Joint Source-Channel Coding Reliability Function for Single and Multi-Terminal Communication Systems

Yangfan Zhong

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements
for the degree of

Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
May 2008

Dedicated to

my parents

Abstract

Traditionally, source coding (data compression) and channel coding (error protection) are performed separately and sequentially, resulting in what we call a tandem (separate) coding system. In practical implementations, however, tandem coding might involve a large delay and a high coding/decoding complexity, since one needs to remove the redundancy in the source coding part and then insert certain redundancy in the channel coding part. On the other hand, joint source-channel coding (JSCC), which coordinates source and channel coding or combines them into a single step, may offer substantial improvements over the tandem coding approach.

This thesis deals with the fundamental Shannon-theoretic limits for a variety of communication systems via JSCC. More specifically, we investigate the reliability function (which is the largest rate at which the coding probability of error vanishes exponentially with increasing blocklength) for JSCC for the following discrete-time communication systems: (i) discrete memoryless systems; (ii) discrete memoryless systems with perfect channel feedback; (iii) discrete memoryless systems with source side information; (iv) discrete systems with Markovian memory; (v) continuous-valued (particularly Gaussian) memoryless systems; (vi) discrete asymmetric 2-user source-channel systems.

For the above systems, we establish upper and lower bounds for the JSCC reliability function and we analytically compute these bounds. The conditions for which the upper and lower bounds coincide are also provided. We show that the conditions are satisfied for a large class of source-channel systems, and hence exactly determine the reliability function. We next provide a systematic comparison between the JSCC reliability function

and the tandem coding reliability function (the reliability function resulting from separate source and channel coding). We show that the JSCC reliability function is substantially larger than the tandem coding reliability function for most cases. In particular, the JSCC reliability function is close to twice as large as the tandem coding reliability function for many source-channel pairs. This exponent gain provides a theoretical underpinning and justification for JSCC design as opposed to the widely used tandem coding method, since JSCC will yield a faster exponential rate of decay for the system error probability and thus provides substantial reductions in complexity and coding/decoding delay for real-world communication systems.

Declaration

The work in this thesis is based on research carried out at the Department of Mathematics and Statistics, Queen's University. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my work conducted jointly with my supervisor Prof. Fady Alajaji and co-supervisor Prof. L. Lorne Campbell unless referenced to the contrary in the text.

Acknowledgements

I have been extremely fortunate to have had Professor Fady Alajaji as my supervisor during my five years of graduate studies. His enthusiasm for top quality research and dedication to both my professional and personal developments are truly inspiring. It has been a great honor to be co-supervised by Professor L. Lorne Campbell. I have benefited enormously from his comments and suggestions on most of the chapters in this thesis. I sincerely thank them for their great guidance and continuous support throughout my graduate studies at Queen's University.

I am grateful to my thesis committee members, Professor En-hui Yang from the University of Waterloo, Professor Il-Min Kim from the Department of Electrical and Computer Engineering, Professor Tamás Linder, and Professor Ernst Kani, for their comments and suggestions with respect to this thesis. In particular, I would like to thank Professor Linder for his guidance in parts of my research. I thoroughly enjoyed our collaboration on watermarking problems (which do not appear in this thesis).

I want to thank all my friends and groupmates for their support and good cheer. Special thanks go to our graduate secretary, Jennifer Read, whose attention to detail and efficiency makes everything run smoothly.

Last but not least, I am deeply thankful to my parents and my brother for their unconditional love and support.

Bibliographic Note

Parts of Chapter 3 were presented at the 23nd Biennial Symposium on Communications [108] and 2007 IEEE International Symposium on Information Theory [113].

Parts of Chapter 4 appears in [109].

Chapter 5 was presented in parts at the 2004 IEEE International Symposium on Information Theory [103], and appears in [106] and [107].

Parts of Chapter 6 were presented at the 10th Canadian Workshop on Information Theory [112].

Chapter 7 was presented in parts at the 2005 IEEE International symposium on Information Theory [105] and appears in [109].

Chapter 8 was presented in parts at the 2006 IEEE International Symposium on Information Theory [110].

Chapter 9 was presented in parts at the 2007 IEEE International Symposium on Information Theory [113].

Chapter 10 was presented in parts at the the 22nd Biennial Symposium on Communications [102] and the 9th Canadian Workshop on Information Theory [104] and appears in [106], [107], and [109].

Contents

Chapter 1

Introduction

Traditionally, source and channel coding have been treated separately, resulting in what we call a *tandem (or separate)* coding system. This is because Shannon in 1948 [83] showed that separate source and channel coding incurs no loss of optimality (in terms of reliable transmissibility) provided that the coding blocklength goes to infinity. In practical implementations, however, there is a price to pay in delay and complexity, for extremely long blocklength. To begin, we note that *joint source-channel coding* (JSCC) might be expected to offer improvements for the combination of a source with significant redundancy and a channel with significant noise, since, for such a system, tandem coding would involve source coding to remove redundancy and then channel coding to insert redundancy. It is a natural conjecture that this is not the most efficient approach (even if the blocklength is allowed to grow without bound). Indeed, Shannon [83] made this point as follows:

> $\cdots$ *However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise. For example, in a noiseless telegraph channel one could save about* 50% *in time by proper encoding of the messages. This is not done and most of the redundancy of English remains in the channel symbols. This has the advantage, however, of allowing considerable noise in the channel. A*

> *sizable fraction of the letters can be received incorrectly and still reconstructed by the context. In fact this is probably not a bad approximation to the ideal in many cases* $\cdots$

The study of JSCC dates back to as early as the 1960's. Over the years, many works have introduced JSCC techniques and illustrated (analytically or numerically) their benefits (in terms of both performance improvement and increased robustness to variations in channel noise) over tandem coding for given source and channel conditions and fixed complexity and/or delay constraints. In JSCC systems, the designs of the source and channel codes are either well coordinated or combined into a single step. Examples of (both constructive and theoretical) previous lossless and lossy JSCC investigations include:

