three types of error events, namely $(\hat{\boldsymbol{u}}_1, \boldsymbol{u}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $(\boldsymbol{u}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) \neq (\boldsymbol{u}_1, \boldsymbol{u}_2)$. We respectively denote these types of error by $\tau \in \{\{1\}, \{2\}, \{1, 2\}\}$. Using that $\min\{1, a + b\} \leq \min\{1, a\} + \min\{1, b\}$, we further bound $\bar{\epsilon}^n$ as

$$\bar{\epsilon}^n \leq \sum_{\tau} \bar{\epsilon}^n_\tau, \tag{4.64}$$

where

$$\bar{\epsilon}^n_\tau \leq \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \min\left\{1, \sum_{\substack{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau \\ \hat{\boldsymbol{x}}_\tau : \frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\boldsymbol{u}}_\tau,\boldsymbol{u}_{\tau^c})W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)}\geq 1}} \bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau)\right\}. \tag{4.65}$$

Like Section 2.4.1, we group the outer and inner summations in (4.65) based on their empirical distributions. Let $\hat{P}_{\underline{U}\underline{X}Y}$ denotes a possible empirical distribution of $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y})$. Since there is no constraint on $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y})$, $\hat{P}_{\underline{U}\underline{X}Y}$ runs over the set of all possible empirical distributions, $\mathcal{P}^n_{\underline{U}\times\underline{X}\times Y}$. On the other hand, since based on the MAP criterion, $(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau)$ leads to error, the empirical distribution of $(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau)$ denoted by $\tilde{P}_{\underline{U}\underline{X}Y}$ is restricted to the set $\mathcal{L}_\tau^n$, defined as

$$\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y}) \triangleq \Big\{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}} : \tilde{P}_{U_{\tau^c}X_{\tau^c}Y} = \hat{P}_{U_{\tau^c}X_{\tau^c}Y}, \\ \mathbb{E}_{\tilde{P}}[\lambda(\underline{U}, \underline{X}, Y)] \geq \mathbb{E}_{\hat{P}}[\lambda(\underline{U}, \underline{X}, Y)]\Big\}. \tag{4.66}$$

As a result, we can write the summations in equation (4.65) respectively as

$$\sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) = \sum_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} \sum_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})\in\mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}), \tag{4.67}$$

and

$$\sum_{\substack{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau \\ \hat{\boldsymbol{x}}_\tau : \frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\boldsymbol{u}}_\tau, \boldsymbol{u}_{\tau^c})W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)} \geq 1}} \bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau) = \\ \sum_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} \sum_{(\hat{\boldsymbol{u}}_\tau,\hat{\boldsymbol{x}}_\tau)\in\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau), \tag{4.68}$$

where $\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\cdot)$ is defined by (1.40).

Since the conditional distribution $\bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau)$ has the same value for all $(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})$, we have

$$\sum_{(\hat{\boldsymbol{u}}_\tau,\hat{\boldsymbol{x}}_\tau)\in\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau) = |\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})|\bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau). \tag{4.69}$$

Considering (1.40) and the fact that $\tilde{P}_{U_{\tau^c}X_{\tau^c}Y} = \hat{P}_{U_{\tau^c}X_{\tau^c}Y}$ in $\mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})$ in (4.66), we have the following upper bound

$$\left|\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})\right| = \frac{\left|\mathcal{T}^n(\tilde{P}_{\underline{U}\underline{X}Y})\right|}{\left|\mathcal{T}^n(\tilde{P}_{U_{\tau^c}X_{\tau^c}Y})\right|} \leq \frac{e^{nH(\tilde{P}_{\underline{U}\underline{X}Y})+o(n)}}{e^{nH(\hat{P}_{U_{\tau^c}X_{\tau^c}Y})}}, \tag{4.70}$$

where $o(n)$ is a sequence satisfying $\lim_{n\to\infty}\frac{o(n)}{n} = 0$. In addition, using equation (1.37) for conditional distributions, for all $(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})$, we have the following identity on the conditional probability

$$\bar{Q}^n_{\tau,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau) = e^{n\sum_{\underline{u},\underline{x},y}\tilde{P}_{\underline{U}\underline{X}Y}(\underline{u},\underline{x},y)\log\bar{Q}_{\tau,\tilde{P}_{U_\tau}}(x_\tau|u_\tau)}. \tag{4.71}$$

Combining inequality (4.70) and identity (4.71) and into (4.69), we obtain the following inequality

$$\sum_{(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c} \boldsymbol{x}_{\tau^c} \boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau, \pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau) \leq$$
$$e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau, \tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c} X_{\tau^c} Y}) - H(\tilde{P}_{U_\tau})\right) + o(n)}. \tag{4.72}$$

Further upper bounding the right hand side of equation (4.72) by the maximum over the empirical probability distributions $\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})$, we have

$$\sum_{(\hat{\boldsymbol{u}}_\tau, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c} \boldsymbol{x}_{\tau^c} \boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau, \pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau) \leq$$
$$\max_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau, \tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c} X_{\tau^c} Y}) - H(\tilde{P}_{U_\tau})\right) + o(n)}. \tag{4.73}$$

Moreover, in view of (1.39), the second summation of the right hand side of (4.67) can be expressed as

$$\sum_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \leq e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1, \hat{P}_{U_1}} \bar{Q}_{2, \hat{P}_{U_2}} W\right)}, \tag{4.74}$$

where $\hat{P}_{U_\nu}$ denotes the marginal distribution of $\hat{P}_{\underline{U}}$, for $\nu = 1, 2$. Similarly to (4.73), we may upper bound the right hand side of (4.74) by the maximum over the empirical distributions $\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}$, i. e.

$$\sum_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \leq \max_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1, \hat{P}_{U_1}} \bar{Q}_{2, \hat{P}_{U_2}} W\right)}. \tag{4.75}$$

Putting back the results obtained in equations (4.75) and (4.73) into the respective inner and outer summations (4.67) and (4.68), we obtain that the average error probability (4.65) can be bounded as

$$\bar{\epsilon}^n_\tau \leq \sum_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}} \max_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1, \hat{P}_{U_1}} \bar{Q}_{2, \hat{P}_{U_2}} W)\right)} \min \Bigg\{ 1,$$
$$\sum_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} \max_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau, \tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c} X_{\tau^c} Y}) - H(\tilde{P}_{U_\tau})\right) + o(n)} \Bigg\}, \tag{4.76}$$

where by using the fact that the cardinality of the sets $\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y})$ and $\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}$ behave polynomially with the codeword length n, and satisfy $\left|\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y})\right| \leq \left|\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}\right| \leq e^{o(n)}$, we find that

$$\bar{\epsilon}_\tau^n \leq \max_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W)\right)+o(n)} \min\Bigg\{1,$$
$$\max_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y})-H(\tilde{P}_{U_\tau})\right)+o(n)}\Bigg\}, \quad (4.77)$$

Using the identity $\min\{1, e^a\} = e^{[a]^+}$, we may write equation (4.77) as

$$\bar{\epsilon}_\tau^n \leq e^{-nE_\tau^n+o(n)}, \quad (4.78)$$

where

$$E_\tau^n = \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W)$$
$$+\left[\min_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y})} D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right]^+. \quad (4.79)$$

Since the average error probability over the ensemble is bounded by the summation over the error events, we further upper bound the summation by the worst type of error, i. e.

$$\sum_\tau \bar{\epsilon}_\tau^n \leq e^{-n\min_\tau E_\tau^n+o(n)}. \quad (4.80)$$

Hence, from (4.64), we conclude that $\bar{\epsilon}^n$ is upper bounded by the right hand side of (4.80), i. e.

$$\bar{\epsilon}^n \leq e^{-n\min_\tau E_\tau^n+o(n)}. \quad (4.81)$$

Using the following properties

$$\liminf_{n\to\infty}(a_n+b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \quad (4.82)$$
$$\liminf_{n\to\infty}\min\{a_n,b_n\} = \min\Big\{\liminf_{n\to\infty} a_n, \liminf_{n\to\infty} b_n\Big\}, \quad (4.83)$$

we obtain that $\bar{\epsilon}^n$ asymptotically satisfies

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\bar{\epsilon}^n) \geq$$
$$\min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \liminf_{n\to\infty} \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W)$$
$$+\liminf_{n\to\infty}\left[\min_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}_\tau^n(\hat{P}_{\underline{U}\underline{X}Y})} D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right]^+. \quad (4.84)$$

We note that the inequality

$$\liminf_{n\to\infty} \max\{a_n, b_n\} \geq \max\Big\{\liminf_{n\to\infty} a_n, \liminf_{n\to\infty} b_n\Big\}, \tag{4.85}$$

implies that

$$\liminf_{n\to\infty} [a_n]^+ \geq \Big[\liminf_{n\to\infty} a_n\Big]^+. \tag{4.86}$$

We further note that the set of all empirical distributions is dense in the set of all possible probability distributions, and that the functions involved in (4.84) are uniformly continuous over their arguments. Hence, we may replace the optimization over empirical distributions by an optimization over the set of all possible distributions. Using (4.86) in (4.84), we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\bar{\epsilon}^n_\tau) \geq \min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{P}_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W)$$
$$+\Bigg[\min_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}_\tau(\hat{P}_{\underline{U}\underline{X}Y})} D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\Bigg]^+, \tag{4.87}$$

where $\mathcal{L}_\tau(\hat{P}_{\underline{U}\underline{X}Y})$ is defined in (4.3). By renaming the right hand side of (4.87) as E_1^{mds}, we conclude the proof.

4.3.3 Proof of (4.4)

In Section 4.3.3.1, by following the same method given in [41, Th. 1], we showed that

$$\bar{\epsilon} \geq \max_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \frac{1}{4}\sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})$$
$$\min\Bigg\{1, \sum_{\boldsymbol{u}'_\tau\neq\boldsymbol{u}_\tau}\mathbb{P}\Bigg[\frac{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}')W^n(\boldsymbol{y}|\underline{\boldsymbol{X}}')}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\underline{\boldsymbol{x}})}\geq 1\Bigg]\Bigg\}, \tag{4.88}$$

or equivalently

$$\bar{\epsilon} \geq \max_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \frac{1}{4}\sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})$$
$$\min\Bigg\{1, \sum_{\substack{\boldsymbol{u}'_\tau\neq\boldsymbol{u}_\tau \\ \boldsymbol{x}'_\tau:\frac{P^n_{\underline{U}}(\boldsymbol{u}'_\tau,\boldsymbol{u}_{\tau^c})W^n(\boldsymbol{y}|\boldsymbol{x}'_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\underline{U}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\boldsymbol{x}_1,\boldsymbol{x}_2)}\geq 1}} \bar{Q}^n_{\tau,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau)\Bigg\}, \tag{4.89}$$

where (4.89) follows from the fact that the inner probability in (4.88) is equal with the summation over all codewords which are distributed according $\bar{Q}_{\tau,\pi(\boldsymbol{u}'_\tau)}$ and give rise to an error according to the MAP criterion.

Next, we group the outer and inner summations in (4.89) based on the empirical distributions of $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y})$ and $(\boldsymbol{u}'_\tau, \boldsymbol{x}'_\tau)$, respectively, and then sum over all possible empirical distributions, respectively denoted by $\hat{P}_{\underline{U}\underline{X}Y}$ and $\tilde{P}_{\underline{U}\underline{X}Y}$. In view of (4.67) and (4.68), we have

$$\sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) = \sum_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} \sum_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}), \quad (4.90)$$

and

$$\sum_{\substack{\boldsymbol{u}'_\tau \neq \boldsymbol{u}_\tau \\ \boldsymbol{x}'_\tau : \frac{P^n_{\underline{U}}(\boldsymbol{u}'_\tau, \boldsymbol{u}_{\tau^c}) W^n(\boldsymbol{y}|\boldsymbol{x}'_\tau, \boldsymbol{x}_{\tau^c})}{P^n_{\underline{U}}(\underline{\boldsymbol{u}}) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)} \geq 1}} \bar{Q}^n_{\tau,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau|\boldsymbol{u}'_\tau) = \sum_{\substack{\tilde{P}_{\underline{U}\underline{X}Y} \in \\ \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})}} \sum_{\substack{(\boldsymbol{u}'_\tau, \boldsymbol{x}'_\tau) \in \\ \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})}} \bar{Q}^n_{\tau,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau|\boldsymbol{u}'_\tau), \quad (4.91)$$

where the set $\mathcal{L}^n_\tau$ is defined by (4.66).

To compute the right hand of (4.90) and (4.91), we use [15, Lemma 2.3] and [15, Lemma 2.6] which lead to

$$\sum_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \geq e^{-n\left(D\left(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W\right)\right)+o(n)}, \quad (4.92)$$

and since the type of $\boldsymbol{u}'_\tau$, $\pi(\boldsymbol{u}'_\tau)$ can be written as $\tilde{P}_{U_\tau}$, we have

$$\sum_{\substack{(\boldsymbol{u}'_\tau, \boldsymbol{x}'_\tau) \in \\ \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})}} \bar{Q}^n_{\tau,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau|\boldsymbol{u}'_\tau) \geq e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right)+o(n)}, \quad (4.93)$$

where $o(n)$ is a sequence satisfying $\lim_{n\to\infty} \frac{o(n)}{n} = 0$.

Putting back (4.92) and (4.93) into (4.89), the average error probability is bounded as

$$\bar{\epsilon} \geq \max_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \frac{1}{4} \sum_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,\hat{P}_{U_1}}\bar{Q}_{2,\hat{P}_{U_2}}W\right)+o(n)}$$

$$\min\left\{1, \sum_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right)+o(n)}\right\}. \quad (4.94)$$

Lower bounding the right hand of (4.94) by considering only the maximum terms in each summation, using the identity $\min\{1, e^a\} = e^{[a]^+}$, taking logarithms on both sides of (4.94) and multiplying result by $-\frac{1}{n}$ we obtain

$$-\frac{1}{n}\bar{\epsilon} \leq \min_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \min_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}^n_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}} D\left(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1,\hat{P}_{U_1}} \bar{Q}_{2,\hat{P}_{U_2}} W\right) + \frac{o(n)}{n}$$
$$\left[\min_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^n_\tau(\hat{P}_{\underline{U}\underline{X}Y})} D\left(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau,\tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c} X_{\tau^c} Y}\right) - H(\tilde{P}_{U_\tau}) + \frac{o(n)}{n}\right]^+ . \tag{4.95}$$

Now we take lim sup from both sides of (4.95) and use the following properties

$$\limsup_{n\to\infty} \min\{a_n, b_n\} \leq \min\left\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\right\}, \tag{4.96}$$
$$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{4.97}$$
$$\limsup_{n\to\infty} \max\{a_n, b_n\} = \max\left\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\right\}, \tag{4.98}$$

we obtain

$$\limsup_{n\to\infty} -\frac{1}{n}\bar{\epsilon} \leq \min_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \min_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{P}_{\underline{\mathcal{U}} \times \underline{\mathcal{X}} \times \mathcal{Y}}} D\left(\hat{P}_{\underline{U}\underline{X}Y} || P_{\underline{U}} \bar{Q}_{1,\hat{P}_{U_1}} \bar{Q}_{2,\hat{P}_{U_2}} W\right) +$$
$$\left[\min_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}_\tau(\hat{P}_{\underline{U}\underline{X}Y})} D\left(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau,\tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c} X_{\tau^c} Y}\right) - H(\tilde{P}_{U_\tau})\right]^+ , \tag{4.99}$$

where in (4.99) we used the facts that the set of all empirical distributions is dense in the set of all possible probability distributions, and that the functions involved in (4.95) are uniformly continuous over their arguments. Hence, we may replace the optimization over empirical distributions by an optimization over the set of all possible distributions.

4.3.3.1 Proof of Equation (4.89)

As mentioned before, the error probability for a given pair of codebooks is denoted by $\epsilon(\mathcal{C}^1, \mathcal{C}^2)$ where

$$\epsilon(\mathcal{C}^1, \mathcal{C}^2) \triangleq \mathbb{P}\Big[(\boldsymbol{U}_1', \boldsymbol{U}_2') \neq (\boldsymbol{U}_1, \boldsymbol{U}_2)\Big] \geq \max_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \mathbb{P}[\boldsymbol{U}_\tau' \neq \boldsymbol{U}_\tau], \tag{4.100}$$

where in (4.100), we group the error events into three types of error events and we used the fact that the probability of union of some events is greater than than the probability of each individual event and specifically the more probable one. Thus, in view of (4.100), we found that

$$\epsilon(\mathcal{C}^1, \mathcal{C}^2) \geq \max_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \mathbb{P}[\boldsymbol{U}'_\tau \neq \boldsymbol{U}_\tau]. \tag{4.101}$$

Now, for given codebooks $\mathcal{C}^1$ and $\mathcal{C}^2$ and for type τ error, let $B_{0,\tau}$ be the event that one or more codewords yield a strictly higher metric than the transmitted one, and let $B_{l,\tau}$ be the event that the transmitted codeword yields a metric which is equal with l other codewords. In view of (4.101), for type τ error, we have

$$\epsilon(\mathcal{C}^1, \mathcal{C}^2) \geq \max_{\tau \in \{\{1\},\{2\},\{1,2\}\}} \mathbb{P}\big[\boldsymbol{U}'_\tau \neq \boldsymbol{U}_\tau\big] = \mathbb{P}[B_{0,\tau}] + \sum_{l=1}^{|\mathcal{U}_\tau|-1} \mathbb{P}[B_{l,\tau}] \frac{l}{l+1} \tag{4.102}$$

$$\geq \mathbb{P}[B_{0,\tau}] + \frac{1}{2} \sum_{l=1}^{|\mathcal{U}_\tau|-1} \mathbb{P}[B_{l,\tau}] \tag{4.103}$$

$$= \frac{1}{2}\epsilon'_\tau(\mathcal{C}^1, \mathcal{C}^2) + \frac{1}{2}\mathbb{P}[B_{0,\tau}] \tag{4.104}$$

$$\geq \frac{1}{2}\epsilon'_\tau(\mathcal{C}^1, \mathcal{C}^2), \tag{4.105}$$

where (4.103) follows by noting to the fact that the $\{\frac{n}{1+n}\}$ is an increasing sequence, in (4.104) we defined $\epsilon'_\tau(\mathcal{C}^1, \mathcal{C}^2) \triangleq \mathbb{P}[B_{0,\tau}] + \sum_{l=1}^{|\mathcal{U}_\tau|-1} \mathbb{P}[B_{l,\tau}]$ as the error probability of a decoder which decodes ties as errors and the inequality in (4.105) follows by lower bounding (4.104) by the first of the two terms. Averaging (4.105) over the random-coding distribution, we obtain

$$\bar{\epsilon} \geq \max_\tau \frac{1}{2}\mathbb{E}\left(\mathbb{P}\Big[\bigcup_{\boldsymbol{U}'_\tau \neq \boldsymbol{U}_\tau} \Big\{\frac{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{U}}')W^n(\boldsymbol{Y}|\underline{\boldsymbol{X}}'(\underline{\boldsymbol{U}}'))}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{U}})W^n(\boldsymbol{Y}|\underline{\boldsymbol{X}}(\underline{\boldsymbol{U}}))} \geq 1 \Big| \underline{\boldsymbol{U}}, \underline{\boldsymbol{X}}, \boldsymbol{Y}\Big\}\Big]\right) \tag{4.106}$$

$$\geq \max_\tau \frac{1}{4} \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \min\left\{1, \sum_{\boldsymbol{u}'_\tau \neq \boldsymbol{u}_\tau} \mathbb{P}\left[\frac{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}}')W^n(\boldsymbol{y}|\underline{\boldsymbol{X}}')}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\underline{\boldsymbol{x}})} \geq 1\right]\right\}, \tag{4.107}$$

where $\underline{\boldsymbol{X}}'$ has the same distribution as $\underline{\boldsymbol{X}}$ but is independent of $\boldsymbol{Y}$ and (4.107) follows by the approach given in [42]. The verification of (4.107) is explained in the following.

In [42], it is shown that for an arbitrary sequence of probabilistic events $A_1, ..., A_k$, we have

$$\mathbb{P}\Big[\bigcup_{i=1}^k A_i\Big] \geq \sum_{i=1}^k \frac{\mathbb{P}[A_i]^2}{\sum_{j=1}^k \mathbb{P}[A_i \cap A_j]}, \tag{4.108}$$

where for each event A_j, we have $\mathbb{P}[A_j] \leq \frac{1}{k} + \mathbb{P}[A_j]$ which leads to $\mathbb{P}[A_i]\mathbb{P}[A_j] \leq \mathbb{P}[A_i]\Big(\frac{1}{k} + \mathbb{P}[A_j]\Big)$. By summing both sides of last inequality over j we find that

$$\sum_{j=1}^{k} \mathbb{P}[A_i]\mathbb{P}[A_j] \leq \mathbb{P}[A_i]\Big(1 + \sum_{j=1}^{k} \mathbb{P}[A_j]\Big). \tag{4.109}$$

In addition, for the case that the events are pairwise independent we have $\mathbb{P}[A_i \bigcap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j]$. Substituting (4.109) into (4.108), we find that if the events are pairwise independent we have

$$\mathbb{P}\Big[\bigcup_{i=1}^{k} A_i\Big] \geq \sum_{i=1}^{k} \frac{\mathbb{P}[A_i]}{1 + \sum_{j=1}^{k} \mathbb{P}[A_j]} = \frac{\sum_{i=1}^{k} \mathbb{P}[A_i]}{1 + \sum_{j=1}^{k} \mathbb{P}[A_j]} \geq \frac{1}{2} \min\Big\{1, \sum_{i=1}^{k} \mathbb{P}[A_i]\Big\}, \tag{4.110}$$

where the last inequality in (4.110) follows by applying the inequality $\frac{\alpha}{1+\alpha} \geq \frac{1}{2}\min\{1, \alpha\}$. Applying (4.110) into (4.106), we obtain the lower bound given in (4.107).

4.3.4 Proof of Lemma 4.2

Recalling $\mathcal{P}^n_{\underline{\mathcal{U}}}$ is the set of all empirical distributions on a joint vector in $\underline{\mathcal{U}}^n$, and $\mathcal{T}^n(\tilde{P}_{\underline{U}})$ is the set of all joint sequences in $\underline{\mathcal{U}}^n$ with empirical distribution $\tilde{P}_{\underline{U}}$, $\underline{\mathcal{U}}^n$ can be partitioned by all possible empirical distributions, i. e. $\underline{\mathcal{U}}^n = \bigcup_{\tilde{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}}} \mathcal{T}^n(\tilde{P}_{\underline{U}})$. Since all $\underline{\boldsymbol{u}}$ belonging to the set $\mathcal{T}^n(\tilde{P}_{\underline{U}})$ has the same probability, the set $\underline{\mathcal{U}}^n$ can be partitioned into two classes $\mathcal{A}^1_\nu(\gamma_\nu)$ and $\mathcal{A}^2_\nu(\gamma_\nu)$ as

$$\mathcal{A}^1_\nu(\gamma_\nu) \triangleq \Big\{\underline{\boldsymbol{u}} \in \bigcup_{\tilde{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}}} \mathcal{T}^n(\tilde{P}_{\underline{U}}) \;:\; P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) \geq \gamma^n_\nu\Big\}, \tag{4.111}$$

$$\mathcal{A}^2_\nu(\gamma_\nu) \triangleq \Big\{\underline{\boldsymbol{u}} \in \bigcup_{\tilde{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}}} \mathcal{T}^n(\tilde{P}_{\underline{U}}) \;:\; P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) < \gamma^n_\nu\Big\}, \tag{4.112}$$

for a given $\gamma_\nu \in [0, 1]$ where $\nu = 1, 2$.

By letting ν^c as the complement of $\nu \in \{1, 2\}$, and noting that for $\underline{\boldsymbol{u}} \in \mathcal{T}^n(\tilde{P}_{\underline{U}})$, the sequence $\boldsymbol{u}_\nu$ contains exactly $n \sum_{u_{\nu^c}} \tilde{P}_{\underline{U}}(\underline{u})$ occurrences of u_ν, the probability of $\boldsymbol{u}_\nu$ is $P_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) = \prod_{u_\nu \in \mathcal{U}_\nu} P_{U_\nu}(u_\nu)^{n \sum_{u_{\nu^c}} \tilde{P}_{\underline{U}}(\underline{u})}$. Let $\mathcal{T}^n(\tilde{P}_{\underline{U}}) \subseteq \mathcal{A}^1_\nu(\gamma_\nu)$, for $\underline{\boldsymbol{u}}^n \in \mathcal{T}^n(\tilde{P}_{\underline{U}})$, the condition $P^n_{\boldsymbol{U}_\nu}(\boldsymbol{u}_\nu) \geq \gamma^n_\nu$ can be written as $\prod_{u_\nu \in \mathcal{U}_\nu} P_{U_\nu}(u_\nu)^{n \sum_{u_{\bar{\nu}}} \tilde{P}_{\underline{U}}(\underline{u})} \geq \gamma^n_\nu$ where by taking logarithm from both sides, it is simplified as $\sum_{\underline{u}} \tilde{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) \geq \log(\gamma_\nu)$. Using the same reasoning for $\mathcal{T}^n(\tilde{P}_{\underline{U}}) \subseteq \mathcal{A}^2_\nu(\gamma_\nu)$, the sets $\mathcal{A}^1_\nu(\gamma_\nu)$ and $\mathcal{A}^2_\nu(\gamma_\nu)$ can be rewritten as

$$\mathcal{A}^1_\nu(\gamma_\nu) = \left\{ \tilde{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}} : \sum_{\underline{u}} \tilde{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) \geq \log(\gamma_\nu) \right\}, \tag{4.113}$$

$$\mathcal{A}^2_\nu(\gamma_\nu) = \left\{ \tilde{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}} : \sum_{\underline{u}} \tilde{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) < \log(\gamma_\nu) \right\}, \tag{4.114}$$

where in (4.113) and (4.114), we express $\mathcal{A}^1_\nu(\gamma_\nu)$ and $\mathcal{A}^2_\nu(\gamma_\nu)$ in terms of empirical distributions.

As n tends to infinity, since the set of all empirical distributions is dense in the set of all possible probability distributions $\mathcal{P}_{\underline{\mathcal{U}}}$, the sets $\mathcal{A}^1_\nu(\gamma_\nu)$ and $\mathcal{A}^2_\nu(\gamma_\nu)$, respectively tend to $\mathcal{B}^1_\nu(\gamma_\nu)$ and $\mathcal{B}^2_\nu(\gamma_\nu)$ given by (4.9) and (4.10), and hence Lemma 4.2 is proved.

4.3.5 Proof of Proposition 4.2

To prove Proposition 4.2, we start by finding the dual form of the following problem.

$$\min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}} : \hat{P}_{\underline{U}} \in \mathcal{B}^{i_1}_1(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}^{i_2}_2(\gamma_2)} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau|U_{\tau^c}}), \tag{4.115}$$

by applying Lagrange duality theory to the minimization problem. We use λ_1 and λ_2 as the Lagrange multipliers, respectively associate with the constraints $\hat{P}_{\underline{U}} \in \mathcal{B}^{i_1}_1(\gamma_1)$ and $\hat{P}_{\underline{U}} \in \mathcal{B}^{i_2}_2(\gamma_2)$.

We simplify the objective function of (4.115). Since $D(\hat{P}_{U_{\tau^c}} || V_{U_{\tau^c}}) \geq 0$, for any $V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}$, we have $\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \hat{P}_{U_{\tau^c}}(u_{\tau^c}) \geq \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log V_{U_{\tau^c}}(u_{\tau^c})$. Multiplying both sides of the inequality by -1 and adding $-H(\hat{P}_{\underline{U}})$ to the both sides of it, we find that $\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{\hat{P}_{U_{\tau^c}}(u_{\tau^c})} \leq \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}$. Recalling the definition of $H(\hat{P}_{U_\tau|U_{\tau^c}})$, the left hand side of the inequality is $-H(\hat{P}_{U_\tau|U_{\tau^c}})$ meaning that $-H(\hat{P}_{U_\tau|U_{\tau^c}}) \leq \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}$. From the last inequality, we conclude that the right hand side of the inequality is always greater than $-H(\hat{P}_{U_\tau|U_{\tau^c}})$ and only is equal to $-H(\hat{P}_{U_\tau|U_{\tau^c}})$ when $V_{U_{\tau^c}}(u_{\tau^c}) = P_{U_{\tau^c}}(u_{\tau^c})$ for all values of $u_{\tau^c} \in \mathcal{U}_{\tau^c}$, i. e. $\min_{V_{U_{\tau^c}}} \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})} = -H(\hat{P}_{U_\tau|U_{\tau^c}})$. By applying this fact to the the objective function of (4.115), we obtain

$$D(\hat{P}_{\underline{U}}||P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau|U_{\tau^c}}) = \min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}. \tag{4.116}$$

Applying (4.116) to the objective function of (4.115), we find that

$$\min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) - \rho H(\hat{P}_{U_\tau|U_{\tau^c}}) = \min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}}$$
$$\min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}. \tag{4.117}$$

Now, we apply Lagrange duality theory to the inner minimization over $\hat{P}_{\underline{U}}$ in (4.117). Considering the constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1)$ and $\hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)$, and in view of definitions (4.9) and (4.10), the Lagrangian associated with the primal is given by

$$\Lambda(V_{U_{\tau^c}}, \hat{P}_{\underline{U}}, \theta, \lambda_1, \lambda_2) = D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}$$
$$+\theta \left(1 - \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u})\right) + (-1)^{i_1} \lambda_1 \left(\sum_{\underline{u}, \underline{x}, y} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_1}(u_1) - \log \gamma_1\right)$$
$$+(-1)^{i_2} \lambda_2 \left(\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_2}(u_2) - \log \gamma_2\right), \tag{4.118}$$

where λ_1, λ_2 and θ are respectively the Lagrange multipliers associated with the inequalities constraints $\hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1)$, $\hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)$ and the sum of any probability distribution over its alphabet is one.

Noting that the objective function and the inequalities constrains given by (4.9) and (4.10) are convex with respect to $\hat{P}_{\underline{U}}$, and the equality constraint is affine, strong duality conditions are satisfied. Thus, the primal optimal objective and the dual optimal objective are equal,

$$\min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1) \cap \mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})}$$
$$= \min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} \max_{\theta} \min_{\hat{P}_{\underline{U}}} \Lambda(V_{U_{\tau^c}}, \hat{P}_{\underline{U}}, \theta, \lambda_1, \lambda_2), \tag{4.119}$$

where we recall that for $\nu = 1, 2$, the condition $\lambda_\nu \geq 0$ in (4.119) associated with inequality constraint $(-1)^{i_\nu} \left(\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log P_{U_\nu}(u_\nu) - \log(\gamma_\nu) < 0\right)$.

Since the strong duality holds, in view of the KKT conditions, by setting $\frac{\partial \Lambda}{\partial \hat{P}_{\underline{U}}} = 0$, and applying the constraint $\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) = 1$, we obtain

$$\Lambda(V_{U_{\tau^c}}, \lambda_1, \lambda_2) = -(-1)^{i_1}\lambda_1 \log \gamma_1 - (-1)^{i_2}\lambda_2 \log \gamma_2 \\ - (1+\rho) \log \left(\sum_{\underline{u}} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} P_{U_1}(u_1)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} P_{U_2}(u_2)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}} V_{U_{\tau^c}}(u_{\tau^c})^{\frac{\rho}{1+\rho}} \right), \tag{4.120}$$

where $\Lambda(V_{U_{\tau^c}}, \lambda_1, \lambda_2) = \max_\theta \min_{\hat{P}_{\underline{U}}} \Lambda(V_{U_{\tau^c}}, \hat{P}_{\underline{U}}, \theta, \lambda_1, \lambda_2)$. Inserting (4.120) into (4.119), we find

$$\min_{V_{U_{\tau^c}}} \min_{\substack{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \\ \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})} \\ = \min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} \Lambda(V_{U_{\tau^c}}, \lambda_1, \lambda_2). \tag{4.121}$$

We note that in (4.121), $V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}$ and $\lambda_\nu \in [0, +\infty)$ for $\nu = 1, 2$. Since $\mathcal{P}_{U_{\tau^c}}$ is a compact convex set, $[0, +\infty)$ is a convex set, $\Lambda(V_{U_{\tau^c}}, \lambda_1, \lambda_2)$ is concave on $[0, +\infty)$ and convex on $\mathcal{P}_{U_{\tau^c}}$, (4.121) satisfies the Sion's minimax theorem. Thus, we swap the maximization over λ_ν with minimization over $V_{U_{\tau^c}}$ which leads to

$$\min_{V_{U_{\tau^c}}} \min_{\substack{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \\ \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1), \hat{P}_{\underline{U}} \in \mathcal{B}_2^{i_2}(\gamma_2)}} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})} \\ = \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} \min_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} \Lambda(V_{U_{\tau^c}}, \lambda_1, \lambda_2). \tag{4.122}$$

Next, to solve the minimization over $V_{U_{\tau^c}}$ in the right hand side of (4.122), by inserting $\Lambda(\cdot)$ given by (4.120) into (4.122), we find that

$$\min_{V_{U_{\tau^c}}} \min_{\hat{P}_{\underline{U}} \in \mathcal{P}_{\underline{\mathcal{U}}}: \hat{P}_{\underline{U}} \in \mathcal{B}_1^{i_1}(\gamma_1) \cap \mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}} || P_{\underline{U}}) + \rho \sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log \frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})} = \\ \max_{\lambda_1 \geq 0, \lambda_2 \geq 0} -(1+\rho) \log \left(\max_{V_{U_{\tau^c}} \in \mathcal{P}_{U_{\tau^c}}} \sum_{\underline{u}} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_1}(u_1)}{\gamma_1} \right)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} \right. \\ \left. \times \left(\frac{P_{U_2}(u_2)}{\gamma_2} \right)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}} V_{U_{\tau^c}}(u_{\tau^c})^{\frac{\rho}{1+\rho}} \right), \tag{4.123}$$

where in (4.123), considering the fact that logarithm is an increasing function, we took the minimization over $V_{U_{\tau^c}}$ inside the logarithm.

Now, we can apply Lemma A.2 into the optimization problem given by the right hand side of (4.123). By defining

$$e(u_{\tau^c}) = \sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_1}(u_1)}{\gamma_1}\right)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} \left(\frac{P_{U_2}(u_2)}{\gamma_2}\right)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}}, \quad (4.124)$$

we let $V_Y(y) = V_{U_{\tau^c}}(u_{\tau^c})$ and $e(y) = e(u_{\tau^c})$ in Lemma A.2. Thus, the optimal $V_{U_{\tau^c}}$ is derived as

$$V_{U_{\tau^c}}(u_{\tau^c}) = \frac{\left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} P_{U_1}(u_1)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} P_{U_2}(u_2)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}}\right)^{1+\rho}}{\sum_{\bar{u}_{\tau^c}} \left(\sum_{\bar{u}_\tau} P_{\underline{U}}(\bar{u}_1,\bar{u}_2)^{\frac{1}{1+\rho}} P_{U_1}(\bar{u}_1)^{-\frac{(-1)^{i_1}\lambda_1}{1+\rho}} P_{U_2}(\bar{u}_2)^{-\frac{(-1)^{i_2}\lambda_2}{1+\rho}}\right)^{1+\rho}}. \quad (4.125)$$

In addition, in view of (A.14) in Lemma A.2, the optimization problem inside (4.123) is equal by $\left(\sum_{u_{\tau^c}} e(u_{\tau^c})^{1+\rho}\right)^{\frac{1}{1+\rho}}$, i. e.

$$\min_{V_{U_{\tau^c}}} \min_{\hat{P}_{\underline{U}}\in\mathcal{P}_{\underline{\mathcal{U}}}:\hat{P}_{\underline{U}}\in\mathcal{B}_1^{i_1}(\gamma_1)\cap\mathcal{B}_2^{i_2}(\gamma_2)} D(\hat{P}_{\underline{U}}||P_{\underline{U}}) + \rho\sum_{\underline{u}} \hat{P}_{\underline{U}}(\underline{u}) \log\frac{\hat{P}_{\underline{U}}(\underline{u})}{V_{U_{\tau^c}}(u_{\tau^c})} = -E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}) \quad (4.126)$$

where in (4.126), in view of $E_{s,\tau,i_1,i_2}(\cdot)$ is defined by (4.12), we used the fact that $\max_\lambda -f(\lambda) = -\min_\lambda f(\lambda)$. By replacing the left hand side of (4.117) with the left hand side of (4.126), we conclude the proof.

4.3.6 Proof of Proposition 4.3

We start by proving the following Lemma.

Lemma 4.3. *Let $E_0(\rho)$ be a continues function of ρ. Considering $E_{s,\tau,i_1,i_2}(\cdot)$ given by (4.12), for $\nu = 1, 2$, the function $\max_\rho E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=1}$ is non-decreasing with respect to γ_ν, and the function $\max_\rho E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=2}$ is non-increasing with respect to γ_ν.*

Proof. For $\nu = 1, 2$, from (4.9) and (4.10), we note that by letting $\gamma'_\nu > \gamma''_\nu$, we have $\mathcal{B}_\nu^1(\gamma'_\nu) \subseteq \mathcal{B}_\nu^1(\gamma''_\nu)$ and $\mathcal{B}_\nu^2(\gamma'_\nu) \supseteq \mathcal{B}_\nu^2(\gamma''_\nu)$. Thus, for all $\rho \in [0, 1]$ by letting

$i_\nu = 1$ in (4.11), we conclude that for γ'_ν the minimization problem of (4.11) is done over smaller set than for γ''_ν, which leads to $-E_{s,\tau,i_1,i_2}(\cdot,\gamma'_\nu)\Big|_{i_\nu=1} \geq -E_{s,\tau,i_1,i_2}(\cdot,\gamma''_\nu)\Big|_{i_\nu=1}$ for all values of $\rho \geq 0$. Similarly, for $i_\nu = 2$, since the minimization problem of (4.11) for γ'_ν is done over larger set than γ''_ν, we have $-E_{s,\tau,i_1,i_2}(\cdot,\gamma'_\nu)\Big|_{i_\nu=2} \leq -E_{s,\tau,i_1,i_2}(\cdot,\gamma''_\nu)\Big|_{i_\nu=2}$.