1. coding theorems on JSCC and separation principle [24], [36], [42], [46], [49], [50], [95];
2. source codes that are robust against channel errors such as optimal (or sub-optimal) quantizer design for noisy channels [5], [13], [39], [40], [45], [61], [62], [64], [68], [74], [89], [90], [93];
3. channel codes that exploit the source's natural redundancy (if no source coding is applied) or its residual redundancy (if source coding is applied) [4], [47], [67], [81], [114];
4. zero-redundancy channel codes with optimized codeword assignment for the transmission of source encoder indices over noisy channels (e.g., [39], [99]);
5. unequal error protection source and channel codes where the rates of the source and channel codes are adjusted to provide various levels of protection to the source data depending on its level of importance and the channel conditions (e.g., [51], [71]);
6. uncoded source-channel matching where the source is uncoded, directly matched to the channel and optimally decoded (e.g., [3], [44], [87], [98]).

The above references are far from exhaustive as the field of JSCC has been quite active, particularly over the last 20 years.

In order to learn more about the performance of the best codes as a function of blocklength, much research has focused on the *reliability function* for source or channel coding (see, e.g., [19], [32], [42], [57], [66], [97]). Throughout the thesis, the reliability function refers to either the *error exponent* of (asymptotically) lossless coding or the *excess distortion exponent* of lossy coding. Roughly speaking, the error exponent (respectively, the excess distortion exponent) E is a number with the property that the probability of decoding error (respectively, the probability of exceeding a prescribed distortion level) of a good code is approximately 2^{-En} for codes of large blocklength n. Thus the error exponent (respectively, the excess distortion exponent) can be used to estimate the trade-off between probability of error (respectively, probability of excess distortion) and blocklength; in such way, we can use the reliability function as a tool to compare the performance of tandem coding and JSCC. While jointly coding the source and channel offers no advantages over tandem coding in terms of reliable transmissibility of the source over the channel (i.e., for the case of memoryless systems as well as the wider class of information stable [49] single-user systems), it is possible that the same error performance can be achieved for smaller blocklengths via optimal JSCC coding.

The JSCC reliability function has only been partially studied in the past. The first quantitative result on the JSCC reliability function for communication systems consisting of a discrete source and a discrete channel was a lower bound on the (lossless) JSCC error exponent derived in 1964 by Gallager [42, pp. 534–535]. This result also indicates that JSCC can lead to a larger exponent than the tandem coding exponent, the exponent resulting from separately performing and concatenating optimal source and channel coding. In 1980, Csiszár [30] established a lower bound (based on the random-coding channel error exponent) and an upper bound (in terms of source and channel error exponents) for the JSCC error exponent for a discrete memoryless source (DMS) and a discrete memoryless channel (DMC). He pointed out that the bounds are tight for a large class of DMS-DMC pairs, hence determining the JSCC error exponent exactly. He extended this work in 1982 [31] to obtain a new expurgated lower bound (based on the expurgated channel exponent) for the above system under some conditions, and to deal with lossy coding relative to a distortion threshold.

In practical applications, however, not just DMS-DMC systems are treated. As most real-world data sources (e.g., multimedia sources) and communication channels (e.g., wireless channels) exhibit statistical dependency or memory, it is of natural interest to study the JSCC reliability function for systems with memory, since the determination of the reliability function (or its bounds), particularly in terms of computable parametric expressions, may lead to the identification of important information-theoretic design criteria for the construction of powerful JSCC techniques that fully exploit the source-channel memory. In addition, since there are a lot of real-world communication systems dealing with the compression and transmission of analog signals instead of digital data, it is natural and important to study the JSCC reliability function for the transmission of a continuous alphabet source over a channel with continuous input/output alphabets. For instance, it is of interest to know the best performance (e.g., excess distortion probability) that a source-channel code can achieve if a stationary memoryless Gaussian source is coded and transmitted through an additive white Gaussian noise channel. On the other hand, with the rapid development of wired and wireless communication networks, increasing attentions are drawn to JSCC for multi-terminal source-channel systems. Therefore, it is of interest to study the JSCC reliability function for multi-terminal systems.

Following Csiszár's work, we study the JSCC reliability function in the thesis for various single-user source-channel systems and the asymmetric 2-user source-channel system. To explore the potential advantages of JSCC over traditional tandem coding, we next provide a systematic comparison between the JSCC reliability function and the tandem coding reliability function. We demonstrate that JSCC substantially outperforms tandem coding in terms of reliability function for most source-channel systems.

1.1 General Overview

The structure of the thesis is shown in Fig. 1.1. Each arrow "→" represents a logic flow from chapter to chapter. Chapters 2 is a review chapter for the known results on the source and channel reliability functions. Chapters 3 and 4 provide technical background for the thesis. Chapter 3 is used for establishing upper and lower bounds for the JSCC reliability function, and Chapter 4 plays a central role for the analysis of these bounds. Chapters 5, 6, 7, and 8 deal with the JSCC reliability functions for single-user (point-to-point) systems. More specifically, we study analytical computation of Csiszár's lower and upper bounds for the JSCC reliability function for discrete memoryless systems in Chapter 5. Chapter 6 is an extension of Chapter 5, where the JSCC reliability functions for discrete memoryless systems with feedback and source side information are investigated. In Chapters 7 and 8, we establish and analyze lower and upper bounds for the JSCC reliability functions for discrete systems with Markovian memory and memoryless systems with continuous alphabets, respectively. Chapter 9 deals with the reliability function for multi-terminal systems consisting of two correlated sources and an asymmetric 2-user channel. In Chapter 10, we study the benefits of JSCC over tandem coding in terms of the reliability function for the single-user systems treated in Chapters 5–8 and a class of multi-user systems consisting of two correlated sources and an asymmetric multiple-access channel addressed in Chapter 9. Chapter 11 is the conclusion chapter.

1.2 Chapter By Chapter Overview

In Chapter 2, we review the basic material regarding the source and channel reliability functions. The topics covered in Chapter 2 are: source error exponent and source excess distortion exponent for memoryless sources, channel error exponent with or without an input cost constraint, random-coding exponent, expurgated exponent, sphere-packing exponent, and the important bounds for the channel error exponent. In particular, we separately discuss the source excess distortion exponent under a squared-error distortion for the memoryless Gaussian source (MGS) and the channel error exponent with a quadratic power input constraint for the memoryless channel with additive Gaussian noise (which we refer to as the memoryless Gaussian channel, MGC).

Chapter 3 contains the background on the method of types which will be used to establish

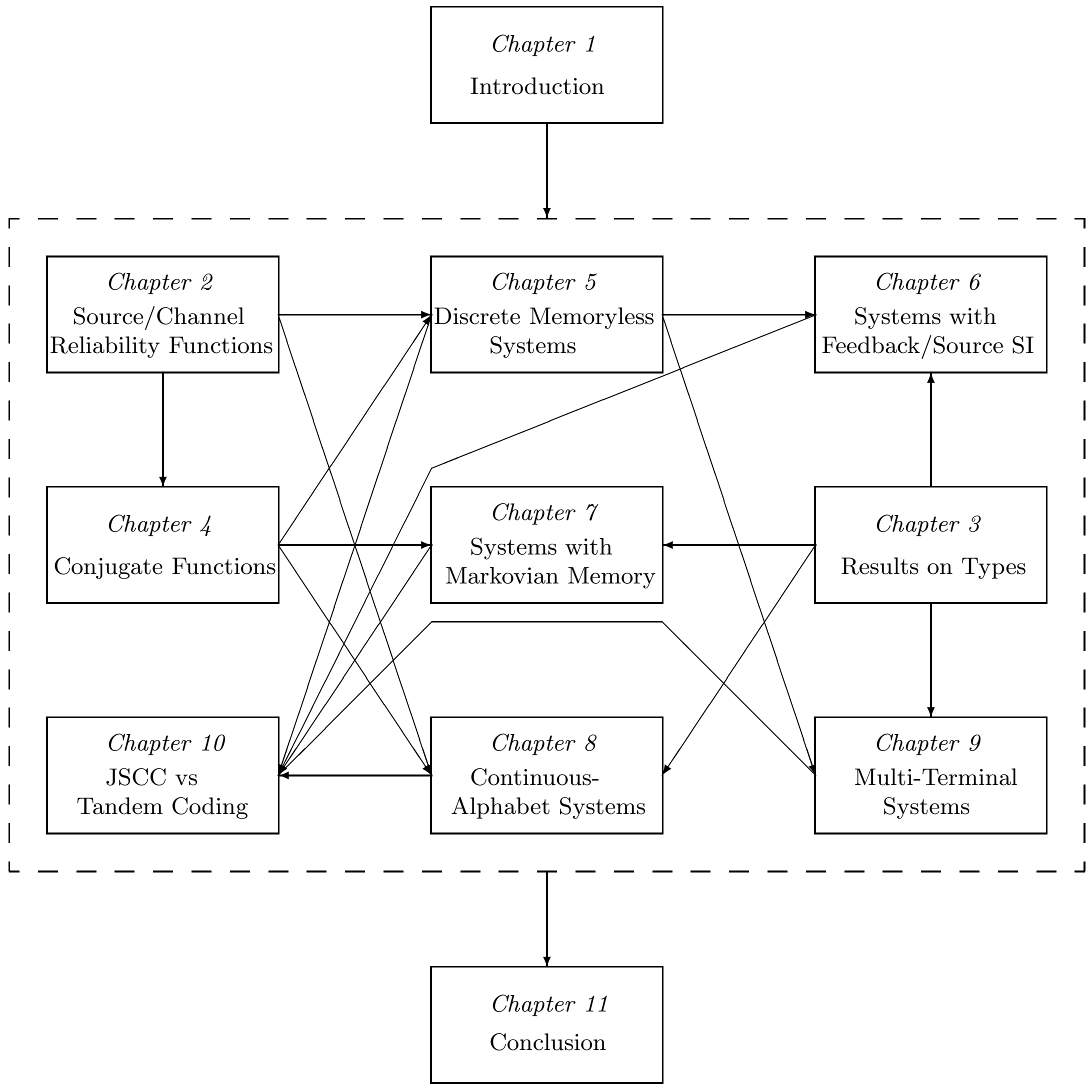

Figure 1.1: Organization of the thesis.

upper and lower bounds for the JSCC reliability function in later chapters. We first go over the basic definitions and properties of discrete types, joint types and conditional types as well as the corresponding type classes (type sets). We then develop a generalized joint type packing lemma. Two continuous type classes, the Gaussian-type class and the Laplacian-type class are introduced. Finally, we present type covering lemmas for the discrete type set, the Gaussian-type class, and the Laplacian-type class.

In Chapter 4, we introduce conjugate functions (which are also called convex/concave Fenchel transforms, or convex/concave Fenchel-Legendre transforms in the literature) and the Fenchel duality theorem (which will be used widely in later chapters to obtain equivalent (dual) forms for the JSCC reliability function). We apply these conjugacy properties on the source and channel reliability functions, and we show that the source/channel reliability functions and the corresponding source/channel functions are actually related by pairs of Fenchel transforms.

In Chapter 5, we examine the computation of Csiszár's lower and upper bounds for the JSCC error exponent of a communication system consisting of a DMS and a DMC by using the results obtained in Chapter 4. For Csiszár's JSCC random-coding bound and JSCC sphere-packing bound, we provide equivalent expressions for these bounds which can be readily computed for arbitrary source-channel pairs via Arimoto's algorithm. We derive sufficient and necessary conditions for which the bounds coincide. These conditions are satisfied by a large class of DMS-DMC pairs, and hence determine the exponent exactly. When the channel's distribution satisfies a symmetry property, the bounds admit closed-form parametric expressions. We also treat Csiszár's JSCC expurgated bound by using a similar approach, and we derive a sufficient and necessary condition for which the expurgated bound is strictly larger than the JSCC random-coding bound. We then examine this condition to DMS's and equidistant channels. Finally, we study the computation of the JSCC excess distortion exponent under the Hamming distortion measure.