Hence, let ν^c be the complement index of $\nu \in \{1,2\}$ and $E_0(\rho)$ be a function of ρ. For a given γ_{ν^c}, we assume $\gamma'_\nu > \gamma''_\nu$. Thus, regardless the value of i_{ν^c}, the maximum of $E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot,\gamma'_\nu)\Big|_{i_\nu=1}$ is not smaller than $E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot,\gamma''_\nu)\Big|_{i_\nu=1}$ meaning that $\max_\rho E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=1}$ is non-decreasing with respect to γ_ν. The same reasoning concludes $\max_\rho E_0(\rho) - E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=2}$ is non-increasing with respect to γ_ν. □

Now, in view of (4.20), we define $F_{\tau,i_1,i_2}(\underline{\gamma})$ as

$$F_{\tau,i_1,i_2}(\underline{\gamma}) = \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i^c_\tau}) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma}), \tag{4.127}$$

where $f_{i_1,i_2}(\underline{\gamma}) = \min_\tau F_{\tau,i_1,i_2}(\underline{\gamma})$. We note that $F_{\tau,i_1,i_2}(\underline{\gamma})$ is of the form $\max_{\rho\in[0,1]} E_0(\rho) - E_{s,\tau,i_1,i_2}(\rho, P_{\underline{U}}, \underline{\gamma})$ in Lemma 4.3. In view of Lemma 4.3, $F_{\tau,1,i_2}$ and $F_{\tau,2,i_2}$ are respectively non-decreasing and non-increasing with respect to γ_1. Similarly, regardless the value of i_1, $F_{\tau,i_1,1}$ and $F_{\tau,i_1,2}$ are respectively non-decreasing and non-increasing with respect to γ_2.

Considering $f_{i_1,i_2}(\underline{\gamma}) = \min_\tau F_{\tau,i_1,i_2}(\underline{\gamma})$, by applying the fact that the minimum of monotonic functions is monotonic, $f_{i_1,i_2}(\underline{\gamma})$ defined by (4.20) is non-decreasing and non-increasing with respect to γ_ν, respectively when $i_\nu = 1$ and $i_\nu = 2$, for $\nu = 1, 2$.

Next, to find the optimal $\underline{\gamma}$ maximizing (4.21), we express E^{md} as

$$\max_{\gamma_1} \max_{\gamma_2} \min_{i_2} \min_{i_1} f_{i_1,i_2}(\underline{\gamma}), \tag{4.128}$$

where for a fixed γ_1, the optimization problem $\max_{\gamma_2} \min_{i_2} \min_{i_1} f_{i_1,i_2}(\underline{\gamma})$ satisfies Lemma A.8 with $\gamma = \gamma_2$, $i = i_2$, and $k_i(\gamma) = \min_{i_1} f_{i_1,i}(\gamma_1, \gamma)$. Therefore, the optimal $\gamma_2^\star$ satisfies

$$\min_{i_1=1,2} f_{i_1,1}(\gamma_1, \gamma_2^\star) = \min_{i_1=1,2} f_{i_1,2}(\gamma_1, \gamma_2^\star), \tag{4.129}$$

whenever (4.129) has solution. Otherwise, we have $\gamma_2^\star = 0$ when $f_{i_1,1}(\gamma_1, 0) > f_{i_1,2}(\gamma_1, 0)$, or $\gamma_2^\star = 1$ when $f_{i_1,1}(\gamma_1, 0) \leq f_{i_1,2}(\gamma_1, 0)$.

Now, applying $\gamma_2 = \gamma_2^\star$, the problem $\max_{\gamma_1} \min_{i_1} \min_{i_2} f_{i_1,i_2}(\gamma_1, \gamma_2^\star)$ satisfies Lemma A.8 with $\gamma = \gamma_1$, $i = i_1$, and $k_i(\gamma) = \min_{i_2} f_{i,i_2}(\gamma, \gamma_2^\star)$. Hence, $\gamma_1^\star$

maximizing (4.21) satisfies

$$\min_{i_2=1,2} f_{1,i_2}(\gamma_1^\star, \gamma_2^\star) = \min_{i_2=1,2} f_{2,i_2}(\gamma_1^\star, \gamma_2^\star), \tag{4.130}$$

and in the case (4.130) does not have solution, $\gamma_1^\star = 0$ when $f_{1,i_2}(0,\gamma_2) > f_{2,i_2}(0,\gamma_2)$, or $\gamma_1^\star = 1$ otherwise. Combining (4.129) and (4.130) we obtain (4.23).

4.3.7 Proof of Proposition 4.4

To prove Proposition 4.4, we use the properties of $E_{s,\tau,i_1,i_2}(\cdot)$ function. Like always, $\nu \in \{1,2\}$, and ν^c denotes the complement index of ν among the set $\{1,2\}$.

Let $\gamma_\nu \in (\max_{u_\nu} P_{U_\nu}(u_\nu), 1]$. In view of (4.9), regardless the value of i_{ν^c}, the minimization problem given by the left hand side of (4.11), is done over an empty set when $i_\nu = 1$ which leads to $E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=1} = -\infty$. However, for the case $\gamma_1 \in (\max_{u_1} P_{U_1}(u_1), 1]$ and $\gamma_2 \in (\max_{u_2} P_{U_2}(u_2), 1]$, if we have $i_1 = i_2 = 2$, by considering (4.10), the problem is simplified to a minimization problem without any constraint over distribution $\hat{P}_{\underline{U}}$, leading to $E_{s,\tau,2,2}(\rho, P_{\underline{U}}, \gamma_\nu) = E_{s,\tau}(\rho, P_{\underline{U}})$.

Similarly, when $\gamma_1 \in [0, \min_{u_1} P_{U_1}(u_1))$ and $\gamma_2 \in [0, \min_{u_2} P_{U_2}(u_2))$, if $i_1 = i_2 = 1$, (4.11) is simplified as a minimization problem without any constraint over distribution $\hat{P}_{\underline{U}}$ meaning that $E_{s,\tau,1,1}(\rho, P_{\underline{U}}, \gamma_\nu) = E_{s,\tau}(\rho, P_{\underline{U}})$. While, regardless the value of i_{ν^c}, for $\gamma_\nu \in [0, \min_{u_\nu} P_{U_\nu}(u_\nu))$, if $i_\nu = 2$, again the minimization is done over an empty set leading to $E_{s,\tau,i_1,i_2}(\cdot)\Big|_{i_\nu=2} = -\infty$. In our analysis, it suffices to consider $\gamma_\nu = 0$ or $\gamma_\nu = 1$ to represent the cases where $E_{s,\tau,i_1,i_2}(\cdot)$ is infinity. Let $\nu = 1,2$, same reasoning yields

$$E_{s,\tau,1,i_2}(\cdot)\Big|_{\gamma_1=1} = E_{s,\tau,2,i_2}(\cdot)\Big|_{\gamma_1=0} = E_{s,\tau,i_1,1}(\cdot)\Big|_{\gamma_2=1} = E_{s,\tau,i_1,2}(\cdot)\Big|_{\gamma_2=0} = -\infty, \tag{4.131}$$

and

$$\begin{cases} E_{s,\tau,1,1}(\cdot)\Big|_{\gamma_1=0,\gamma_2=0} = E_{s,\tau,1,2}(\cdot)\Big|_{\gamma_1=0,\gamma_2=1} = E_{s,\tau}(\rho, P_{\underline{U}}), \\ E_{s,\tau,2,1}(\cdot)\Big|_{\gamma_1=1,\gamma_2=0} = E_{s,\tau,2,2}(\cdot)\Big|_{\gamma_1=1,\gamma_2=1} = E_{s,\tau}(\rho, P_{\underline{U}}). \end{cases} \tag{4.132}$$

From (4.20), we conclude that for the cases given by (4.131) and (4.132), the function $f_{i_1,i_2}(\gamma_1,\gamma_2)$ is either infinity or is the Gallager exponent, i. e.

$$\min_{\tau\in\{\{1\},\{2\},\{1,2\}\}} \max_{\rho\in[0,1]} E_0(\rho, Q_{\tau,i_\tau}, WQ_{\tau^c,i_\tau^c}) - Es,\tau(\rho, P_{\underline{U}}). \tag{4.133}$$

For example, when $\gamma_1, \gamma_2 \in \{0,1\}$, the function $f_{i_1,i_2}(0,0)$ is equal to (4.133) when $i_1 = i_2 = 1$, and is infinity for the rest combinations of i_1 and i_2.

As a result, when $\gamma_1, \gamma_2 \in \{0,1\}$, from (4.131) and (4.132) we find that $f_{i_1,i_2}(\underline{\gamma})$ is finite in only one case, and it is infinity for other combinations of i_1 and i_2, more specificity

$$\min_{i_1,i_2=1,2} f_{i_1,i_2}(0,0) = f_{1,1}(0,0), \qquad \min_{i_1,i_2=1,2} f_{i_1,i_2}(0,1) = f_{1,2}(0,1), \tag{4.134}$$

$$\min_{i_1,i_2=1,2} f_{i_1,i_2}(1,0) = f_{2,1}(1,0), \qquad \min_{i_1,i_2=1,2} f_{i_1,i_2}(1,1) = f_{2,2}(1,1). \tag{4.135}$$

Next, by considering (4.134) and (4.134), we lower bound the achievable exponent given by (4.21). By taking maximization over $\gamma_\nu \in \{0,1\}$, rather than the interval of $[0,1]$, i. e.

$$E^{\mathrm{md}} \geq \max_{\gamma_1,\gamma_2 \in \{0,1\}} \min_{i_1,i_2=1,2} f_{i_1,i_2}(\gamma_1,\gamma_2), \tag{4.136}$$

we can find the following lower bound for E^{md}

$$E^{\mathrm{md}} \geq \max\Big\{ \min_{i_1,i_2=1,2} f_{i_1,i_2}(0,0), \min_{i_1,i_2=1,2} f_{i_1,i_2}(0,1), \\ \min_{i_1,i_2=1,2} f_{i_1,i_2}(1,0), \min_{i_1,i_2=1,2} f_{i_1,i_2}(1,1)\Big\}, \tag{4.137}$$

where by applying (4.134) and (4.135) into the minimizations over i_1 and i_2, we rewrite (4.137) as

$$E^{\mathrm{md}} \geq \max\left\{f_{1,1}(0,0), f_{1,2}(0,1), f_{2,1}(1,0), f_{2,2}(1,1)\right\}. \tag{4.138}$$

Inserting (4.133) into (4.138), we conclude (4.26).

4.3.8 Proof of (4.32)

From (4.12), we consider the following function

$$\min_{\lambda_\tau \geq 0, \eta} \log \sum_{u_{\tau^c}} \left(\frac{P_{U_{\tau^c}}(u_{\tau^c})}{\gamma_{\tau^c}}\right)^{\eta} \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(\frac{P_{U_\tau}(u_\tau)}{\gamma_\tau}\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}}\right)^{1+\rho}, \tag{4.139}$$

where the optimal $\eta^\star$ minimizing (4.139) is either positive, negative or zero. More precisely, by setting the first derivative of (4.139) with respect to η

equal to zero, $\eta^\star$ is obtained as the solution of following equation

$$\log(\gamma_{\tau^c}) = \frac{\sum_{u_{\tau^c}} P_{U_{\tau^c}}(u_{\tau^c})^{\eta^\star} \log\left(P_{U_{\tau^c}}(u_{\tau^c})\right) \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(P_{U_\tau}(u_\tau)\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}} \right)^{1+\rho}}{\sum_{u_{\tau^c}} P_{U_{\tau^c}}(u_{\tau^c})^{\eta^\star} \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(P_{U_\tau}(u_\tau)\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}} \right)^{1+\rho}}. \tag{4.140}$$

Let the optimal $\eta^\star$ satisfying (4.140), be positive. By comparing (4.139) with (4.12), we can easily conclude that (4.139) is equal to $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$. While, since $\eta^\star > 0$, in (4.12) the λ_{τ^c} minimizing $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ is negative, and hence the optimal $\lambda^\star_{\tau^c} = 0$, i. e. when $\eta\star > 0$, the function $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ equals to $E_{s,\tau,i_\tau}(\cdot)$ given by (4.33). Similarly, if $\eta^\star < 0$, $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ will equal to (4.33), while $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ is (4.139). Finally, when $\eta^\star = 0$, both $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ and $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ are equal to (4.33). Thus, $\eta^\star = 0$ plays a critical role in the behaviour of the $E_{s,\tau,i_1,i_2}(\cdot)$.

Given γ_τ, let for $\gamma_{\tau^c} = \gamma_\eta$ we have $\eta^\star = 0$ i. e.

$$\log(\gamma_\eta) = \frac{\sum_{u_{\tau^c}} \log\left(P_{U_{\tau^c}}(u_{\tau^c})\right) \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(P_{U_\tau}(u_\tau)\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}} \right)^{1+\rho}}{\sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} \left(P_{U_\tau}(u_\tau)\right)^{-\frac{(-1)^{i_\tau}\lambda_\tau}{1+\rho}} \right)^{1+\rho}}. \tag{4.141}$$

For $\gamma_{\tau^c} \in [0, \gamma_\eta)$, the optimal $\eta^\star$ maximizing (4.139) is negative, while for $\gamma_{\tau^c} \in (\gamma_\eta, 1]$ we have $\eta^\star > 0$. To be precisely, we recall from Lemma 4.3, $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ and $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ are respectively non-increasing and non-decreasing with respect to γ_{τ^c}. Moreover, in view of (4.131) and (4.132), we can conclude that the function $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ is constant on the interval $[0, \gamma_\eta)$ and equals to (4.33). While, by increasing γ_{τ^c} along the interval $[0, \gamma_\eta)$, the function $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ increases from $-\infty$ to (4.33). At $\gamma_{\tau^c} = \gamma_\eta$ both $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ and $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ are equal to (4.33). An then, by moving γ_{τ^c} along the interval $(\gamma_\eta, 1]$, the function $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=1}$ decreases from (4.33) to $-\infty$, while $E_{s,\tau,i_1,i_2}(\cdot)\big|_{i_{\tau^c}=2}$ equals to (4.33) and is constant on the interval $(\gamma_\eta, 1]$.

As a result, (4.31) can be rewritten as

$$\max\Bigg\{\max_{\substack{\gamma_{\tau^c}\in[0,\gamma_\eta),\\ \gamma_\tau\in[0,1]}}\min_{i_\tau,i_{\tau^c}=1,2}\max_{\rho\in[0,1]} E_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau})-E_{s,\tau,i_1,i_2}(\rho,P_{\underline{U}},\underline{\gamma}),$$
$$\max_{\substack{\gamma_{\tau^c}\in[\gamma_\eta,1],\\ \gamma_\tau\in[0,1]}}\min_{i_\tau,i_{\tau^c}=1,2}\max_{\rho\in[0,1]} E_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau})-E_{s,\tau,i_1,i_2}(\rho,P_{\underline{U}},\underline{\gamma})\Bigg\},\tag{4.142}$$

where as mentioned, in the first and second terms of (4.142) the minimization over i_{τ^c} is attained respectively at $i_{\tau^c}=1$ and $i_{\tau^c}=2$, both leading to

$$\max\Bigg\{\max_{\gamma_\tau\in[0,1]}\min_{i_\tau=1,2}\max_{\rho\in[0,1]} E_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau})-E_{s,\tau,i_\tau}(\rho,P_{\underline{U}},\gamma_\tau),$$
$$\max_{\gamma_\tau\in[0,1]}\min_{i_\tau=1,2}\max_{\rho\in[0,1]} E_0(\rho,Q_{\tau,i_\tau},WQ_{\tau^c,i^c_\tau})-E_{s,\tau,i_\tau}(\rho,P_{\underline{U}},\gamma_\tau)\Bigg\},\tag{4.143}$$

which concludes the proof.

4.3.9 Proof of Proposition 4.5

The proof of Theorem 4.3.9 is very similar to the presented proof in Section 4.3.2. To avoid repetition, here we just mention the main steps. Like before, initially we bound $\bar{\epsilon}^n$, the average error probability over the ensemble for a given block length n. By applying the random coding union bound [23] for joint source channel coding, and then grouping error events, the average error probability is bounded as

$$\bar{\epsilon}^n \le \sum_\tau \bar{\epsilon}^n_\tau,\tag{4.144}$$

where

$$\epsilon^n_\tau \le \sum_{\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y}} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})\min\Bigg\{1,\sum_{\substack{\hat{\boldsymbol{u}}_\tau\neq\boldsymbol{u}_\tau\,\hat{\boldsymbol{x}}_\tau:\\ \frac{P^n_{\underline{\boldsymbol{U}}}(\hat{\boldsymbol{u}}_\tau,\boldsymbol{u}_{\tau^c})W^n(\boldsymbol{y}|\hat{\boldsymbol{x}}_\tau,\boldsymbol{x}_{\tau^c})}{P^n_{\underline{\boldsymbol{U}}}(\underline{\boldsymbol{u}})W^n(\boldsymbol{y}|\boldsymbol{x}_\tau,\boldsymbol{x}_{\tau^c})}\ge 1}}\bar{Q}^n_{\tau,p,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau)\Bigg\}.\tag{4.145}$$

Like before, we group the outer and inner summations in (4.145) based on the empirical distributions of $(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})$ and $(\hat{\boldsymbol{u}}_\tau,\hat{\boldsymbol{x}}_\tau)$, respectively and then sum over all possible empirical distributions. Due to the fact that we study conditional constant-composition ensemble, the all possible empirical distributions of $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y})$ denoted by $\hat{P}_{\underline{U}\underline{X},Y}$ are restricted to the types of empirical distributions of $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y})$ whose distribution of x_ν given u_ν are the type-p conditional distribution defined in (4.41). It means that $\hat{P}_{\underline{U}\underline{X}Y}$ belongs to the set

$$\mathcal{S}_n^{\text{gccc}} \triangleq \Big\{ \hat{P}_{\underline{U}\underline{X}Y} : \hat{P}_{\underline{U}\underline{X}Y} = \hat{P}_{\underline{U}} \bar{Q}_{1,p,\hat{P}_{U_1}} \bar{Q}_{2,p,\hat{P}_{U_2}} \hat{P}_{Y|\underline{U}\underline{X}}, \\ \hat{P}_{\underline{U}} \in \mathcal{P}^n_{\underline{\mathcal{U}}}, \hat{P}_{Y|\underline{U}\underline{X}} \in \mathcal{P}^n_{\mathcal{Y}|\underline{\mathcal{U}}\times\underline{\mathcal{X}}} \Big\}, \tag{4.146}$$

where $\hat{P}_{U_\nu}$ is the marginal distribution of $\hat{P}_{\underline{U}}$. Moreover, the empirical distribution of the inner summation in (4.145) is denoted by $\tilde{P}_{\underline{U}\underline{X},Y}$ and is restricted to the set

$$\mathcal{L}_{\tau,n}^{\text{gccc}}(\hat{P}_{\underline{U}\underline{X}Y}) \triangleq \Bigg\{ \tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{S}_n^{\text{gccc}} : \tilde{P}_{U_{\tau^c}X_{\tau^c}Y} = \hat{P}_{U_{\tau^c}X_{\tau^c}Y}, \\ \mathbb{E}_{\tilde{P}} \lambda(\underline{U}, \underline{X}, Y) \geq \mathbb{E}_{\hat{P}} \lambda(\underline{U}, \underline{X}, Y) \Bigg\}. \tag{4.147}$$