One may ask whether the reliability of transmission of an information source over a communication channel can be enhanced by feedback of the channel output or side information related to the source. In Chapter 6, we study the JSCC error exponent for discrete memoryless systems with perfect (noiseless and instantaneous) channel output feedback and source side information (SI) at decoder, respectively. It is seen that feedback does not affect the reliable transmission region of a DMS over a DMC, i.e., the JSCC theorem is the same when there is feedback. Nevertheless, we show that feedback can enlarge the JSCC error exponent. In particular, we establish upper and lower bounds for the JSCC error exponent with feedback. A sufficient condition for which the exponent is determined exactly is presented for binary input channels with a symmetric distribution (in the Gallager sense). By numerically comparing this lower bound with the upper bound for the JSCC error exponent without feedback, it is demonstrated that the JSCC error exponent can be obviously increased in the presence of feedback. For the system with source SI at decoder, we employ the method of types to establish a lower bound for the JSCC error exponent. As a consequence, a JSCC theorem on the reliable transmissibility of the source over the channel is obtained. For binary sources and symmetric channels, we derive a sufficient condition for which the SI at the decoder can strictly improve the JSCC error exponent. Numerical results show that the availability of the SI at decoder can enlarge the region for reliable transmissibility and increase the JSCC error exponent for a wide class of source-channel parameters.

In Chapter 7, we investigate the JSCC error exponent for reliably transmitting a discrete stationary ergodic Markov (SEM) source over a discrete channel with additive SEM noise (which is referred to as the SEM channel). We first establish a sphere-packing type upper bound for the JSCC error exponent in terms of the Rényi entropy rates of the source and noise processes. We next investigate the analytical computation of the exponent by comparing our bound with Gallager's lower bound when the latter one is specialized to the SEM source-channel system. We also note that both bounds can be represented in Csiszár's form, as the minimum of the sum of the source and channel error exponents. It is seen that the JSCC error exponent can be exactly determined by the two bounds for a large class of SEM source-channel pairs. As for the discrete memoryless case, a conceptual upper bound for the JSCC error exponent in terms of SEM source and channel error exponents is established by using Markov types. This upper bound might not be computable, but it is useful for the comparison of the JSCC error exponent with the tandem coding error exponent, which is addressed in Chapter 10.

Chapter 8 deals with the JSCC excess distortion exponent for memoryless communication systems with continuous alphabets. We first establish upper and lower bounds for the

JSCC excess distortion exponent for systems consisting of an MGS under the squared-error distortion fidelity criterion and an MGC with a quadratic power constraint at the channel input. A sufficient and necessary condition for which the two bounds coincide is provided, thus exactly determining the exponent. This condition is observed to hold for a wide range of source-channel parameters. The extension for the bounds to transmitting memoryless Laplacian sources over the MGC under the magnitude-error distortion is next carried out. We also establish a lower bound for the JSCC excess distortion exponent for a certain class of continuous source-channel pairs when the distortion measure is a metric.

In Chapter 9, we study the exponential behavior of the probabilities of error for certain multi-terminal systems. We consider transmitting two discrete memoryless correlated sources (CS), consisting of a common and a private source, over a discrete memoryless multi-terminal channel with two transmitters and two receivers. At the transmitter side, the common source is observed by both encoders but the private source can only be accessed by one encoder. At the receiver side, both decoders need to reconstruct the common source, but only one decoder needs to reconstruct the private source. We hence refer to this system by the asymmetric 2-user source-channel system. We derive a universally achievable JSCC error exponent pair for the 2-user system by using the method of types. We next investigate the largest convergence rate of asymptotic exponential decay of the system (overall) probability of erroneous transmission, i.e., the system JSCC error exponent. We obtain lower and upper bounds for the exponent. As a consequence, we establish the JSCC theorem with single letter characterization and we show that the separation principle holds for the asymmetric 2-user scenario. We next specialize our results to systems consisting of two discrete memoryless CS and an asymmetric multiple-access channel (AMAC), and systems consisting of two discrete memoryless CS and an asymmetric broadcast channel (ABC). Finally, we evaluate the upper and lower bounds for the system JSCC error exponent for the CS-AMAC system when the channel admits some symmetric distribution. It is shown that the upper and lower bounds coincide for many binary CS-AMAC pairs.

The advantage of JSCC over traditional tandem coding in terms of the reliability function is explored in Chapter 10. We first derive a formula for the tandem coding error exponent, which applies if the source and channel are separately coded, for arbitrary discrete systems. We then use our results to provide a systematic comparison between the JSCC error exponent and the tandem error exponent for discrete memoryless systems and discrete systems with Markovian memory (SEM systems). It is shown that the JSCC exponent can at most double the tandem coding exponent. We establish conditions for which the JSCC exponent is strictly larger than the tandem error exponent. Numerical examples indicate that the JSCC exponent substantially outperforms the tandem coding exponent (particularly, the JSCC exponent is close to twice of the tandem coding exponent) for a large class of DMS-DMC pairs and SEM source-channel pairs. This gain translates into a power saving larger than 2 dB for a binary source transmitted over additive white Gaussian noise channels and Rayleigh fading channels with finite output quantization. As an extension, we show that our formula for tandem exponent remains valid for discrete memoryless systems with channel output feedback and source SI, and the joint exponent is superior to the corresponding tandem exponent for many cases. We also establish a formula for the tandem coding excess distortion exponent for Gaussian systems with squared-error distortion measure. By numerically comparing the lower bound of the joint exponent and the upper bound of the tandem exponent, it is observed that, as for the discrete systems, JSCC considerably outperforms tandem coding for many MGS-MGC pairs. For the asymmetric 2-user coding scenario studied in Chapter 9, we derive a formula for the tandem coding error exponent as for the point-to-point systems. Numerical examples show that for a large class of systems consisting of two correlated sources and an asymmetric multiple-access channel with additive noise, the JSCC error exponent, as for the point-to-point systems, considerably outperforms the corresponding tandem coding error exponent.

Chapter 11 provides a summary of the thesis contributions and contains directions for future research.

Chapter 2

Preliminaries: Source and Channel Reliability Functions

This chapter contains basic material on source and channel coding. Fundamental information quantities such as entropy and mutual information, and source and channel reliability functions will be introduced.

In Section 2.1, we first give an overview for the notation and conventions which will be used throughout the thesis. We introduce the source error exponent and source excess distortion exponent in Section 2.2. We then address the channel error exponent (with and without input cost constraints) together with different types of lower and upper bounds in Section 2.3. As a special case, the excess distortion exponent for the memoryless Gaussian source (MGS) and the error exponent for the memoryless Gaussian channel (MGC, namely, a continuous channel with additive memoryless Gaussian noise) is separately treated in Section 2.4. We finally draw a conclusion in Section 2.5.

2.1 Notation and Conventions

For any finite alphabet $\mathcal{X}$, the set of all probability distributions (probability mass functions (pmf)) on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$; for any finite alphabets $\mathcal{X}, \mathcal{Y}$, the set of all conditional distributions $V_{Y|X}$ is denoted by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. For any alphabet $\mathcal{X}$ and $k \in \mathbb{N}$, let the Cartesian product of k $\mathcal{X}$'s be denoted by $\mathcal{X}^k$. To simplify notation, $\mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{Y})$ can also be denoted by $\mathcal{P}(\mathcal{X}^2 \times \mathcal{Y}^2)$.

For finite alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ with joint distribution $P_{XYZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$, for simplicity we employ P_X, P_{XY}, $P_{YZ|X}$, etc, to denote the corresponding marginal and conditional probabilities induced by P_{XYZ} unless otherwise indicated. Conversely, $P_X P_{YZ|X}$ denotes a joint (product) distribution on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ with marginal distribution P_X and conditional distribution $P_{YZ|X}$.

Given distributions P_X and $W_{Y|X}$, let $P_X^{(n)}$ and $W_{Y|X}^{(n)}$ be their n-dimensional product distributions; in other words, $P_X^{(n)}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$ and $W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n W_{Y|X}(y_i|x_i)$, where $\mathbf{x} \triangleq (x_1, \cdots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} \triangleq (y_1, \cdots, y_n) \in \mathcal{Y}^n$. Note that $P_X^{(n)}$ ($W_{Y|X}^{(n)}$) is different from P_{X^n} ($W_{Y^n|X^n}$), where the later denotes a generic probability distribution on $\mathcal{X}^n$ (conditional distribution on $\mathcal{X}^n \times \mathcal{Y}^n$).

For any finite set $\mathcal{X}$, the size of $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The expectation of the random variable (RV) X under the distribution P_X is denoted by $\mathbb{E}_{P_X}(X)$ or $\mathbb{E}(X)$ if P_X is clear from the context. For a given set A, A^c denotes the complement of A. Given a matrix (vector) A, A^t denotes its transposition.

When we say that the probability of some events, say $\Pr(A)$, is taken under a pmf or a probability density function (pdf) P_{X^n} on $\mathcal{X}^n$, this can be interpreted as $\Pr(A) = \sum_{\mathcal{X}^n} P_{X^n}(\mathbf{x}) \mathbb{1}\{A\}$ or $\Pr(A) = \int_{\mathcal{X}^n} P_{X^n}(\mathbf{x}) \mathbb{1}\{A\} d\mathbf{x}$, respectively, where $\mathbb{1}\{\cdot\}$ is the indicator function. By convention, we define throughout the thesis $0 \log_2 0 \triangleq 0$, $\log_2 0 = -\infty$, $\frac{x}{0} \triangleq +\infty$ for $x > 0$, and $\inf\{x : x \in \emptyset\} = +\infty$ unless otherwise indicated, where $\emptyset$ denotes the empty set.

2.2 Source Reliability Function

Let Q_S be a (stationary) memoryless source with finite alphabet $\mathcal{S}$ and a generic pmf $Q_S \in \mathcal{P}(\mathcal{S})$. The probability distribution of a k-length source sequence $\mathbf{s} \triangleq (s_1, s_2, ..., s_k) \in \mathcal{S}^k$ is given by $Q_{S^k}(\mathbf{s}) = Q_S^{(k)}(\mathbf{s}) = \prod_{i=1}^k Q_S(s_i)$.

2.2.1 Error Exponent for DMSs

Given a discrete memoryless source (DMS) Q_S, the entropy of the source is given by (e.g., [29])

$$H_{Q_S}(S) = -\sum_{s\in\mathcal{S}} Q_S(s)\log_2 Q_S(s).$$

We will simply write the entropy as $H(S)$ if Q_S is clear from the context.

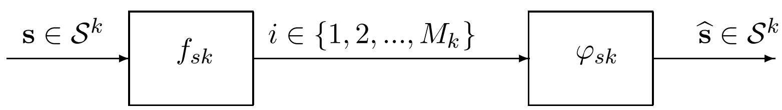

Figure 2.1: Memoryless source coding system.

A (k, M_k) block source code for a DMS Q_S is a pair of mappings (see Fig. 2.1):

$$f_{sk} : \mathcal{S}^k \longrightarrow \{1, 2, ..., M_k\}$$

and

$$\varphi_{sk} : \{1, 2, ..., M_k\} \longrightarrow \mathcal{S}^k.$$

The code rate is defined by

$$R_k \triangleq \frac{1}{k}\log_2 M_k \quad \text{bits/source symbol.}$$

The probability of erroneously reconstructing the source via the (k, M_k) block source code (f_{sk}, φ_{sk}) is given by

$$P_{se}^{(k)}(Q_S, R_k) \triangleq \sum_{\mathbf{s}:\mathbf{s}\neq\varphi_{sk}(f_{sk}(\mathbf{s}))} Q_S^{(k)}(\mathbf{s}). \tag{2.1}$$

We refer to $P_{se}^{(k)}$ by the probability of error for coding the source Q_S.