Thus, equation (4.145) can be rewritten as

$$\bar{\epsilon}_\tau^n \leq \sum_{\hat{P}_{\underline{U}\underline{X}Y} \in \mathcal{S}_n^{\text{gccc}}} \sum_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \\ \min \Bigg\{ 1, \sum_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}_{\tau,n}^{\text{gccc}}(\hat{P}_{\underline{U}\underline{X}Y})} \sum_{(\hat{\boldsymbol{u}}_\tau\, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau,p,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau) \Bigg\}. \tag{4.148}$$

Since for all $(\hat{\boldsymbol{u}}_\tau\, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})$, distribution $\bar{Q}^n_{\tau,p,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau)$ has the same value, the inner sum of (4.148) is

$$\sum_{(\hat{\boldsymbol{u}}_\tau\, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau,p,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau) = |\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})| \bar{Q}^n_{\tau,p,\pi(\hat{\boldsymbol{u}}_\tau)}(\hat{\boldsymbol{x}}_\tau | \hat{\boldsymbol{u}}_\tau), \tag{4.149}$$

where $|\mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})|$ is defined in (1.40). Noting $(\hat{\boldsymbol{u}}_\tau\, \hat{\boldsymbol{x}}_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})$, we can write $\bar{Q}^n_{\tau,p,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau | \boldsymbol{u}'_\tau)$ as the number of occurrence of the symbols, i. e.

$$\bar{Q}^n_{\tau,p,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau | \boldsymbol{u}'_\tau) = e^{n \sum_{x_\tau, u_\tau} \tilde{P}_{U_\tau}(u_\tau) \bar{Q}_{\tau,p,\tilde{P}_{U_\tau}}(x_\tau|u_\tau) \log \bar{Q}_{\tau,p,\tilde{P}_{U_\tau}}(x_\tau|u_\tau)}. \tag{4.150}$$

As before, combining (4.150) and (4.70) into (4.149), we obtain

$$\sum_{(\boldsymbol{u}'_\tau, \boldsymbol{x}'_\tau) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})} \bar{Q}^n_{\tau,p,\pi(\boldsymbol{u}'_\tau)}(\boldsymbol{x}'_\tau | \boldsymbol{u}'_\tau) \leq \\ \max_{\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^{\text{cc}}_{\tau,n}(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y} || \tilde{P}_{U_\tau} \bar{Q}_{\tau,p,\tilde{P}_{U_\tau}} \hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau}) \right) + o(n)}, \tag{4.151}$$

where we bounded (4.151) by the maximum over the empirical probability distributions $\tilde{P}_{\underline{U}\underline{X}Y} \in \mathcal{L}^{\text{gccc}}_{\tau,n}(\hat{P}_{\underline{U}\underline{X}Y})$.

Additionally, to compute the summation over $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{T}^n_{\boldsymbol{u}_{\tau^c}\boldsymbol{x}_{\tau^c}\boldsymbol{y}}(\tilde{P}_{\underline{U}\underline{X}Y})$ in (4.148), we follow the same steps deriving (4.75). Thus, we have

$$\sum_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{x}},\boldsymbol{y})\in\mathcal{T}^n(\hat{P}_{\underline{U}\underline{X}Y})} P^n_{\underline{\boldsymbol{U}}\underline{\boldsymbol{X}}\boldsymbol{Y}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{x}}, \boldsymbol{y}) \leq \max_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{S}^{\text{gccc}}_n} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}W)\right)}. \tag{4.152}$$

Combining (4.151) and (4.152) into (4.148), taking the polynomial number of types into account and recognizing that a probability must be at most 1, the average error probability is bounded by

$$\begin{aligned}\bar{\epsilon}^n_\tau \leq & \max_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{S}^{\text{gccc}}_n} e^{-n\left(D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}W)\right)+o(n)} \\ & \min\left\{1, \max_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}^{\text{gccc}}_{\tau,n}(\hat{P}_{\underline{U}\underline{X}Y})} e^{-n\left(D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,p,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y})-H(\tilde{P}_{U_\tau})\right)+o(n)}\right\}.\end{aligned} \tag{4.153}$$

Like before, by taking logarithm from both sides of (4.153), using the identity $\min\{1, e^a\} = e^{[a]^+}$, we find that

$$-\frac{1}{n}\log(\bar{\epsilon}^n_\tau) \geq E^n_\tau - \frac{o(n)}{n}, \tag{4.154}$$

where

$$\begin{aligned}E^n_\tau = & \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{S}^{\text{gccc}}_n} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}W)+ \\ & \left[\min_{\tilde{P}_{\underline{U}\underline{X}Y}\in\mathcal{L}^{\text{gccc}}_{\tau,n}(\hat{P}_{\underline{U}\underline{X}Y})} D(\tilde{P}_{\underline{U}\underline{X}Y}||\tilde{P}_{U_\tau}\bar{Q}_{\tau,\tilde{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\tilde{P}_{U_\tau})\right].\end{aligned} \tag{4.155}$$

Now, by setting $U = U_\tau$, $X = X_\tau$, $W = P_{U_{\tau^c}}\bar{Q}\tau^c, p, \hat{P}_{U_{\tau^c}}W$ and $Y = U_{\tau^c}X_{\tau^c}Y$ in Lemma A.1, (4.155) is bounded as

$$\begin{aligned}E^n_\tau \geq & \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{S}^{\text{gccc}}_n} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}W)+ \\ & \left[D(\hat{P}_{\underline{U}\underline{X}Y}||\hat{P}_{U_\tau}\bar{Q}_{\tau,p,\hat{P}_{U_\tau}}\hat{P}_{U_{\tau^c}X_{\tau^c}Y}) - H(\hat{P}_{U_\tau})\right].\end{aligned} \tag{4.156}$$

Recalling (4.147), we have $\hat{P}_{U_{\tau^c}X_{\tau^c}Y} = \hat{P}_{U_{\tau^c}}\bar{Q}_{\tau^c,p,\hat{P}_{\underline{U}_{\tau}^c}}\hat{P}_{Y|U_{\tau^c}X_{\tau^c}}$ and thanks to (4.146), $\hat{P}_{\underline{U}\underline{X}Y} = \hat{P}_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}\hat{P}_{Y|\underline{X}\underline{U}}$. Applying these facts to (4.156), (4.154) is bounded as

$$-\frac{1}{n}\log(\bar{\epsilon}_{\tau}^{n}) \geq -\frac{o(n)}{n} + \min_{\hat{P}_{\underline{U}\underline{X}Y}\in\mathcal{S}_{n}^{\mathrm{gccc}}} D(\hat{P}_{\underline{U}\underline{X}Y}||P_{\underline{U}}\bar{Q}_{1,p,\hat{P}_{U_1}}\bar{Q}_{2,p,\hat{P}_{U_2}}W)$$

$$+\left[I\left(X;Y\middle|U_{\tau^c},X_{\tau^c}\right) - H(\hat{P}_{U_{\tau}|U_{\tau^c}})\right]^{+}. \tag{4.157}$$

Since both sides of (4.157) are bounded, by taking lim inf from both sides, and then using the fact that $\cup_n\mathcal{P}^{n}_{\underline{\mathcal{U}}\times\underline{\mathcal{X}}\times\mathcal{Y}}$ is dense in the set of all distributions, we conclude the proof.

Chapter 5

Conclusions and Future Work

In this thesis, for the point-to-point and the multiple-access channels, we studied a number of random coding ensembles reviewed in the following.

1. **The iid ensemble**: for this coding scheme, codewords of each user are generated independently according to an identical input distribution.

2. **The icd ensemble**: for this coding scheme, codewords of each user are generated independently according to a conditional probability distribution.

3. **The message-dependent ensemble**: for this coding scheme, the source outputs of each user are partitioned into countable classes and are encoded by the codes that depend on the class index.

4. **The constant-composition ensemble**, in which for each user, codewords are drawn from the set of sequences with a given empirical distribution.

5. **The conditional constant-composition ensemble**: for this coding scheme, codewords are drawn from the conditional constant-composition sequences. In other words, codewords are generated such that for the message $\boldsymbol{u}_\nu$ and its corresponding codeword $\boldsymbol{x}_\nu$, the ratio of the number of occurrences of joint symbols $(u_\nu, x_\nu) \in \mathcal{U}_\nu \times \mathcal{X}_\nu$ in the joint sequences $(\boldsymbol{u}_\nu, \boldsymbol{x}_\nu)$ to the number of occurrences of u_ν in $\boldsymbol{u}_\nu$, remains constant.

6. **The generalized constant-composition ensemble**, in which the ideas of the message-dependent and the conditional constant-composition ensembles are merged. For each user, the source messages are assigned into disjoint classes, and codewords are drawn from the set of sequences with a given empirical distribution that depends on the class index.

7. **The generalized conditional constant-composition ensemble**, in which the ideas of the message-dependent and the conditional constant-composition ensembles are merged.

We found that for single-user communication and also for the MAC with independent sources, among the studied ensembles, the generalized constant-composition ensemble has the largest exponent. Thus, for the MAC with independent sources, in terms of error exponent, there is no benefit to apply statistical dependency between the messages and codewords. On the other hand, the results show that for the MAC with correlated sources, considering statistical dependency between the messages and codewords is essential.

In this thesis, we mainly focused on the messaged-dependent ensemble, where to generate codewords, we assign a set of input distributions rather than one input distribution. Hence, the optimal number of input distributions are another problem to answer. For single-user communication, in Proposition 2.7, we showed that two input distributions are sufficient for the message-dependent random-coding exponent. However, for the MAC with independent sources, we could only show that for the error type $\tau = \{1\}$ or $\tau = \{2\}$, the sufficient number of input distributions for each user is two. Unfortunately, the proof in Section 2.4.6 cannot be easily generalized to the error type $\tau = \{1, 2\}$. On the other hand, we could not find an example showing that assigning more than two input distributions has benefit for the exponent. As a result, the sufficient number of input distributions is for the message-dependent exponent is still an open problem.

To generalize the results to continuous alphabet, we may use the idea of cost-constrained random-coding ensemble [5, Ch. 7]. In fact, for user $\nu = 1, 2$, we can consider the cost-constrained ensemble characterized by the following conditional distribution

$$P_{\boldsymbol{X}_\nu|\boldsymbol{U}_\nu}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \frac{1}{\mu_{\nu,n}} \prod_{j=1}^{n} \bar{Q}_\nu(x_{j,\nu}|u_{j,\nu}) \mathbb{1}\Big\{\boldsymbol{x}_\nu \in \mathcal{E}_n^\nu(\boldsymbol{u}_\nu)\Big\}, \tag{5.1}$$

where

$$\mathcal{E}_n^\nu(\boldsymbol{u}) \triangleq \Big\{\boldsymbol{x}_\nu : \Big|a_{l_\nu}^n(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu) - \phi_{l_\nu}(\boldsymbol{u}_\nu)\Big| \le \delta_\nu\Big\}, \tag{5.2}$$

and $\mu_{\nu,n}$ is a normalizing constant, δ_ν is a positive constant and $\phi_{l_\nu}(\boldsymbol{u}_\nu) = \sum_{\boldsymbol{x}_\nu} \bar{Q}_\nu^n(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) a_{l_\nu}^n(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu)$. In addition, for each $l_\nu = 1, 2, ..., L_\nu$, the function $a_{l,\nu}(x_\nu, u_\nu)$ is a real-valued function known as a cost function where $a_{l_\nu}^n(\boldsymbol{x}, \boldsymbol{u}) = \sum_j^n a_{l_\nu}(x_{j,\nu}, u_{j,\nu})$.

As discussed in [5, p. 324], the ensemble of codes whose all codewords satisfies the constraint is included in the class of codes for which the average over the codewords satisfies the constraint. Thus, to find the achievable exponent, by noting that $a^n_{l_\nu}(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu) = \sum_j^n a_{l_\nu}(x_{j,\nu}, u_{j,\nu})$, we may simplify $\phi_{l_\nu}(\boldsymbol{u}_\nu)$ as

$$\phi_{l_\nu}(\boldsymbol{u}_\nu) = \sum_{(x_{1,\nu},...,x_{n,\nu})} \prod_{j_\nu=1}^{n} \bar{Q}_\nu(x_{j,\nu}|u_{j,\nu})\Big(a_{l_\nu}(x_{1,\nu}, u_{1,\nu}) + ... + a_{l_\nu}(x_{n,\nu}, u_{n,\nu})\Big) \tag{5.3}$$

$$= \sum_{x_{1,\nu}} a_l(x_{1,\nu}, u_{1,\nu})\bar{Q}(x_{1,\nu}|u_{1,\nu}) \sum_{(x_{2,\nu},...,x_{n,\nu})} \bar{Q}^{n-1}_\nu(x_{2,\nu}...x_{n,\nu}|u_{2,\nu}...u_{n,\nu}) + ...$$
$$+ \sum_{x_{n,\nu}} a_{l_\nu}(x_{n,\nu}, u_{n,\nu})\bar{Q}_\nu(x_{n,\nu}|u_{n,\nu}) \sum_{(x_{1,\nu},...,x_{n-1,\nu})} \bar{Q}^{n-1}_\nu(x_{1,\nu}, ..., x_{n-1}|u_{1,\nu}, ..., u_{n-1,\nu}) \tag{5.4}$$

$$= \sum_j \sum_{x_{j,\nu}} a(x_{j,\nu}, u_{j,\nu})\bar{Q}_\nu(x_{j,\nu}|u_{j,\nu}) = \sum_j \phi_{l_\nu}(u_{j,\nu}) = n\phi_{l_\nu}(u_\nu), \tag{5.5}$$

where in (5.3), we used the fact that $a^n_{l_\nu}(\boldsymbol{x}_\nu, \boldsymbol{u}_\nu) = \sum_j^n a_{l_\nu}(x_{j,\nu}, u_{j,\nu})$, in (5.4) we broke the summation over $(x_{1,\nu}, ..., x_{n,\nu})$ into summation over $x_{i,\nu}$ and summation over $(x_{1,\nu}, ..., x_{i-1,\nu}, x_{i+1,\nu}, ..., x_{n,\nu})$ and in (5.5), we used the identities that $\sum_{\boldsymbol{x}_\nu} \bar{Q}^{n-1}_\nu(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = 1$ and $\phi_{l_\nu}(u_\nu) = \mathbb{E}_{\bar{Q}_\nu}[a_{l_\nu}(x_\nu, u_\nu)]$. For the cost-constrained random coding, it may be proved that the following exponent is achievable

$$\min_\tau \max_{\rho\in[0,1]} \sup_{\{r_{l_1}, r_{l_2}\}} -\log\Bigg(\sum_{y,x_{\tau^c},u_{\tau^c}}\Bigg(\sum_{u_\tau,x_\tau} P_{\underline{U}}(\underline{u})^{\frac{1}{1+\rho}} e^{\sum_{l_1}\bar{r}_{l_1}\left(a_{l_1}(x_1,u_1)-\phi_{l_1}(u_1)\right)}$$
$$\times e^{\sum_{l_2}\bar{r}_{l_2}\left(a_{l_2}(x_2,u_2)-\phi_{l_2}(u_2)\right)}\bar{Q}_\tau(x_\tau|u_\tau)W(y|\underline{x})^{\frac{1}{1+\rho}}\bar{Q}_{\tau^c}(x_{\tau^c}|u_{\tau^c})^{\frac{1}{1+\rho}}\Bigg)^{1+\rho}\Bigg). \tag{5.6}$$

We may compare the cost-constrained exponent given by (5.6) with conditional constant-composition and icd exponents. Discussion about (5.6) and its application for continuous alphabets are left for future works.

In this thesis, we only derived achievable exponents for the MAC by using the idea of random coding. In Section 4.3.3, we showed that the primal form of the achievable exponent given by (4.2) is ensemble tightness. However, we do not find an upper bound for the error exponent satisfied by any code. Generally, finding converse bounds for the MAC is more difficult than single-user communication. It may be simpler to start by separate source-channel coding, for the MAC with independent sources. In [31], by using Csiszár's techniques [15], sphere-packing and minimum-distance exponents were found.

Hypothesis-testing method can be another approach to find converse bound. However, since here we have three types of error, the test should be chosen more carefully. On the other hand, due to [43], the performance of the maximal error and the average error for the MAC are not as easy as the point-to-point channel.

Appendix A

General Lemmas

In this Appendix, we provide a number of general lemmas that will be used through the thesis. Throughout this Appendix, we consider a discrete memoryless source characterized by a probability distribution P_U.