Shannon's lossless source coding theorem [32, 83] states that for a DMS Q_S, only $H_{Q_S}(S) + \varepsilon$ ($\varepsilon > 0$) bits per source symbol are needed to encode the source with arbitrarily small probability of error $P_{se}^{(k)}$, provided that the blocklength k of encoded source symbols is allowed to be sufficiently large. The (asymptotically) lossless source coding error exponent was developed to determine the asymptotics of the smallest possible probability of incorrect decoding as a function of the coding rate.

Definition 2.1 For any $R > 0$, the source error exponent $e(R, Q_S)$ of the DMS Q_S is defined as the supremum of the set of all numbers e for which there exists a sequence of (k, M_k) block codes (f_{sk}, φ_{sk}) with

$$e \leq \liminf_{k\to\infty} -\frac{1}{k} \log_2 P_{se}^{(k)}(Q_S, R_k) \tag{2.2}$$

and

$$R \geq \limsup_{k\to\infty} R_k. \tag{2.3}$$

For probability distributions $P_S, Q_S \in \mathcal{P}(\mathcal{S})$, denote the Kullback-Leibler divergence (relative entropy) between P_S and Q_S by (e.g., [29])

$$D(P_S \parallel Q_S) = \sum_{s\in\mathcal{S}} P_S(s) \log_2 \frac{P_S(s)}{Q_S(s)}.$$

It has been shown in [32, 63] by a combinatorial approach (the method of types) that the source error exponent for a DMS Q_S is equal to

$$e(R, Q_S) = \begin{cases} \min_{P_S: H_{P_S}(S)\geq R} D(P_S \parallel Q_S) & \text{if } 0 < R \leq \log_2 |\mathcal{S}|, \\ \infty & \text{if } R > \log_2 |\mathcal{S}|, \end{cases} \tag{2.4}$$

and the above exponent is shown to be universally achievable [32], i.e., (2.4) can be achieved by a sequence of source codes (f_{sk}, φ_{sk}) constructed without any knowledge of the source distribution Q_S. Immediately, it can be verified that $e(R, Q_S)$ is zero for $0 < R \leq H_{Q_S}(S)$, and is strictly increasing, convex and hence continuous in R for $H_{Q_S}(S) \leq R \leq \log_2 |\mathcal{S}|$. Thus, a precise formula for $e(R, Q_S)$ is given by

$$e(R, Q_S) = \begin{cases} 0 & \text{if } 0 < R \leq H_{Q_S}(S), \\ \min_{P_S: H_{P_S}(S)=R} D(P_S \parallel Q_S) & \text{if } H_{Q_S}(S) \leq R \leq \log_2 |\mathcal{S}|, \\ \infty & \text{if } R > \log_2 |\mathcal{S}|. \end{cases} \tag{2.5}$$

Furthermore, it has been shown in [19] that $e(R, Q_S)$ given by (2.5) involves an equivalent parametric form

$$e(R, Q_S) = \sup_{\rho\geq 0} [\rho R - E_s(\rho, Q_S)], \tag{2.6}$$

where

$$E_s(\rho, Q_S) \triangleq (1+\rho)\log_2 \sum_{s\in\mathcal{S}} Q_S(s)^{\frac{1}{1+\rho}} \tag{2.7}$$

is called Gallager's source function, as the parametric form of $e(R, Q_S)$ was first obtained by Gallager [42] (see also [57]).

2.2.2 Excess Distortion Exponent for Memoryless Sources

Let $d : \mathcal{S} \times \mathcal{S} \to [0, \infty)$ be a single-letter distortion function. The distortion measure on $\mathcal{S}^k \times \mathcal{S}^k$ is defined as

$$d^{(k)}(\mathbf{s}, \mathbf{s}') \triangleq \frac{1}{k}\sum_{i=1}^{k} d(s_i, s_i')$$

for any $\mathbf{s} \triangleq (s_1, ..., s_k) \in \mathcal{S}^k, \mathbf{s}' \triangleq (s_1', ..., s_k') \in \mathcal{S}^k$. For RV's X and Y admitting a joint pmf $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, the mutual information between X and Y is given by (e.g., [29])

$$I_{P_{XY}}(X;Y) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x,y)\log_2 \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}.$$

The mutual information will be simply written as $I(X;Y)$ if P_{XY} is clear from the context.

Given a distortion threshold $\Delta > 0$, the rate-distortion function for the DMS Q_S is given by (e.g., [15])

$$R(Q_S, \Delta) = \inf_{P_{S'|S}:\mathbb{E}d(S,S')\leq\Delta} I_{Q_S P_{S'|S}}(S; S'), \tag{2.8}$$

where the infimum is taken over all the conditional distributions $P_{S'|S} \in \mathcal{P}(\mathcal{S}|\mathcal{S})$ subject to $\mathbb{E}d(S, S') \leq \Delta$.