Lemma A.1. *Assume E^n be*

$$E^n = \min_{\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W) + \Big[\min_{\tilde{P}_{UXY} \in \mathcal{K}^n(\hat{P}_{UXY})} D(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y) - H(\tilde{P}_U)\Big]^+, \qquad (A.1)$$

where

$$\mathcal{K}^n(\hat{P}_{UXY}) \triangleq \Big\{\tilde{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}} : \tilde{P}_Y = \hat{P}_Y, \mathbb{E}_{\tilde{P}}[\lambda(U,X,Y)] \geq \mathbb{E}_{\hat{P}}[\lambda(U,X,Y)]\Big\}, \qquad (A.2)$$

and $\lambda(U,X,Y) = P_U(U)W(Y|X)$. It can be proved that

$$E^n \geq \min_{\hat{P}_{UXY} \in \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W) + \Big[D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y) - H(\hat{P}_U)\Big]^+. \qquad (A.3)$$

Proof. To prove (A.3), firstly we assume

$$D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U) \geq D\left(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y\right) - H(\hat{P}_U), \qquad (A.4)$$

which leads to

$$\Big[D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U)\Big]^+ \geq \Big[D\left(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y\right) - H(\hat{P}_U)\Big]^+. \qquad (A.5)$$

By adding $D\left(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W\right)$ on the both sides of (A.5), (A.3) will be proved. In the alternative case

$$D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U) \le D\left(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y\right) - H(\hat{P}_U), \quad \text{(A.6)}$$

by noting (A.2), we have $\mathbb{E}_{\tilde{P}}\lambda(U,X,Y) \ge \mathbb{E}_{\hat{P}}\lambda(U,X,Y)$ or

$$\sum_{u,x,y}\tilde{P}_{UXY}(u,x,y)\log\left(P_U(u)W(y|x)\right) \ge \sum_{u,x,y}\hat{P}_{UXY}(u,x,y)\log\left(P_U(u)W(y|x)\right). \quad \text{(A.7)}$$

Subtracting (A.7) from (A.6), we obtain

$$\sum_{u,x,y}\tilde{P}_{UXY}(u,x,y)\log\left(\frac{\tilde{P}_{UXY}(u,x,y)}{P_U(u)W(y|x)\bar{Q}_{\tilde{P}_U}(x|u)\hat{P}_Y(y)}\right) \le \sum_{u,x,y}\hat{P}_{UXY}(u,x,y)\log\left(\frac{\hat{P}_{UXY}(u,x,y)}{P_U(u)W(y|x)\bar{Q}_{\hat{P}_U}(x|u)\hat{P}_Y(y)}\right). \quad \text{(A.8)}$$

Moreover, in view of (A.2), $\tilde{P}_Y = \hat{P}_Y$ which leads to the fact that $H(\tilde{P}_Y) = H(\hat{P}_Y)$ or equivalently

$$\sum_{u,x,y}\tilde{P}_{UXY}(u,x,y)\log\hat{P}_Y(y) = \sum_{u,x,y}\hat{P}_{UXY}(u,x,y)\log\hat{P}_Y(y). \quad \text{(A.9)}$$

By adding (A.9) to the both sides of (A.8), we obtain

$$\sum_{u,x,y}\tilde{P}_{UXY}(u,x,y)\log\left(\frac{\tilde{P}_{UXY}(u,x,y)}{P_U(u)W(y|x)\bar{Q}_{\tilde{P}_U}(x|u)}\right) \le \sum_{u,x,y}\hat{P}_{UXY}(u,x,y)\log\left(\frac{\hat{P}_{UXY}(u,x,y)}{P_U(u)W(y|x)\bar{Q}_{\hat{P}_U}(x|u)}\right). \quad \text{(A.10)}$$

Noting to the definition of the relative entropy, (A.10) can be expressed as

$$D(\tilde{P}_{UXY}||P_U\bar{Q}_{\tilde{P}_U}W) \le D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W). \quad \text{(A.11)}$$

By adding $\left[D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U)\right]^+$ on the both sides of (A.11), we obtain

$$D(\tilde{P}_{UXY}||P_U\bar{Q}_{\tilde{P}_U}W) + \left[D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U)\right]^+ \le D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W) + \left[D\left(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\right) - H(\tilde{P}_U)\right]^+. \quad \text{(A.12)}$$

Since $\mathcal{K}^n(\hat{P}_{UXY}) \subset \mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}$, we proved that whether $D\Big(\tilde{P}_{UXY}||\tilde{P}_U\bar{Q}_{\tilde{P}_U}\hat{P}_Y\Big) - H(\tilde{P}_U)$ be lower than $D\left(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y\right) - H(\hat{P}_U)$ or greater, we have

$$E^n \geq \min_{\hat{P}_{UXY}\in\mathcal{P}^n_{\mathcal{U}\times\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{UXY}||P_U\bar{Q}_{\hat{P}_U}W) + \Big[D(\hat{P}_{UXY}||\hat{P}_U\bar{Q}_{\hat{P}_U}\hat{P}_Y) - H(\hat{P}_U\Big]^+. \qquad \text{(A.13)}$$

□

Lemma A.2. *Let V_Y be a probability distribution and $e(y)$ be a positive function such that for $\rho \in [0,1]$, the quantity $\sum_y e(y)V_Y(y)^{\frac{\rho}{1+\rho}}$ is a concave function of $V_Y(y)$. Then, we have*

$$\max_{V_Y} \sum_y e(y)V_Y(y)^{\frac{\rho}{1+\rho}} = \left(\sum_y e(y)^{1+\rho}\right)^{\frac{1}{1+\rho}}, \qquad \text{(A.14)}$$

where the optimal V_Y maximizing (A.14) is obtained as $V_Y(y) = \dfrac{e(y)^{1+\rho}}{\sum_{\bar{y}} e(\bar{y})^{1+\rho}}$.

Proof. Recalling that $\sum_y V_Y(y) = 1$, the Lagrangian associated with the optimization problem in (A.14) can be written as

$$\Lambda(V_Y,\theta) = \sum_y e(y)V_Y(y)^{\frac{\rho}{1+\rho}} + \theta(1 - \sum_y V_Y(y)). \qquad \text{(A.15)}$$

In view of the KKT condition, setting the partial derivative of $\Lambda(V_Y,\theta)$ with respect to $V_Y(y)$ equal to zero, yields

$$\frac{\rho}{1+\rho} e(y)V_Y(y)^{\frac{-1}{1+\rho}} - \theta = 0. \qquad \text{(A.16)}$$

Solving (A.16) with respect to $V_Y(y)$, applying the constraint that $\sum_y V_Y(y) = 1$, the optimal value of $V_Y(y)$ is derived as $V_Y(y) = \dfrac{e(y)^{1+\rho}}{\sum_{\bar{y}} e(\bar{y})^{1+\rho}}$. Inserting the optimal $V_Y(y)$ into the left hand side of (A.14) proves Lemma A.2. □

Lemma A.3. *For a given channel W with input distribution Q, we have*

$$\min_{\hat{P}_{XY}\in\mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||Q\hat{P}_Y) = E_0(\rho,Q,W), \qquad \text{(A.17)}$$

where $E_0(\rho,Q,W) = -\log\Big(\sum_y \Big(\sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\Big)$.

Proof. Firstly, we show that

$$D(\hat{P}_{XY}||Q\hat{P}_Y) = \min_{V_Y} D(\hat{P}_{XY}||QV_Y), \tag{A.18}$$

where V_Y is an arbitrary probability assignment over the alphabet $\mathcal{Y}$. To prove (A.18), we use Lemma A.4. Setting $Z = X$ and $P_Z = Q$ in Lemma A.4, (A.18) will prove.

Next, by substituting (A.18) into the left hand side of (A.17), it remains to show that

$$\min_{V_Y} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) = E_0(\rho, Q, W). \tag{A.19}$$

In order to prove (A.19), we start by applying Lagrange duality theory to the inner minimization over $\hat{P}_{XY}$ in (A.19). By recognizing that the sum of the probabilities of all possible outcomes must be 1, the Lagrangian of optimization problem over $\hat{P}_{XY}$ can be expressed as

$$\Lambda(\hat{P}_{XY}, \theta) = D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) + \theta\Big(1 - \sum_{x,y} \hat{P}_{XY}(x,y)\Big), \tag{A.20}$$

where since the objective function is convex with respect to $\hat{P}_{XY}$ and the constraint $\sum_{x,y} \hat{P}_{XY}(x,y) = 1$ is affine, strong duality holds which leads to

$$\min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) = \max_{\theta} \min_{\hat{P}_{XY}} \Lambda(\hat{P}_{XY}, \theta). \tag{A.21}$$

Using the definition of the relative entropy, the Lagrangian is simplified as

$$\Lambda(\hat{P}_{XY}, \theta) = \sum_{x,y} \hat{P}_{XY}(x,y) \log \frac{\hat{P}_{XY}(x,y)^{1+\rho}}{Q(x)^{1+\rho} W(y|x) V_Y(y)^{\rho}} + \theta\Big(1 - \sum_{x,y} \hat{P}_{XY}(x,y)\Big). \tag{A.22}$$

Since strong duality holds, we can proceed by analyzing the necessary KKT conditions. Setting $\frac{\partial \Lambda(\hat{P}_{XY})}{\partial \hat{P}_{XY}(x,y)} = 0$ yields

$$\log \frac{\hat{P}_{XY}(x,y)^{1+\rho}}{Q(x)^{1+\rho} W(y|x) V_Y(y)^{\rho}} + (1+\rho) - \theta = 0, \tag{A.23}$$

leading to

$$\hat{P}_{XY}(x,y) = e^{\frac{\theta-(1+\rho)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}. \tag{A.24}$$

Summing both sides of (A.24) over x, y and applying $\sum_{x,y} \hat{P}_{XY}(x,y) = 1$, we obtain

$$1 = e^{\frac{\theta-(1+\rho)}{1+\rho}} \sum_{x,y} Q(x)W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}. \tag{A.25}$$

Putting back $e^{\frac{\theta-(1+\rho)}{1+\rho}}$ obtained in (A.25) into (A.24), the optimal $\hat{P}_{XY}$ is given by

$$\hat{P}_{XY}(x,y) = \frac{Q(x)W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}}{\sum_{\bar{x},\bar{y}} Q(\bar{x})W(\bar{y}|\bar{x})^{\frac{1}{1+\rho}} V_Y(\bar{y})^{\frac{\rho}{1+\rho}}}. \tag{A.26}$$

Substituting (A.26) into (A.22), yields

$$\max_{\theta} \min_{\hat{P}_{XY}} \Lambda(\hat{P}_{XY}, \theta) = -(1+\rho)\log\left(\sum_{x,y} Q(x)W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \tag{A.27}$$

where by putting back (A.27) into (A.21), (A.19) can be written as

$$\begin{aligned}\min_{V_Y} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} & D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) \\ &= \min_{V_Y} -(1+\rho)\log\left(\sum_{x,y} Q(x)W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right). \end{aligned} \tag{A.28}$$

Since the function in the log term of (A.28) is a concave function with respect to V_Y and the logarithm is an increasing function, (A.28) can be simplified as

$$\begin{aligned}\min_{V_Y} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} & D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) \\ &= -(1+\rho)\log\left(\max_{V_Y} \sum_{x,y} Q(x)W(y|x)^{\frac{1}{1+\rho}} V_Y(y)^{\frac{\rho}{1+\rho}}\right), \end{aligned} \tag{A.29}$$

where the optimization problem of the right hand side of (A.29) is solved by using Lemma A.2.

Setting $e(y) = \sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}$ in Lemma A.2, from (A.29) we obtain

$$\begin{aligned}\min_{V_Y} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} & D(\hat{P}_{XY}||QW) + \rho D(\hat{P}_{XY}||QV_Y) \\ &= -(1+\rho)\log\left(\sum_y \left(\sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right)^{\frac{1}{1+\rho}}. \end{aligned} \tag{A.30}$$

Applying the identity that $(1+\rho)\log(b^{\frac{1}{1+\rho}}) = \log(b)$ to the right hand side of (A.30), in view of the definition of $E_0(\cdot)$ and (A.19), we conclude (A.17). $\square$

Lemma A.4. *Suppose P_Z be a probability distribution and let $\hat{P}_{ZY}$ be an arbitrary joint distribution where $\hat{P}_Y$ is its marginal distribution. It can be proved that*

$$D(\hat{P}_{ZY}||P_Z\hat{P}_Y) = \min_{V_Y} D(\hat{P}_{ZY}||P_Z V_Y), \tag{A.31}$$

where V_Y is an arbitrary probability assignment over the alphabet $\mathcal{Y}$.

Proof. It suffices to show that $D(\hat{P}_{ZY}||P_Z\hat{P}_Y) \leq D(\hat{P}_{ZY}||P_Z V_Y)$ with equality if $\hat{P}_Y(y) = V_Y(y)$ for all y. Subtracting $D(\hat{P}_{ZY}||P_Z V_Y)$ from $D(\hat{P}_{ZY}||P_Z\hat{P}_Y)$ leads to

$$\begin{aligned} D(\hat{P}_{ZY}||P_Z\hat{P}_Y) - D(\hat{P}_{ZY}||P_Z V_Y) &= \\ \sum_{z,y} \hat{P}_{ZY}(z,y) \log \frac{V_Y(y)}{\hat{P}_Y(y)} &= -D(V_Y||\hat{P}_y) \leq 0, \end{aligned} \tag{A.32}$$

where (A.32) follows from the fact that the relative entropy is non-negative with equality when $V_Y(y) = \sum_z \hat{P}_{ZY}(z,y)$ for all $y \in \mathcal{Y}$. Thus, from (A.32) we conclude that $D(\hat{P}_{ZY}||P_Z\hat{P}_Y) \leq D(\hat{P}_{XY}||P_Z V_Y)$ and equality holds if $V_Y(y) = \hat{P}_Y(y)$ for all $y \in \mathcal{Y}$. As a result, $D(\hat{P}_{ZY}||P_Z\hat{P}_Y) = \min_{V_Y} D(\hat{P}_{ZY}||P_Z V_Y)$. □

Lemma A.5. *The function $\frac{\sum_u P_U(u)^{\frac{1}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1}{1+\rho}}}$ is continuous, non-decreasing and one-to-one with respect to $\frac{1}{1+\rho}$. Thus, for a given $\gamma_\ell \in [0,1]$ for $\ell = 1, ..., L$, let $\frac{1}{1+\rho^*_{\gamma_\ell}}$ be the solution of the following equation*

$$\frac{\sum_u P_U(u)^{\frac{1}{1+\rho^*_{\gamma_\ell}}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1}{1+\rho^*_{\gamma_\ell}}}} = \log(\gamma_\ell). \tag{A.33}$$

Since $\frac{\sum_u P_U(u)^{\frac{1}{1+\rho}} \log P_U(u)}{\sum_u P_U(u)^{\frac{1}{1+\rho}}}$ is non-decreasing with respect to $\frac{1}{1+\rho}$, we have

$$\gamma_{\ell+1} \leq \gamma_\ell \Rightarrow \frac{1}{1+\rho^*_{\gamma_{\ell+1}}} \leq \frac{1}{1+\rho^*_{\gamma_\ell}}. \tag{A.34}$$

Lemma A.6. *Let $E(\rho, Q_1)$ and $E(\rho, Q_2)$ be two concave and continuous functions of ρ. Consider the following optimization problem*

$$\max_{\gamma\in[0,1]} \min_{i=1,2} \max_{\rho\in[0,1]} E(\rho, Q_i) - E_{s,i}(\rho, P_U, \gamma), \tag{A.35}$$

where for $i = 1, 2$, $E_{s,1}(\cdot)$ and $E_{s,2}(\cdot)$ are given by (2.33) and (2.34). The optimal $\gamma^\star$ maximizing (A.35), satisfies

$$\max_{\rho\in[0,1]} E(\rho, Q_1) - E_{s,1}(\rho, P_U, \gamma^\star) = \max_{\rho\in[0,1]} E(\rho, Q_2) - E_{s,2}(\rho, P_U, \gamma^\star). \quad \text{(A.36)}$$

When (A.36) has no solutions, the optimal $\gamma^\star$ is either zero or one. In fact, if $\max_{\rho\in[0,1]} E(\rho, Q_1) - E_{s,1}(\rho, P_U, 0) > \max_{\rho\in[0,1]} E(\rho, Q_2) - E_{s,2}(\rho, P_U, 0)$, then $\gamma^\star = 0$, otherwise $\gamma^\star = 1$. Moreover, for the expression given by (A.35), we have

$$\max_{\gamma\in[0,1]} \min_{i=1,2} \max_{\rho} E(\rho, Q_i) - E_{s,i}(\rho, P_U, \gamma) = \max_{\rho\in[0,1]} \bar{E}(\rho) - E_s(\rho, P_U), \quad \text{(A.37)}$$

where $E_s(\cdot)$ is given by (1.9), and we have

$$\bar{E}(\rho) = \sup_{\substack{\rho_1,\rho_2,\lambda\in[0,1]:\\ \lambda\rho_1+(1-\lambda)\rho_2=\rho}} \left\{\lambda \max_{i=1,2} E(\rho_1, Q_i) + (1-\lambda) \max_{i=1,2} E(\rho_2, Q_i)\right\}. \quad \text{(A.38)}$$

Proof. From Lemma A.7, we conclude that with respect to γ, both functions $\max_\rho E(\rho, Q_1) - E_{s,1}(\rho, P_U, \gamma)$ and $\max_\rho E(\rho, Q_2) - E_{s,2}(\rho, P_U, \gamma)$ are non-decreasing and non-increasing, respectively. Thus, (A.35) satisfies Lemma A.8 with $k_1(\gamma) = \max_\rho E(\rho, Q_1) - E_{s,1}(\rho, P_U, \gamma)$ and $k_2(\gamma) = \max_\rho E(\rho, Q_2) - E_{s,2}(\rho, P_U, \gamma)$. Therefore, the optimal $\gamma^\star$ satisfies (A.36).

In order to prove (A.37), without loss of generality, we write (A.35) as

$$\max_{\gamma\in[0,1]} \min_{i=1,2} \max_{\rho_i\in[0,1]} E(\rho_i, Q_i) - E_{s,i}(\rho_i, P_U, \gamma), \quad \text{(A.39)}$$

where the optimal $\gamma^\star$ is obtained by solving (A.36). In addition, corresponding to $\gamma^\star$, let $\rho_{\gamma^\star}$, given by (2.25), be the tangent point to $E_s(\cdot)$ function. Letting $\gamma^\star$ as the optimal value maximizing (A.39), we define $\rho_1^\star$ and $\rho_2^\star$ as follows

$$\rho_1^\star = \arg\max_{\rho_1\in[0,1]} E(\rho_1, Q_1) - E_{s,1}(\rho_1, P_U, \gamma^\star), \quad \text{(A.40)}$$

$$\rho_2^\star = \arg\max_{\rho_2\in[0,1]} E(\rho_2, Q_2) - E_{s,2}(\rho_2, P_U, \gamma^\star), \quad \text{(A.41)}$$

where $\gamma^\star$ is derived by solving (A.36).

Therefore, (A.36) can be written as

$$E(\rho_1^\star, Q_1) - E_{s,1}(\rho_1^\star, P_U, \gamma^\star) = E(\rho_2^\star, Q_2) - E_{s,2}(\rho_2^\star, P_U, \gamma^\star), \quad \text{(A.42)}$$

where without loss of generality we assume $\rho_1^\star < \rho_{\gamma^\star} < \rho_2^\star$. Recalling that $E_{s,i}(\cdot)$ for $i = 1, 2$ is either the Gallager's source function, or a straight line tangent to it. We assume both $\rho_1^\star$ and $\rho_2^\star$, respectively are located at the straight line parts of $E_{s,1}(\cdot)$ and $E_{s,2}(\cdot)$. Without this assumption, in view of Figure 2.2 and Figure 2.3, (A.36) is equal to

$$\max\left\{\max_{\rho\in[0,1]} E(\rho, Q_1) - E_s(\rho, P_U), \max_{\rho\in[0,1]} E(\rho, Q_2) - E_s(\rho, P_U)\right\}. \quad \text{(A.43)}$$

Since both $\rho_1^\star$ and $\rho_2^\star$, are located at the straight line part, in view of (2.33) and (2.34), (A.36) can be written as

$$\begin{aligned} E(\rho_1^\star, Q_1) - E_s\left(\rho_{\gamma^\star}, P_U\right) - E_s'\left(\rho_{\gamma^\star}\right)\left(\rho_1^\star - \rho_{\gamma^\star}\right) = \\ E(\rho_2^\star, Q_2) - E_s\left(\rho_{\gamma^\star}, P_U\right) - E_s'\left(\rho_{\gamma^\star}\right)\left(\rho_2^\star - \rho_{\gamma^\star}\right). \end{aligned} \quad \text{(A.44)}$$

By solving (A.44) with respect to $E_s'\left(\rho_{\gamma^\star}\right)$, we find that

$$E_s'\left(\rho_{\gamma^\star}\right) = \frac{E(\rho_1^\star, Q_1) - E(\rho_2^\star, Q_2)}{\rho_1^\star - \rho_2^\star}. \quad \text{(A.45)}$$

Inserting the the right hand side of (A.45), in to $E_s'\left(\rho_{\gamma^\star}\right)$ appeared in the left hand side of (A.44), (A.39) is derived as

$$\begin{aligned} &\max_{\gamma\in[0,1]} \min_{i=1,2} \max_{\rho_i\in[0,1]} E(\rho_i, Q_i) - E_{s,i}(\rho, P_U, \gamma) = \\ &\left(1 - \frac{\rho_1^\star - \rho_{\gamma^\star}}{\rho_1^\star - \rho_2^\star}\right) E(\rho_1^\star, Q_1) + \frac{\rho_1^\star - \rho_{\gamma^\star}}{\rho_1^\star - \rho_2^\star} E(\rho_2^\star, Q_2) - E_s(\rho_{\gamma^\star}, P_U), \end{aligned} \quad \text{(A.46)}$$

where the sum of coefficients of $E(\rho_1^\star, Q_1)$ and $E(\rho_2^\star, Q_2)$ is one. By defining $\lambda = \frac{\rho_1^\star - \rho_{\gamma^\star}}{\rho_1^\star - \rho_2^\star}$, recalling the definitions of $\rho_1^\star$, $\rho_2^\star$ and $\gamma^\star$, and expressing $\rho_1^\star$ and $\rho_2^\star$ in terms of λ and $\rho_{\gamma^\star}$, (A.46) is obtained as

$$\begin{aligned} &\max_{\gamma\in[0,1]} \min_{i=1,2} \max_{\rho_i\in[0,1]} E(\rho_i, Q_i) - E_{s,i}(\rho, P_U, \gamma) = \\ &\max_{\rho_\gamma} \max_{\substack{\rho_1,\rho_2,\lambda\in[0,1]: \\ (1-\lambda)\rho_1+\lambda\rho_2=\rho_\gamma}} (1-\lambda)\, E(\rho_1, Q_1) + \lambda E(\rho_2, Q_2) - E_s(\rho_\gamma, P_U). \end{aligned} \quad \text{(A.47)}$$

By comparing (A.47) with (A.38), we conclude (A.37).

□

Lemma A.7. *Let $E(\rho)$ be a function of ρ. Considering (2.33) and (2.34), the function $f_1(\gamma) = \max_{\rho\in[0,1]} E(\rho) - E_{s,1}(\rho, P_U, \gamma)$ is non-decreasing with respect to γ and $f_2(\gamma) = \max_{\rho\in[0,1]} E(\rho) - E_{s,2}(\rho, P_U, \gamma)$ is non-increasing with respect to γ.*

Proof. Let $\gamma, \gamma' \in [0,1]$ where $\gamma \leq \gamma'$, or equivalently $\frac{1}{1+\rho_\gamma} \leq \frac{1}{1+\rho_{\gamma'}}$, where ρ_γ is defined in (2.25). Considering (2.33) we conclude that for all values of $\rho \in [0,1]$ we have $E_{s,1}(\rho, P_U, \gamma) \geq E_{s,1}(\rho, P_U, \gamma')$. Thus, the maximum of $E(\rho) - E_{s,1}(\rho, P_U, \gamma)$ is not greater than the maximum of $E(\rho) - E_{s,1}(\rho, P_U, \gamma')$ meaning that $f_1(\gamma) \leq f_1(\gamma')$ or that $f_1(\gamma)$ is non-decreasing in γ.

Similarly, if $\gamma \leq \gamma'$, by considering (2.34) we conclude that for all values of $\rho \in [0,1]$ we have $E_{s,2}(\rho, P_U, \gamma) \leq E_{s,2}(\rho, P_U, \gamma')$. Using the same reasoning, we have $f_2(\gamma) \geq f_2(\gamma')$, or equivalently that $f_2(\gamma)$ is non-increasing in γ.

□

Lemma A.8. *Let $k_1(\gamma)$ and $k_2(\gamma)$ be respectively continuous non-decreasing and non-increasing functions with respect to $\gamma \in [0,1]$. The optimal $\gamma^\star$ maximizing $\min_{i=1,2} k_i(\gamma)$ satisfies the following equation*

$$k_1(\gamma^\star) = k_2(\gamma^\star). \tag{A.48}$$

When (A.48) does not have any solution, we have $\gamma^\star = 0$ if $k_1(0) > k_2(0)$, and $\gamma^\star = 1$ otherwise.

Proof. In fact, the relative behaviour of a non-decreasing function with a non-increasing function can be categorized in three cases.

1. We focus on the first case where $k_1(0) < k_2(0)$ and $k_1(1) > k_2(1)$, i. e. there exists a $\gamma^\star$ such that $k_1(\gamma^\star) = k_2(\gamma^\star)$. In this case, the function $\min_i k_i(\gamma)$ is non-decreasing from $[0, \gamma^\star)$, and non-increasing from $(\gamma^\star, 1]$. Thus, the maximum over γ of $\min_i k_i(\gamma)$ occurs at $\gamma = \gamma^\star$.

2. If $k_1(0) < k_2(0)$ and $k_1(1) < k_2(1)$, $k_1(\gamma)$ and $k_2(\gamma)$ do not cross in $\gamma \in [0,1]$. Hence, we have $\min_i k_i(\gamma) = k_1(\gamma)$ and obviously since it is an non-decreasing function the maximum over γ occurs at $\gamma = \gamma^\star = 1$.

3. When $k_1(0) \geq k_2(0)$, we have $\min_i k_i(\gamma) = k_2(\gamma)$ and hence $\gamma^\star = 0$.

□

Lemma A.9. *For a given point-to-point channel W with input distribution input distribution Q, we have*

$$E_0^{\mathrm{cc}}(\rho, Q, W) \geq E_0(\rho, Q, W), \tag{A.49}$$

where $E_0^{\mathrm{cc}}(\cdot)$ and $E_0(\cdot)$ are respectively given by (2.56) and (1.14). More precisely,

$$\max_{\bar{\alpha}(x): \sum_x \bar{\alpha}(x) Q(x) = 0} -\log\left(\sum_y \left(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right) \geq$$
$$-\log\left(\sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right), \tag{A.50}$$

where the equality holds for the optimal Q maximizing (1.14).

Proof. To prove (A.50), we start by obtaining the optimal $\bar{\alpha}^{\star}(x)$ maximizing the left hand side of (A.50). Since logarithm is a non-decreasing function, the optimal $\bar{\alpha}^{\star}(x)$ maximizing

$$\max_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q(x)=0} -\log\left(\sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}\right), \tag{A.51}$$

has the same value with the one minimizing

$$\min_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q(x)=0} \sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}. \tag{A.52}$$

Since the objective function of (A.52) is convex with respect to $\bar{\alpha}(x)$ and the constraint $\sum_x \bar{\alpha}(x)Q(x) = 0$ is affine, we have

$$\min_{\bar{\alpha}(x):\sum_x \bar{\alpha}(x)Q(x)=0} \sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho} =$$
$$\max_{\mu} \min_{\bar{\alpha}(x)} \sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho} + \mu\left(-\sum_x \bar{\alpha}(x)Q(x)\right), \tag{A.53}$$

where by taking derivative with respect to $\bar{\alpha}(x)$, the optimal $\bar{\alpha}(x)^{\star}$ satisfies

$$\mu Q(x) = \sum_y e^{\frac{\bar{\alpha}(x)^{\star}}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_x e^{\frac{\bar{\alpha}(x)^{\star}}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{\rho}, \tag{A.54}$$

where by summing both sides of (A.54) with respect to x and considering the fact $\sum_x Q(x) = 1$, we obtain

$$\mu = \sum_y \Big(\sum_x e^{\frac{\bar{\alpha}(x)^{\star}}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{1+\rho}. \tag{A.55}$$

Comparing (A.55) with (2.56), we have

$$E_0^{\mathrm{cc}}(\rho, Q, W) = -\log(\mu). \tag{A.56}$$

Additionally, removing $Q(x) \neq 0$ from both sides of (A.54) and taking logarithm from both sides of it, yields

$$\frac{\bar{\alpha}(x)^{\star}}{1+\rho} = \log(\mu) - \log\left(\sum_y W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_x e^{\frac{\bar{\alpha}(x)^{\star}}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{\rho}\right), \tag{A.57}$$

where by applying the constraint $\sum_x \bar{\alpha}(x)^{\star} Q(x) = 0$ to (A.57), we find that

$$\log(\mu) = \sum_x Q(x) \log\left(\sum_y W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_x e^{\frac{\bar{\alpha}(x)^{\star}}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}}\Big)^{\rho}\right), \tag{A.58}$$

and by comparing (A.58) with (A.56), E_0^{cc} will equal to

$$E_0^{\text{cc}}(\rho, Q, W) = -\sum_x Q(x) \log \left(\sum_y W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_x e^{\frac{\bar{\alpha}(x)^\star}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}} \Big)^\rho \right). \tag{A.59}$$

Next, applying Jensen's inequality to (A.59) yields

$$E_0^{\text{cc}}(\rho, Q, W) \geq -\log \left(\sum_x Q(x) \sum_y W(y|x)^{\frac{1}{1+\rho}} \Big(\sum_x e^{\frac{\bar{\alpha}(x)^\star}{1+\rho}} Q(x) W(y|x)^{\frac{1}{1+\rho}} \Big)^\rho \right), \tag{A.60}$$

which is equal to $E_0(\cdot)$ function, i. e.

$$E_0^{\text{cc}}(\rho, Q, W) \geq E_0(\rho, Q, W), \tag{A.61}$$

where as shown in [20, Eq. (31)], for the optimal input distribution maximizing $E_0(\cdot)$, we have equality in (A.61).

□

Lemma A.10. *Let $h(u, x, y)$ be a positive function of variables (u, x, y) and suppose $k(x, u)$ be a positive function of (x, u) such that $\sum_{u,x} k(x, u) = 1$. Consider the following optimization problem*

$$\min_{\bar{\alpha}(x): \sum_{x,u} k(x,u)\bar{\alpha}(x) = 0} \sum_y \left(\sum_{u,x} e^{\frac{\bar{\alpha}(x)}{1+\rho}} h(u, x, y) \right)^{1+\rho}, \tag{A.62}$$

where $\rho \in [0, 1]$. The optimal value of $\bar{\alpha}(x)$ which minimizes the objective function in (A.62) is denoted by $\bar{\alpha}^*(x)$ and satisfies

$$\bar{\alpha}^*(x) = (1+\rho) \log \frac{\sum_{\bar{u}} k(x, \bar{u})}{\sum_{\bar{y}} \left(\sum_{\bar{u},\bar{x}} h(\bar{u}, \bar{x}, \bar{y}) e^{\frac{\bar{\alpha}^*(\bar{x})}{1+\rho}} \right)^\rho \sum_{\hat{u}} h(\hat{u}, x, \bar{y})}$$
$$\times \left(\sum_{\tilde{x},\tilde{u}} k(\tilde{x}, \tilde{u}) \log \frac{\sum_{\bar{y}} \left(\sum_{\bar{u},\bar{x}} h(\bar{u}, \bar{x}, \bar{y}) e^{\frac{\bar{\alpha}^*(\bar{x})}{1+\rho}} \right)^\rho \sum_{\hat{u}} h(\hat{u}, \tilde{x}, \bar{y})}{\sum_{\bar{u}} k(\tilde{x}, \bar{u})} \right). \tag{A.63}$$

In addition, the optimal value of $\tilde{\alpha}(u, x)$ which minimizes the following problem

$$\min_{\tilde{\alpha}(u,x): \sum_{x,u} k(x,u)\tilde{\alpha}(u,x) = 0} \sum_y \left(\sum_{u,x} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} h(u, x, y) \right)^{1+\rho}, \tag{A.64}$$

is denoted by $\tilde{\alpha}^*(u,x)$ *and satisfies*

$$\tilde{\alpha}^*(u,x) = (1+\rho)\log \frac{k(x,u)}{\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\tilde{\alpha}^*(\bar{u},\bar{x})}{1+\rho}}\Big)^{\rho} h(u,x,\bar{y})}$$
$$\times\Big(\sum_{\tilde{x},\tilde{u}} k(\tilde{x},\tilde{u})\log \frac{\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\tilde{\alpha}^*(\bar{u},\bar{x})}{1+\rho}}\Big)^{\rho} h(\tilde{u},\tilde{x},\bar{y})}{k(\tilde{x},\tilde{u})}\Big). \tag{A.65}$$

Proof. To solve the optimization problems given in Lemma A.10 we apply Lagrange duality theory. Since the objective functions given in (A.62) and (A.64) are convex with respect to $\bar{\alpha}(x)$ and $\tilde{\alpha}(u,x)$, respectively and the constraints $\sum_{x,u} k(x,u)\bar{\alpha}(x) = 0$ and $\sum_{x,u} k(x,u)\tilde{\alpha}(u,x) = 0$ are affine, the strong duality holds for both problems.

To prove Lemma A.10, firstly we focus on the optimization problem given (A.62). Since the primal problem given in (A.62) satisfies the strong duality conditions, we have

$$\min_{\bar{\alpha}(x):\sum_{x,u} k(x,u)\bar{\alpha}(x)=0} \sum_y \Big(\sum_{u,x} e^{\frac{\bar{\alpha}(x)}{1+\rho}} h(u,x,y)\Big)^{1+\rho} = \max \min_{\bar{\alpha}(x)} \hat{\mathcal{L}}_2(\bar{\alpha},\nu), \tag{A.66}$$

where $\min_{\bar{\alpha}(x)} \hat{\mathcal{L}}_2(\bar{\alpha},\nu)$ is the Lagrange dual function of the primary problem (A.62) and the Lagrangian is given by

$$\hat{\mathcal{L}}_2(\bar{\alpha},\nu) = \sum_y \Big(\sum_{u,x} e^{\frac{\bar{\alpha}(x)}{1+\rho}} h(u,x,y)\Big)^{1+\rho} + \nu\Big(0 - \sum_{x,u} k(x,u)\bar{\alpha}(x)\Big), \tag{A.67}$$

where ν is the Lagrange multiplier associated with the given constraint $\sum_{x,u} k(x,u)\bar{\alpha}(x) = 0$. In order to determine $\max\min_{\bar{\alpha}(x)} \hat{\mathcal{L}}_2(\bar{\alpha},\nu)$, in view of the KKT conditions, for the optimal values of $(\bar{\alpha},\nu)$, we have $\frac{\partial \hat{\mathcal{L}}_2(\bar{\alpha},\nu)}{\partial \bar{\alpha}} = 0$, which leads to

$$\sum_{\bar{y}} \Big(\sum_{\bar{u},\bar{x}} e^{\frac{\bar{\alpha}(\bar{x})}{1+\rho}} h(\bar{u},\bar{x},\bar{y})\Big)^{\rho} e^{\frac{\bar{\alpha}(x)}{1+\rho}} \sum_{\hat{u}} h(\hat{u},x,\bar{y}) - \nu \sum_{\bar{u}} k(x,\bar{u}) = 0. \tag{A.68}$$

By solving (A.68) with respect to $\bar{\alpha}(x)$, we obtain

$$\bar{\alpha}(x) = (1+\rho)\log \frac{\nu\sum_{\bar{u}} k(x,\bar{u})}{\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\bar{\alpha}(\bar{x})}{1+\rho}}\Big)^{\rho} \sum_{\hat{u}} h(\hat{u},x,\bar{y})}. \tag{A.69}$$

To apply the constraint $\sum_{x,u} k(x,u)\bar{\alpha}(x) = 0$, by multiplying both sides of (A.69) by $k(x,u)$, summing over (u,x) and using the fact that $\log(ab) =$

$\log(a) - \log(\frac{1}{b})$, we get

$$\log(\nu)\sum_{x,u} k(x,u) = \sum_{x,u} k(x,u)\log\frac{\sum_{\bar{y}}\left(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\bar{\alpha}(\bar{x})}{1+\rho}}\right)^{\rho}\sum_{\hat{u}} h(\hat{u},x,\bar{y})}{\sum_{\bar{u}} k(x,\bar{u})}. \tag{A.70}$$

Considering the identity that $\sum_{x,u} k(u,x) = 1$, optimal ν is derived as

$$\nu = \exp\left(\sum_{\tilde{x},\tilde{u}} k(\tilde{x},\tilde{u})\log\frac{\sum_{\bar{y}}\left(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\bar{\alpha}(\bar{x})}{1+\rho}}\right)^{\rho}\sum_{\hat{u}} h(\hat{u},\tilde{x},\bar{y})}{\sum_{\bar{u}} k(\tilde{x},\bar{u})}\right). \tag{A.71}$$

Putting back ν obtained in (A.71) into (A.69) and using the fact that $e^{\log a} = a$, the optimal value of $\bar{\alpha}$ is derived in (A.63) as $\bar{\alpha}^*(x)$.

Now, in order to find the optimal $\tilde{\alpha}(u,x)$ which minimizes (A.64), again we apply Lagrange duality theory. since the approach is exactly the same, we omit some details.

The Lagrange dual function to the primary problem (A.64) is

$$\hat{\mathcal{L}}_3(\tilde{\alpha},\mu) = \sum_y\left(\sum_{u,x} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} h(u,x,y)\right)^{1+\rho} + \mu\Big(0 - \sum_{x,u} k(x,u)\tilde{\alpha}(u,x)\Big), \tag{A.72}$$

where μ is the Lagrange multiplier associated with the given constraint $\sum_{x,u} k(x,u)\tilde{\alpha}(u,x) = 0$. In view of the KKT conditions, by setting $\frac{\partial\hat{\mathcal{L}}_3(\tilde{\alpha},\mu)}{\partial\tilde{\alpha}} = 0$, we have

$$\sum_y\left(\sum_{u,x} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} h(u,x,y)\right)^{\rho} e^{\frac{\tilde{\alpha}(u,x)}{1+\rho}} h(u,x,y) - \mu k(x,u) = 0. \tag{A.73}$$

Solving (A.73) with respect to $\tilde{\alpha}(u,x)$, applying $\sum_{x,u} k(x,u)\tilde{\alpha}(u,x) = 0$ and $\sum_{x,u} k(u,x) = 1$, after some mathematical manipulations, the optimal value of $\tilde{\alpha}(u,x)$ is derived by (A.65) as $\tilde{\alpha}^*(u,x)$. □

Lemma A.11. *Let $h(u,x,y)$ and $k(x,u)$ in Lemma A.10 are chosen such that for all values of (u,x,y), we have $\frac{h(u,x,y)}{k(x,u)} = \frac{\sum_u h(u,x,y)}{\sum_u k(x,u)}$. Then $\tilde{\alpha}^*(u,x) = \bar{\alpha}^*(x)$ for all $u \in \mathcal{U}$ and the quantity of (A.62) is equal with the quantity given in (A.64).*