A (k, M_k) block source code for a DMS Q_S is a pair of mappings (see Fig. 2.1):

$$f_{sk} : \mathcal{S}^k \longrightarrow \{1, 2, ..., M_k\}$$

and

$$\varphi_{sk} : \{1, 2, ..., M_k\} \longrightarrow \mathcal{S}^k.$$

The code rate is defined by

$$R_k \triangleq \frac{1}{k}\log_2 M_k \quad \text{bits/source symbol.}$$

The probability of exceeding a given distortion threshold $\Delta > 0$ for the code $(f_{sk}, \varphi_{sk}, \Delta)$ is given by

$$P_\Delta^{(k)}(Q_S, R_k) \triangleq \sum_{\mathbf{s}: d^{(k)}(\mathbf{s}, \varphi_{sk}(f_{sk}(\mathbf{s}))) > \Delta} Q_S^{(k)}(\mathbf{s}). \tag{2.9}$$

Note that if Q_S is a pdf on the continuous alphabet $\mathcal{S} \subseteq \mathbb{R}$, then the summation is replaced by a proper integration, i.e.,

$$P_\Delta^{(k)}(Q_S, R_k) \triangleq \int_{\mathbf{s}: d^{(k)}(\mathbf{s}, \varphi_{sk}(f_{sk}(\mathbf{s}))) > \Delta} Q_S^{(k)}(\mathbf{s}) d\mathbf{s}. \tag{2.10}$$

We call $P_\Delta^{(k)}(Q_S, R_k)$ the probability of excess distortion for coding the source Q_S.

The lossy source coding theorem (e.g., [42]) for a memoryless source Q_S states that only $R(Q_S, \Delta) + \varepsilon$ ($\varepsilon > 0$) bits per source symbol are needed to reproduce the source within a distortion threshold Δ with arbitrarily small probability of exceeding the distortion threshold Δ, i.e., $P_\Delta^{(k)}(Q_S, R_k)$ asymptotically vanishes with the coding blocklength. The source coding excess distortion exponent describes the asymptotic behavior of the smallest possible probability of excess distortion as a function of the coding rate.

Definition 2.2 For any $R > 0$ and $\Delta > 0$, the excess distortion exponent $e_\Delta(R, Q_S)$ of the source Q_S is defined as the supremum of the set of all numbers e for which there exists a sequence of (k, M_k) block codes $(f_{sk}, \varphi_{sk}, \Delta)$ with

$$e \le \liminf_{k \to \infty} -\frac{1}{k} \log_2 P_\Delta^{(k)}(Q_S, R_k)$$

and

$$R \ge \limsup_{k \to \infty} R_k.$$

It has been shown in [54, 55, 108] that the excess distortion exponent for some particular sources can be expressed in Marton's form [66]. In other words,

$$e_\Delta(R, Q_S) = F(R, Q_S, \Delta) \triangleq \inf_{P_S: R(P_S, \Delta) > R} D(P_S \parallel Q_S), \tag{2.11}$$

holds for the following cases:

1. DMS's with arbitrary distortion measures [66];

2. Memoryless Gaussian sources (MGS's) with squared-error distortion measure [54];

3. (Stationary) memoryless sources whose alphabets are complete metric spaces with a metric distortion measure $d(\cdot,\cdot)$ under the condition that there exists an element $s_o \in \mathcal{S}$ with $\mathbb{E}\exp[td(s,s_o)] < \infty$ for all $t \in (-\infty,+\infty)$ [55].

Note that if P_S and Q_S are pdf's on the continuous alphabet $\mathcal{S} \subseteq \mathbb{R}$ such that P_S is absolutely continuous [53, p. 21] with respect to Q_S (denoted by $P_S \ll Q_S$), then the Kullback-Leibler divergence $D(P_S \parallel Q_S)$ should be interpreted as

$$D(P_S \parallel Q_S) = \int_{\mathcal{S}} P_S(s) \log_2 \frac{P_S(s)}{Q_S(s)} ds.$$

Note also that Cases 2 is not included by Case 3. First, the squared-error distortion is not a metric; second, the condition with respect to the metric and the source distribution does not hold for MGS's with squared-error distortion measure. It follows by definition that the function $F(R, Q_S, \Delta)$ should be an increasing function of R, however, unlike the source error exponent $e(R, Q_S)$, $F(R, Q_S, \Delta)$ is not necessarily convex or even continuous in R [2,32,66]. In fact, $F(R, Q_S, \Delta)$ is an increasing function of R with at most countably many discontinuities. We do not have a parametric form for $F(R, Q_S, \Delta)$ in general. When Q_S is an MGS with a squared-error distortion measure, the explicit analytical form of $F(R, Q_S, \Delta)$ will be given in Section 2.4.

2.3 Channel Reliability Function

Let $W_{Y|X}$ be a memoryless channel with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and probability transition distribution $W_{Y|X}$. If the channel has a continuous output alphabet $\mathcal{Y}$, we only consider continuous channels for which a conditional pdf exists. The conditional pmf (pdf) of receiving $\mathbf{y} \triangleq (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$ at the channel output given that the codeword $\mathbf{x} \triangleq (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ is transmitted is given by $W_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = W^{(n)}_{Y|X}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n W_{Y|X}(y_i|x_i)$.

2.3.1 Error Exponent for DMCs

Given a discrete memoryless channel (DMC) $W_{Y|X}$, the channel capacity is given by (e.g., [29])

$$C(W_{Y|X}) = \max_{P_X \in \mathcal{P}(\mathcal{X})} I_{P_X W_{Y|X}}(X;Y).$$

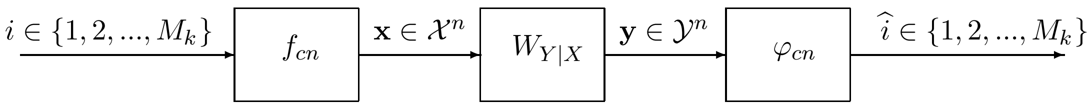

Figure 2.2: Memoryless channel coding system.