Proof. If $\frac{h(u,x,y)}{k(x,u)} = \frac{\sum_u h(u,x,y)}{\sum_u k(x,u)}$, we define $z(x,y)$ as $z(x,y) \triangleq \frac{h(u,x,y)}{k(x,u)} = \frac{\sum_u h(u,x,y)}{\sum_u k(x,u)}$. By replacing $z(x,y)$ instead of $\frac{\sum_u h(u,x,y)}{\sum_u k(x,u)}$ and $\frac{h(u,x,y)}{k(x,u)}$ in (A.63) and (A.65), respectively the quantities of $\bar{\alpha}^*(x)$ and $\tilde{\alpha}^*(u,x)$ are given by

$$\bar{\alpha}^*(x) = -(1+\rho)\log\Big(\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\bar{\alpha}^*(\bar{x})}{1+\rho}}\Big)^{\rho} z(x,\bar{y})\Big)$$
$$\times\Big(\sum_{\tilde{x},\tilde{u}} k(\tilde{x},\tilde{u})\log\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\bar{\alpha}^*(\bar{x})}{1+\rho}}\Big)^{\rho} z(\tilde{x},\bar{y})\Big), \tag{A.74}$$

and

$$\tilde{\alpha}^*(u,x) = -(1+\rho)\log\Big(\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\tilde{\alpha}^*(\bar{u},\bar{x})}{1+\rho}}\Big)^{\rho} z(x,\bar{y})\Big)$$
$$\times\Big(\sum_{\tilde{x},\tilde{u}} k(\tilde{x},\tilde{u})\log\sum_{\bar{y}}\Big(\sum_{\bar{u},\bar{x}} h(\bar{u},\bar{x},\bar{y})e^{\frac{\tilde{\alpha}^*(\bar{u},\bar{x})}{1+\rho}}\Big)^{\rho} z(\tilde{x},\bar{y})\Big), \tag{A.75}$$

where from (A.75), we may note that the first term of (A.75) only depends on x and the second term, is constant. As a result, when $\frac{h(u,x,y)}{k(x,u)} = \frac{\sum_u h(u,x,y)}{\sum_u k(x,u)} = z(x,y)$ holds, the quantity of $\tilde{\alpha}^*(u,x)$ only depends on x and for $u' \neq u$, we have $\tilde{\alpha}^*(u',x) = \tilde{\alpha}^*(u,x)$ which implies $\tilde{\alpha}^*(u,x) = \tilde{\alpha}^*(x)$. Using this fact, by replacing $\tilde{\alpha}^*(u,x)$ with $\tilde{\alpha}^*(x)$ in (A.75) and comparing it with with $\bar{\alpha}^*(x)$ given in (A.74), we conclude that $\tilde{\alpha}^*(u,x) = \tilde{\alpha}^*(x) = \bar{\alpha}^*(x)$ when $\frac{h(u,x,y)}{k(x,u)} = \frac{\sum_u h(u,x,y)}{\sum_u k(x,u)}$. Putting back the optimal value of $\bar{\alpha}^*(x)$ given in (A.74) into objective functions in (A.62) and (A.64), concludes Lemma A.11, i. e. the quantity of (A.62) is equal with the quantity given in (A.64). □

Lemma A.12. *For $\gamma_{i+1}, \gamma_i \in [0,1]$, let $\gamma_{i+1} \le \gamma_i$. Then, for probability distribution P_U we have*

$$\min_{\lambda_{i+1},\lambda_i\ge 0}\left(\sum_{\boldsymbol{u}} P^n_{\boldsymbol{U}}(\boldsymbol{u})^{\frac{1}{1+\rho}}\left(\frac{\gamma_i^n}{P^n_{\boldsymbol{U}}(\boldsymbol{u})}\right)^{\frac{\lambda_i}{1+\rho}}\left(\frac{P^n_{\boldsymbol{U}}(\boldsymbol{u})}{\gamma_{i+1}^n}\right)^{\frac{\lambda_{i+1}}{1+\rho}}\right)^{1+\rho} = e^{E_{s,i}\left(\rho,P^n_{\boldsymbol{U}},\gamma_{i+1},\gamma_i\right)}, \tag{A.76}$$

where $E_{s,i}\left(\rho, P_U, \gamma_{i+1}, \gamma_i\right)$ is given by (2.24).

Proof. Since the objective function of (A.76) is convex with respect to λ_i and λ_{i+1}, we have

$$\frac{\partial}{\partial\lambda_j}\left(\sum_{\boldsymbol{u}} P^n_{\boldsymbol{U}}(\boldsymbol{u})^{\frac{1}{1+\rho}}\left(\frac{\gamma_i^n}{P^n_{\boldsymbol{U}}(\boldsymbol{u})}\right)^{\frac{\lambda_i}{1+\rho}}\left(\frac{P^n_{\boldsymbol{U}}(\boldsymbol{u})}{\gamma_{i+1}^n}\right)^{\frac{\lambda_{i+1}}{1+\rho}}\right)^{1+\rho}\Bigg|_{\lambda_j^*\ge 0} = 0, \tag{A.77}$$

for $j = i, i+1$. We recall that when the solution $\lambda_j^\star$ for $j = i, i+1$ to (A.77) is strictly negative, then $\lambda_j^\star = 0$. Let $\lambda_i^\star > 0$, then (A.77) yields

$$\frac{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-\lambda_i^\star+\lambda_{i+1}^\star}{1+\rho}} \log(P_{\boldsymbol{U}}^n(\boldsymbol{u}))}{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-\lambda_i^\star+\lambda_{i+1}^\star}{1+\rho}}} = \log(\gamma_i^n), \tag{A.78}$$

and if $\lambda_{i+1}^\star > 0$, from (A.77) we have

$$\frac{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-\lambda_i^\star+\lambda_{i+1}^\star}{1+\rho}} \log(P_{\boldsymbol{U}}^n(\boldsymbol{u}))}{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-\lambda_i^\star+\lambda_{i+1}^\star}{1+\rho}}} = \log(\gamma_{i+1}^n). \tag{A.79}$$

Since the left hand sides of (A.78) and (A.79) are same, we conclude that if both $\lambda_i^\star > 0$ and $\lambda_{i+1}^\star > 0$, then $\gamma_i = \gamma_{i+1}$. In other words, if $\gamma_i \neq \gamma_{i+1}$, it is impossible that (A.78) and (A.79) are satisfied at the same time, i. e. it is impossible that $\lambda_i^\star > 0$ and $\lambda_{i+1}^\star > 0$ at the same time.

Thus, we consider three cases including only (A.78) is satisfied, only (A.79) is satisfied and none of the them are satisfied and we define ρ_{γ_i} and $\rho_{\gamma_{i+1}}$ as

$$\frac{1-\lambda_i}{1+\rho} = \frac{1}{1+\rho_{\gamma_i}}, \tag{A.80}$$

$$\frac{1+\lambda_{i+1}}{1+\rho} = \frac{1}{1+\rho_{\gamma_{i+1}}}. \tag{A.81}$$

1. Let only (A.78) is satisfied, i. e. $\lambda_i^\star > 0$ and $\lambda_{i+1}^\star = 0$. Applying $\lambda_i^\star > 0$ into (A.80), yields

$$\frac{1}{1+\rho} - \frac{1}{1+\rho_{\gamma_i}} > 0. \tag{A.82}$$

   Moreover, by inserting $\lambda_{i+1}^\star = 0$ and $\frac{1-\lambda_i^\star}{1+\rho} = \frac{1}{1+\rho_{\gamma_i}}$ into (A.78), we obtain

$$\frac{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_i}}} \log(P_{\boldsymbol{U}}^n(\boldsymbol{u}))}{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_i}}}} = \log(\gamma_i^n), \tag{A.83}$$

   where by considering (1.9), γ_i^n in (A.83) can be expressed as

$$\gamma_i^n = e^{E_s(\rho_{\gamma_i}, P_{\boldsymbol{U}}^n(\boldsymbol{u})) - (1+\rho_{\gamma_i}) E_s'(\rho_{\gamma_i}, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}. \tag{A.84}$$

In addition, setting $\frac{1-\lambda_i^\star}{1+\rho} = \frac{1}{1+\rho_{\gamma_i}}$ and $\lambda_{i+1}^\star = 0$ into the left hand side of (A.76), leads to

$$\min_{\lambda_{i+1},\lambda_i \geq 0} \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma_i^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})} \right)^{\frac{\lambda_i}{1+\rho}} \left(\frac{P_{\boldsymbol{U}}^n(\boldsymbol{u})}{\gamma_{i+1}^n} \right)^{\frac{\lambda_{i+1}}{1+\rho}} \right)^{1+\rho} =$$
$$= \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_i}}} \right)^{1+\rho} \gamma_i^{n\frac{\rho_{\gamma_i}-\rho}{1+\rho}} = e^{(1+\rho)\log\left(P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_i}}} \right)} \gamma_i^{n\frac{\rho_{\gamma_i}-\rho}{1+\rho_{\gamma_i}}}, \quad \text{(A.85)}$$

where we used $a^{(1+\rho)} = e^{(1+\rho)\log(a)}$. Substituting (A.84) into (A.85), yields

$$\min_{\lambda_{i+1},\lambda_i \geq 0} \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma_i^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})} \right)^{\frac{\lambda_i}{1+\rho}} \left(\frac{P_{\boldsymbol{U}}^n(\boldsymbol{u})}{\gamma_{i+1}^n} \right)^{\frac{\lambda_{i+1}}{1+\rho}} \right)^{1+\rho}$$
$$= e^{E_s(\rho_{\gamma_i}, P_{\boldsymbol{U}}^n(\boldsymbol{u})) - (\rho-\rho_{\gamma_i}) E_s'(\rho_{\gamma_i}, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}, \quad \text{(A.86)}$$

for ρs satisfying (A.82), i. e. for

$$\frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_i}}. \quad \text{(A.87)}$$

2. Let only (A.79) is satisfied, i. e. $\lambda_{i+1}^\star > 0$ and $\lambda_i^\star = 0$. Using the same steps as previous case, by defining $\frac{1+\lambda_{i+1}^\star}{1+\rho} = \frac{1}{1+\rho_{\gamma_{i+1}}}$, and applying $\lambda_{i+1}^\star > 0$ into (A.81), for ρs satisfying

$$\frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{i+1}}}, \quad \text{(A.88)}$$

we have

$$\min_{\lambda_{i+1},\lambda_i \geq 0} \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma_i^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})} \right)^{\frac{\lambda_i}{1+\rho}} \left(\frac{P_{\boldsymbol{U}}^n(\boldsymbol{u})}{\gamma_{i+1}^n} \right)^{\frac{\lambda_{i+1}}{1+\rho}} \right)^{1+\rho}$$
$$= e^{E_s(\rho_{\gamma_{i+1}}, P_{\boldsymbol{U}}^n(\boldsymbol{u})) - (\rho-\rho_{\gamma_{i+1}}) E_s'(\rho_{\gamma i+1}, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}, \quad \text{(A.89)}$$

where $\rho_{\gamma_{i+1}}$ satisfies the following equation

$$\frac{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_{i+1}}}} \log(P_{\boldsymbol{U}}^n(\boldsymbol{u}))}{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_{\gamma_{i+1}}}}} = \log(\gamma_{i+1}^n). \quad \text{(A.90)}$$

3. Let none of the equations (A.78) and (A.79) are satisfied, i. e. both λ_{i+1} and λ_i satisfying (A.77) are negative. Applying $\lambda_{i+1} \leq 0$ and $\lambda_{i+1} \leq$ into (A.80) and (A.81), for ρs satisfying

$$\frac{1}{1+\rho_{\gamma_{i+1}}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_i}}, \tag{A.91}$$

we insert $\lambda^\star_{i+1} = \lambda^\star_i = 0$ into the left hand side of (A.76), which leads

$$\min_{\lambda_{i+1},\lambda_i \geq 0} \left(\sum_{\boldsymbol{u}} P^n_{\boldsymbol{U}}(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma^n_i}{P^n_{\boldsymbol{U}}(\boldsymbol{u})} \right)^{\frac{\lambda_i}{1+\rho}} \left(\frac{P^n_{\boldsymbol{U}}(\boldsymbol{u})}{\gamma^n_{i+1}} \right)^{\frac{\lambda_{i+1}}{1+\rho}} \right)^{1+\rho} = e^{E_s(\rho, P^n_{\boldsymbol{U}}(\boldsymbol{u}))}. \tag{A.92}$$

Combining (A.92) for (A.91), (A.89) for (A.88) and (A.86) for (A.87), concludes the proof. □

Lemma A.13. *Let $L_{0s}(\rho) = L_0(\rho) - L_s(\rho)$ where $L_0(\rho)$ is a continuous function and $L_s(\rho)$ is a convex function of ρ. Then,*

$$\bar{L}_{0s}(\rho) \leq \bar{L}_0(\rho) - L_s(\rho), \tag{A.93}$$

where $\bar{L}_{0s}$ and $\bar{L}_0$ denote the concave hull of $L_{0s}(\rho)$ and $L_0(\rho)$, respectively.

Proof. From the definition of concave hull in (2.4), the left hand side of (A.93) is given by

$$\bar{L}_{0s}(\rho) = \sup_{\substack{\rho_1,\rho_2,\theta\in[0,1]: \\ \theta\rho_1+(1-\theta)\rho_2=\rho}} \Big\{ \theta L_{0s}(\rho_1) + (1-\theta) L_{0s}(\rho_2) \Big\}. \tag{A.94}$$

Using the definition of $L_{0s}(\rho)$, the right hand side of (A.94) is simplified as

$$\theta L_{0s}(\rho_1) + (1-\theta) L_{0s}(\rho_2) = \\ \theta L_0(\rho_1) + (1-\theta) L_0(\rho_2) - \theta L_s(\rho_1) - (1-\theta) L_s(\rho_2). \tag{A.95}$$

Since $L_s(\rho)$ is a convex function of ρ, and so $\theta L_s(\rho_1) + (1-\theta) L_s(\rho_2) \geq L_s(\theta\rho_1 + (1-\theta)\rho_2)$, we further obtain that

$$\theta L_{0s}(\rho_1) + (1-\theta) L_{0s}(\rho_2) \leq \theta L_0(\rho_1) + (1-\theta) L_0(\rho_2) - L_s(\rho), \tag{A.96}$$

where we used that $\theta\rho_1 + (1-\theta)\rho_2 = \rho$. Taking supremum from both sides of (A.96), in view of [44, Sec. 2.9], we obtain that

$$\sup_{\substack{\rho_1,\rho_2,\theta\in[0,1]: \\ \theta\rho_1+(1-\theta)\rho_2=\rho}} \Big\{ \theta L_{0s}(\rho_1) + (1-\theta) L_{0s}(\rho_2) \Big\} \leq \\ \sup_{\substack{\rho_1,\rho_2,\theta\in[0,1]: \\ \theta\rho_1+(1-\theta)\rho_2=\rho}} \Big\{ \theta L_0(\rho_1) + (1-\theta) L_0(\rho_2) \Big\} - L_s(\rho), \tag{A.97}$$

concluding the proof. □

Lemma A.14. *Let T and Z be two correlated random variables characterized by $P_{TZ} = P_U$. For a given channel W, source $P_U = P_{TZ}$, and input distribution Q, let E be*

$$E = \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_U||P_U) + D(\hat{P}_{XY}||QW) + \Big[\min_{\tilde{P}_U \in \mathcal{K}_s(\hat{P}_U)} \min_{\tilde{P}_{XY} \in \mathcal{K}_c(\hat{P}_{XY})} D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z}) \Big]^+, \quad \text{(A.98)}$$

where

$$\mathcal{K}_s(\hat{P}_U) \triangleq \Big\{ \tilde{P}_U \in \mathcal{P}_\mathcal{U} : \tilde{P}_Z = \hat{P}_Z,\ \mathbb{E}_{\tilde{P}} \log \big(P_U(U)\big) \geq \mathbb{E}_{\hat{P}} \log \big(P_U(U)\big) \Big\}, \quad \text{(A.99)}$$

and

$$\mathcal{K}_c(\hat{P}_{XY}) \triangleq \Big\{ \tilde{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}} : \tilde{P}_Y = \hat{P}_Y, \mathbb{E}_{\tilde{P}} \log \big(W(Y|X)\big) \geq \mathbb{E}_{\hat{P}} \log \big(W(Y|X)\big) \Big\}. \quad \text{(A.100)}$$

It can be proved that

$$E \geq \min_{\hat{P}_U \in \mathcal{P}_\mathcal{U}} \min_{\hat{P}_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}} D(\hat{P}_U||P_U) + D(\hat{P}_{XY}||QW) + \Big[D(\hat{P}_{XY}||Q\hat{P}_Y) - H(\hat{P}_{T|Z}) \Big]^+. \quad \text{(A.101)}$$

Proof. Firstly, assume for the optimal $\hat{P}_U$, $\hat{P}_{XY}$, $\tilde{P}_U$ and $\tilde{P}_{XY}$ minimizing (A.98), we have

$$D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z}) \geq D(\hat{P}_{XY}||Q\hat{P}_Y) - H(\hat{P}_{T|Z}), \quad \text{(A.102)}$$

which leads to

$$\Big[D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z}) \Big]^+ \geq \Big[D(\hat{P}_{XY}||Q\hat{P}_Y) - H(\hat{P}_{T|Z}) \Big]^+. \quad \text{(A.103)}$$

Adding $D(\hat{P}_U||P_U) + D(\hat{P}_{XY}||QW)$ to the both sides of (A.103), (A.101) is proved. Alternatively, if

$$D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z}) \leq D(\hat{P}_{XY}||Q\hat{P}_Y) - H(\hat{P}_{T|Z}), \quad \text{(A.104)}$$

in view of (A.99), since $\tilde{P}_Z(z) = \hat{P}_Z(z)$, for all $z \in \mathcal{Z}$, we add $-H(\tilde{P}_Z) = -H(\hat{P}_Z)$ to the both sides of (A.104), where since $P_{T|Z} = \frac{P_U}{P_Z}$, we have

$$D(\tilde{P}_{XY}||Q\hat{P}_Y) + \sum_u \tilde{P}_U(u) \log \big(\tilde{P}_U(u)\big) \leq D(\hat{P}_{XY}||Q\hat{P}_Y) + \sum_u \hat{P}_U(u) \log \big(\hat{P}_U(u)\big). \quad \text{(A.105)}$$

Next, by using $\mathbb{E}_{\tilde{P}} \log \Big(P_U(U)\Big) \geq \mathbb{E}_{\hat{P}} \log \Big(P_U(U)\Big)$ and $\mathbb{E}_{\tilde{P}} \log \Big(W(Y|X)\Big) \geq \mathbb{E}_{\hat{P}} \log \Big(W(Y|X)\Big)$, respectively given by (A.99) and (A.100), we find

$$\sum_u \tilde{P}_U(u) \log\left(P_U(u)\right) + \sum_{x,y} \tilde{P}_{XY}(x,y) \log\left(W(y|x)\right) \geq \sum_u \hat{P}_U(u) \log\left(P_U(u)\right) + \sum_{x,y} \hat{P}_{XY}(x,y) \log\left(W(y|x)\right). \quad \text{(A.106)}$$

Subtracting (A.106) from (A.105) leads to

$$\sum_u \tilde{P}_U(u) \log\left(\frac{\tilde{P}_U(u)}{P_U(u)}\right) + \sum_{x,y} \tilde{P}_{XY}(x,y) \log\left(\frac{\tilde{P}_{XY}(x,y)}{W(y|x)Q(x)\hat{P}_Y(y)}\right) \leq \sum_u \hat{P}_U(u) \log\left(\frac{\hat{P}_U(u)}{P_U(u)}\right) + \sum_{x,y} \hat{P}_{XY}(x,y) \log\left(\frac{\hat{P}_{XY}(x,y)}{W(y|x)Q(x)\hat{P}_Y(y)}\right). \quad \text{(A.107)}$$

Moreover, in view of (A.100), $\tilde{P}_Y = \hat{P}_Y$ which yields $H(\tilde{P}_Y) = H(\hat{P}_Y)$ or equivalently

$$\sum_{x,y} \tilde{P}_{XY}(x,y) \log \hat{P}_Y(y) = \sum_{x,y} \hat{P}_{XY}(x,y) \log \hat{P}_Y(y). \quad \text{(A.108)}$$

By adding (A.108) to the both sides of (A.107), we have

$$\sum_u \tilde{P}_U(u) \log\left(\frac{\tilde{P}_U(u)}{P_U(u)}\right) + \sum_{x,y} \tilde{P}_{XY}(x,y) \log\left(\frac{\tilde{P}_{XY}(x,y)}{W(y|x)Q(x)}\right) \leq \sum_u \hat{P}_U(u) \log\left(\frac{\hat{P}_U(u)}{P_U(u)}\right) + \sum_{x,y} \hat{P}_{XY}(x,y) \log\left(\frac{\hat{P}_{XY}(x,y)}{W(y|x)Q(x)}\right). \quad \text{(A.109)}$$

Using the definition of the relative entropy, (A.109) can be expressed as

$$D(\tilde{P}_U||P_U) + D(\tilde{P}_{XY}||QW) \leq D(\hat{P}_U||P_U) + D(\hat{P}_{XY}||QW). \quad \text{(A.110)}$$

By adding $\Big[D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z})\Big]^+$ on the both sides of (A.110), we obtain

$$D(\tilde{P}_U||P_U) + D(\tilde{P}_{XY}||QW) + \Big[D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z})\Big]^+ \leq D(\hat{P}_U||P_U) + D(\hat{P}_{XY}||QW) + \Big[D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z})\Big]^+. \quad \text{(A.111)}$$

Inasmuch as $\mathcal{K}_s(\hat{P}_U) \subset \mathcal{P}_{\mathcal{U}}$ and $\mathcal{K}_c(\hat{P}_{XY}) \subset \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}$, we have proved that whether $D(\tilde{P}_{XY}||Q\hat{P}_Y) - H(\tilde{P}_{T|Z})$ be lower than $D(\hat{P}_{XY}||Q\hat{P}_Y) - H(\hat{P}_{T|Z})$ or greater, we have (A.101). □

Lemma A.15. *Let $E(\rho)$ be a continuous function of ρ and we have $0 = \gamma_{L+1} \leq \gamma_L \leq ... \leq \gamma_2 < \gamma_1 = 1$. Considering (2.24), for $\ell = 2, ..., L$, the function*

$$f_{\ell-1}(\gamma_\ell) = \max_{\rho\in[0,1]} E(\rho) - E_{s,\ell-1}(\rho, P_U, \gamma_{\nu,\ell}, \gamma_{\nu,\ell-1}), \tag{A.112}$$

is non-decreasing with respect to γ_ℓ and

$$f_\ell(\gamma_\ell) = \max_{\rho\in[0,1]} E(\rho) - E_{s,\ell}(\rho, P_U, \gamma_{\nu,\ell+1}, \gamma_{\nu,\ell}), \tag{A.113}$$

is non-increasing with respect to γ_ℓ.

Proof. In order to prove Lemma A.15, we compute $E_{s,\ell-1}(\cdot)$ and $E_{s,\ell}(\cdot)$ for the fixed source-partitioning thresholds $\gamma_{\ell+1} \leq \gamma_\ell \leq \gamma_{\ell-1}$ and also we compute $E_{s,\ell'-1}(\cdot)$ and $E_{s,\ell'}(\cdot)$ for the fixed source-partitioning thresholds $\gamma_{\ell+1} \leq \gamma_{\ell'} \leq \gamma_{\ell-1}$. We will show that if $\gamma_\ell \leq \gamma_{\ell'}$ then $E_{s,\ell-1}(\cdot) \geq E_{s,\ell'-1}(\cdot)$ and $E_{s,\ell}(\cdot) \leq E_{s,\ell'}(\cdot)$. Thus, $\max_\rho E(\rho) - E_{s,\ell-1}(\rho, P_U) \leq \max_\rho E(\rho) - E_{s,\ell'-1}(\rho, P_U)$ and $\max_\rho E(\rho) - E_{s,\ell}(\rho, P_U) \geq \max_\rho E(\rho) - E_{s,\ell'}(\rho, P_U)$ meaning that $f_{\ell-1}(\gamma_\ell)$ and $f_\ell(\gamma_\ell)$ are respectively non-decreasing and non-increasing with respect to γ_ℓ.

To prove $E_{s,\ell-1}(\cdot) \geq E_{s,\ell'-1}(\cdot)$, recalling that $\gamma_\ell < \gamma_{\ell'}$, we have $\frac{1}{1+\rho_{\gamma_\ell}} < \frac{1}{1+\rho_{\gamma_{\ell'}}}$. Hence, for $\gamma_{\ell+1} \leq \gamma_\ell \leq \gamma_{\ell-1}$ we express $E_{s,\ell-1}(\cdot)$ as

$$E_{s,\ell-1}(\rho, P_U, \gamma_{\nu,\ell}, \gamma_{\nu,\ell-1}) = \begin{cases} E_s(\rho_{\gamma_\ell}, P_U) + E_s'(\rho_{\gamma_\ell})(\rho - \rho_{\gamma_\ell}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_\ell}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_\ell}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell'}}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_{\ell'}}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell-1}}}, \\ E_s(\rho_{\gamma_{\ell-1}}, P_U) + E_s'(\rho_{\gamma_{\ell-1}})(\rho - \rho_{\gamma_{\ell-1}}) & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_{\ell-1}}}, \end{cases} \tag{A.114}$$

and for $\gamma_{\ell+1} \leq \gamma_{\ell'} \leq \gamma_{\ell-1}$, $E_{s,\ell'-1}(\cdot)$ can be written as

$$E_{s,\ell'-1}(\rho, P_U, \gamma_{\nu,\ell'}, \gamma_{\nu,\ell'-1}) = \begin{cases} E_s(\rho_{\gamma_{\ell'}}, P_U) + E_s'(\rho_{\gamma_{\ell'}})(\rho - \rho_{\gamma_{\ell'}}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_\ell}}, \\ E_s(\rho_{\gamma_{\ell'}}, P_U) + E_s'(\rho_{\gamma_{\ell'}})(\rho - \rho_{\gamma_{\ell'}}) & \frac{1}{1+\rho_{\gamma_\ell}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell'}}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_{\ell'}}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell-1}}}, \\ E_s(\rho_{\gamma_{\ell-1}}, P_U) + E_s'(\rho_{\gamma_{\ell-1}})(\rho - \rho_{\gamma_{\ell-1}}) & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_{\ell-1}}}, \end{cases} \tag{A.115}$$

where we recall that since E_s function is a convex function in the region of $\rho \in [0,1]$, the tangent line at $\rho_{\gamma_{\ell'}}$ lies below E_s. In addition, since $\frac{1}{1+\rho_{\gamma_\ell}} < \frac{1}{1+\rho_{\gamma_{\ell'}}}$

and the tangent line at ρ_{γ_ℓ} is above the tangent line at $\rho_{\gamma_{\ell'}}$ for $\frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_\ell}}$. As a result, by comparing (A.114) and (A.115), we conclude that for all $\rho \in [0,1]$ we have $E_{s,\ell-1}(\rho, P_U) \geq E_{s,\ell'-1}(\rho, P_U)$.

To prove $E_{s,\ell}(\cdot) \leq E_{s,\ell'}(\cdot)$ for $\gamma_\ell \leq \gamma_{\ell'}$, similarly, for $\gamma_{\ell+1} \leq \gamma_\ell \leq \gamma_{\ell-1}$ we have

$$E_{s,\ell}(\rho, P_U, \gamma_{\nu,\ell+1}, \gamma_{\nu,\ell}) = \begin{cases} E_s(\rho_{\gamma_{\ell+1}}, P_U) + E_s'(\rho_{\gamma_{\ell+1}})(\rho - \rho_{\gamma_{\ell+1}}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_{\ell+1}}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_\ell}}, \\ E_s(\rho_{\gamma_\ell}, P_U) + E_s'(\rho_{\gamma_\ell})(\rho - \rho_{\gamma_\ell}) & \frac{1}{1+\rho_{\gamma_\ell}} < \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell'}}}, \\ E_s(\rho_{\gamma_\ell}, P_U) + E_s'(\rho_{\gamma_\ell})(\rho - \rho_{\gamma_\ell}) & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_{\ell'}}}, \end{cases} \tag{A.116}$$

and for $\gamma_{\ell+1} \leq \gamma_{\ell'} \leq \gamma_{\ell-1}$

$$E_{s,\ell'}(\rho, P_U, \gamma_{\nu,\ell'+1}, \gamma_{\nu,\ell'}) = \begin{cases} E_s(\rho_{\gamma_{\ell+1}}, P_U) + E_s'(\rho_{\gamma_{\ell+1}})(\rho - \rho_{\gamma_{\ell+1}}) & \frac{1}{1+\rho} < \frac{1}{1+\rho_{\gamma_{\ell+1}}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_{\ell+1}}} \leq \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_\ell}}, \\ E_s(\rho, P_U) & \frac{1}{1+\rho_{\gamma_\ell}} < \frac{1}{1+\rho} \leq \frac{1}{1+\rho_{\gamma_{\ell'}}}, \\ E_s(\rho_{\gamma_{\ell'}}, P_U) + E_s'(\rho_{\gamma_{\ell'}})(\rho - \rho_{\gamma_{\ell'}}) & \frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_{\ell'}}}, \end{cases} \tag{A.117}$$

where since $\frac{1}{1+\rho_{\gamma_\ell}} < \frac{1}{1+\rho_{\gamma_{\ell'}}}$ for $\frac{1}{1+\rho} > \frac{1}{1+\rho_{\gamma_{\ell'}}}$ the tangent line of E_s at $\rho_{\ell'}$ is above of the tangent line of E_s at ρ_ℓ, meaning that $E_{s,\ell}(\rho, P_U) \leq E_{s,\ell'}(\rho, P_U)$. □

Lemma A.16. *Let $i = 1,2$, for a given source probability distribution P_U and some $\gamma \in [0,1]$. Then, we have that*

$$\min_{\lambda \geq 0} \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})} \right)^{\frac{(-1)^i \lambda}{1+\rho}} \right)^{1+\rho} = e^{E_{s,i}(\rho, P_U^n, \gamma^n)}, \tag{A.118}$$

where $E_{s,i}(\rho, P_U, \gamma)$ for $i = 1,2$ is given by (2.33) and (2.34).

Proof. In order to prove (A.118), we may note that since the objective function in (A.118) is convex with respect to λ, the optimal $\lambda^\star$ satisfies

$$\frac{\partial}{\partial \lambda} \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}} \left(\frac{\gamma^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})} \right)^{\frac{(-1)^i \lambda}{1+\rho}} \right)^{1+\rho} \Bigg|_{\lambda^\star \geq 0} = 0. \tag{A.119}$$

This leads to

$$\frac{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-(-1)^i\lambda^\star}{1+\rho}} \log(P_{\boldsymbol{U}}^n(\boldsymbol{u}))}{\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1-(-1)^i\lambda^\star}{1+\rho}}} = \log(\gamma^n). \tag{A.120}$$

It is convenient to define ρ_γ through the implicit equation

$$\frac{1-(-1)^i\lambda^\star}{1+\rho} = \frac{1}{1+\rho_\gamma}. \tag{A.121}$$

When the solution to (A.120) is strictly negative, i. e. when

$$(-1)^i\left(\frac{1}{1+\rho} - \frac{1}{1+\rho_\gamma}\right) < 0, \tag{A.122}$$

we have $\lambda^\star = 0$, and hence (A.118) simplifies to

$$\left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}}\left(\frac{\gamma^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})}\right)^{\frac{(-1)^i\lambda}{1+\rho}}\right)^{1+\rho}\Bigg|_{\lambda=0} = \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}}\right)^{1+\rho} = e^{E_s(\rho, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}. \tag{A.123}$$

Otherwise, when the solution to (A.120) is non-negative, i. e. when

$$(-1)^i\left(\frac{1}{1+\rho} - \frac{1}{1+\rho_\gamma}\right) \geq 0 \tag{A.124}$$

and using (A.121), the left hand side of (A.118) satisfies

$$\min_{\lambda\geq 0}\left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}}\left(\frac{\gamma^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})}\right)^{\frac{(-1)^i\lambda}{1+\rho}}\right)^{1+\rho} = \left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_\gamma}}\right)^{1+\rho}\gamma^{n\frac{\rho_\gamma-\rho}{1+\rho}} = e^{(1+\rho)\log\left(P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho_\gamma}}\right)}\gamma^{n\frac{\rho_\gamma-\rho}{1+\rho_\gamma}}, \tag{A.125}$$

where we used $a^{(1+\rho)} = e^{(1+\rho)\log(a)}$. Using (A.121) into (A.120), we may express γ^n in terms of the $E_s(\cdot)$ function and its derivative $E_s'(\cdot)$ as

$$\gamma^n = e^{E_s(\rho_\gamma, P_{\boldsymbol{U}}^n(\boldsymbol{u})) - (1+\rho_\gamma)E_s'(\rho_\gamma, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}, \tag{A.126}$$

Inserting the right hand side of (A.126) into (A.125), we obtain

$$\min_{\lambda\geq 0}\left(\sum_{\boldsymbol{u}} P_{\boldsymbol{U}}^n(\boldsymbol{u})^{\frac{1}{1+\rho}}\left(\frac{\gamma^n}{P_{\boldsymbol{U}}^n(\boldsymbol{u})}\right)^{\frac{(-1)^i\lambda}{1+\rho}}\right)^{1+\rho} = e^{E_s(\rho_\gamma, P_{\boldsymbol{U}}^n(\boldsymbol{u})) - (\rho-\rho_\gamma)E_s'(\rho_\gamma, P_{\boldsymbol{U}}^n(\boldsymbol{u}))}. \tag{A.127}$$

Finally, combining (A.123) and (A.127) respectively for (A.122) and (A.124), and using the definitions (2.33) and (2.34), we conclude the proof. □

Bibliography

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.

[2] A. Feinstein, “Error bounds in noisy channels without memory,” *IEEE Trans. Inf. Theory*, vol. 1, pp. 13–14, Sept. 1955.

[3] C. E. Shannon, “Certain results in coding theory for noisy channels,” *Inf. and Control*, vol. 1, pp. 6–25, Sept. 1957.

[4] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. MIT Press, 1961.

[5] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[6] F. Jelinek, *Probabilistic Information Theory-Discrete and Memoryless Models*. McGraw-Hill Book Company, 1968.

[7] Y. Zhong, F. Alajaji, and L. L. Campbell, “On the joint source-channel coding error exponent for discrete memoryless systems,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1450–1468, Apr. 2006.

[8] I. Csiszár, “Joint source-channel error exponent,” *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 315–328, 1980.

[9] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.

[10] R. Ahlswede, “Multi-way communication channels,” in *IEEE Int. Symp. Inf. Theory (ISIT)*, (Tsahkadsor, Armenian), Sept. 1971.

[11] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channel with correlated sources,” *Bell Syst. Tech. Journal.*, vol. 52, pp. 1037–1076, Sept. 1973.

[12] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, pp. 648–657, Nov. 1980.

[13] R. Gallager, "A perspective on multiple access channels," *IEEE Trans. Inf. Theory*, vol. 31, pp. 124–142, Mar. 1985.

[14] G. Dueck, "A note on the multiple access channel with correlated sources (corresp.)," *IEEE Trans. Inf. Theory*, vol. 27, pp. 232–235, Mar. 1981.

[15] I. Csiszár and J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems.* Cambridge university press, 2011.

[16] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, A. Martinez, and T. Koch, "A derivation of the source-channel error exponent using nonidentical product distributions," *IEEE Trans. Inf. Theory*, vol. 60, pp. 3209–3217, June 2014.

[17] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, pp. 405–417, July 1974.

[18] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. i," *Inf. and Control*, vol. 10, pp. 65–103, Jan. 1967.

[19] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. ii," *Inf. and Control*, vol. 10, pp. 522–552, May 1967.

[20] R. Gallager, *Fixed composition arguments and lower bounds to error probability*, 1994.

[21] I. Bocharova, A. Guillén i Fàbregas, B. Kudryashov, A. Martinez, A. Tauste Campo, and G. Vazquez-Vilar, "Multi-class source-channel coding," *IEEE Trans. Inf. Theory*, vol. 62, pp. 5093–5104, Sept. 2016.

[22] K. Fan, "Minimax theorems," *Proc. of the Nat. Acad. Sci.*, vol. 39, pp. 42–47, Jan. 1953.

[23] H. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.

[24] M. Sion, "On general minimax theorems," *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, 1958.

[25] Y.-S. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 376–386, Mar. 1996.

[26] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1439–1452, Sept. 1996.

[27] J. Scarlett, "Reliable communication under mismatched decoding," PhD thesis, University of Cambridge, Cambridge, Aug. 2014.

[28] E. Haim, Y. Kochman, and U. Erez, "Improving the MAC error exponent using distributed structure," in *IEEE Int. Symp. Inf. Theory (ISIT)*, (St. Petersburg, Russia), July 2011.

[29] N. Cai, "The maximum error probability criterion, random encoder, and feedback, in multiple input channels," *Entropy*, vol. 16, no. 3, pp. 1211–1242, 2014.

[30] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," *IEEE Trans. Inf. Theory*, vol. 60, pp. 5095–5115, Sept. 2014.

[31] A. Nazari, S. S. Pradhan, and A. Anastasopoulos, "Error exponent for multiple access channels: Upper bounds," *IEEE Trans. Inf. Theory*, vol. 61, pp. 3605–3621, July 2015.

[32] L. Farkas and T. Kói, "Random access and source-channel coding error exponents for multiple access channels," *IEEE Trans. Inf. Theory*, vol. 61, pp. 3029–3040, June 2015.

[33] A. Rezazadeh, J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Multiple-access channel with independent sources: Error exponent analysis," in *2018 Information Theory Workshop*, (Guangzhou, China), Nov. 2018.

[34] A. Padakandla, "Communicating correlated sources over a MAC," in *IEEE Int. Symp. Inf. Theory (ISIT)*, (Aachen, Germany), June 2017.

[35] M. Heidari, F. Shirani, and S. S. Pradhan, "New sufficient conditions for multiple-access channel with correlated sources," in *IEEE Int. Symp. Inf. Theory (ISIT)*, (Barcelona, Spain), July 2016.

[36] A. Rezazadeh, J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "An achievable error exponent for the multiple access channel with

correlated sources," in *IEEE Int. Symp. Inf. Theory (ISIT)*, (Aachen, Germany), June 2017.

[37] A. Rezazadeh, J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Joint source-channel coding for the multiple-access channel with correlated sources," in *IEEE Int. Symp. Inf. Theory (ISIT)*, (Paris, France), July 2019.

[38] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser coding techniques for mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 62, pp. 3950–3970, July 2016.

[39] Y. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 376–386, Mar. 1996.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley & Sons, 2nd ed., 2006.

[41] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Ensemble-tight error exponents for mismatched decoders," in *50th Annual Allerton Conference on Communication, Control and Computing*, (Allerton, IL), Nov. 2012.

[42] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 169, pp. 217–220, 1997.

[43] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels," *Probl. Control Inf. Theory*, vol. 7, pp. 11–19, Jan. 1978.

[44] R. Goldberg, *Methods of Real Analysis.* Blaisdell Pub. Co., 1964.