An (n, M_n) block channel code for a DMC $W_{Y|X}$ is a pair of mappings (see Fig. 2.2):

$$f_{cn} : \{1,2,...,M_n\} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_{cn} : \mathcal{Y}^n \longrightarrow \{1,2,...,M_n\}.$$

The code rate is defined as

$$R_n \triangleq \frac{1}{n} \log_2 M_n \quad \text{bits/channel use.}$$

The (average) probability and the maximum probability of decoding error for the (f_{cn}, φ_{cn}) code are respectively given by

$$P_{ec}^{(n)}(W_{Y|X}, R_n) \triangleq \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y})=j} W_{Y|X}^{(n)}(\mathbf{y}|f_{cn}(i)) \tag{2.12}$$

and

$$P_{max,ec}^{(n)}(W_{Y|X}, R_n) \triangleq \max_{1\leq i\leq M_n} \sum_{j=1, j\neq i}^{M_n} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y})=j} W_{Y|X}^{(n)}(\mathbf{y}|f_{cn}(i)). \tag{2.13}$$

We refer to $P_{ec}^{(n)}$ by the probability of error and to $P_{max,ec}^{(n)}$ by the maximum probability of error.

By Shannon's channel coding theorem [29, 83], block codes with arbitrarily small probability of block decoding error exist at any code rate smaller than the channel capacity

$C(W_{Y|X})$. Like the source exponent, the channel error exponent is a quantity that describes the relation between the rate of convergence or decay for the probability of error and the code rate (specifically for rates less than channel capacity).

Definition 2.3 For any $R > 0$, the channel error exponent $E(R, W_{Y|X})$ of the channel $W_{Y|X}$ is defined as the supremum of the set of all numbers E for which there exists a sequence of (n, M_n) block codes (f_{cn}, φ_{cn}) with

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{ec}^{(n)}(W_{Y|X}, R_n)$$

and

$$R \leq \liminf_{n\to\infty} R_n.$$

We remark that, by the definition of channel error exponent and an "expurgated codebook" argument, we can show that the probability of error $P_{ec}^{(n)}$ and the maximal probability of error $P_{max,ec}^{(n)}$ for channel coding lead to the same channel error exponent [85, p. 416]. Thus, an equivalent definition for the channel error exponent follows if $P_{ec}^{(n)}(W_{Y|X}, R_n)$ is replaced by $P_{max,ec}^{(n)}(W_{Y|X}, R_n)$ in the above.

Like channel capacity $C(W_{Y|X})$, $E(R, W_{Y|X})$ is a quantity that depends on the channel characteristics. By definition, $E(R, W_{Y|X})$ is a non-increasing function in $R \leq C(W_{Y|X})$ and is zero for $R > C(W_{Y|X})$, but unlike the source exponent, the error exponent of a DMC is not known for all R. The most familiar bounds to $E(R, W_{Y|X})$ are the random-coding and expurgated lower bounds due to Fano (1961) and Gallager (1965), and the sphere-packing and straight-line upper bounds due to Shannon-Gallager-Berlerkamp (1967) (see [32,42,57]). We next summarize the random-coding exponent, the expurgated exponent, sphere-packing exponent, and their corresponding channel functions.

2.3.2 Random-Coding Exponent

For a given DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and any $R > 0$, define the random-coding exponent for $W_{Y|X}$ by

$$E_r(R, W_{Y|X}) \triangleq \max_{P_X \in \mathcal{P}(\mathcal{X})} E_r(R, P_X, W_{Y|X}) \tag{2.14}$$

where

$$E_r(R, P_X, W_{Y|X}) \triangleq \min_{V_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \left[D(V_{Y|X} \parallel W_{Y|X}|P_X) + \left| I_{P_X V_{Y|X}}(X;Y) - R \right|^+ \right], \quad (2.15)$$

where $|x|^+ = \max\{0, x\}$ and

$$D(V_{Y|X} \parallel W_{Y|X}|P_X) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_X(x) V_{Y|X}(y|x) \log_2 \frac{V_{Y|X}(y|x)}{W_{Y|X}(y|x)}$$

is the Kullback-Leibler divergence between conditional distributions $V_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ conditional on distribution $P_X \in \mathcal{P}(\mathcal{X})$.

It has been shown that $E_r(R, P_X, W_{Y|X})$ has the following parametric form (e.g., [32])

$$E_r(R, P_X, W_{Y|X}) = \max_{0\le\rho\le1} [E_o(\rho, P_X, W_{Y|X}) - \rho R], \quad (2.16)$$

where

$$E_o(\rho, P_X, W_{Y|X}) \triangleq -\log_2 \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} P_X(x) W_{Y|X}^{\frac{1}{1+\rho}}(y|x) \right)^{1+\rho}, \qquad \rho \ge 0. \quad (2.17)$$

Thus, the random-coding exponent can also be written by

$$E_r(R, W_{Y|X}) = \max_{0\le\rho\le1} [E_o(\rho, W_{Y|X}) - \rho R], \quad (2.18)$$

where

$$E_o(\rho, W_{Y|X}) \triangleq \max_{P_X\in\mathcal{P}(\mathcal{X})} E_o(\rho, P_X, W_{Y|X}) \quad (2.19)$$

is called Gallager's channel function.

2.3.3 Expurgated Exponent

For a given DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, define a (not necessarily finite-valued) distortion measure on $\mathcal{X} \times \mathcal{X}$ by

$$d_{W_{Y|X}}(x, \widetilde{x}) \triangleq -\log_2 \sum_{y\in\mathcal{Y}} \sqrt{W_{Y|X}(y|x) W_{Y|X}(y|\widetilde{x})},$$

which is called the Bhattacharya distance between two channel input symbols x and $\widetilde{x}$ in $\mathcal{X}$. For any $R > 0$, define the expurgated exponent for $W_{Y|X}$ by

$$E_{ex}(R, W_{Y|X}) \triangleq \max_{P_X\in\mathcal{P}(\mathcal{X})} E_{ex}(R, P_X, W_{Y|X}) \quad (2.20)$$

where

$$E_{ex}(R, P_X, W_{Y|X}) \triangleq \min_{\substack{Q_{X\tilde{X}}: Q_X = Q_{\tilde{X}} = P_X \\ I(X;\tilde{X}) \le R}} \left[\mathbb{E} d_{W_{Y|X}}(X, \tilde{X}) + I_{Q_{X\tilde{X}}}(X; \tilde{X}) - R\right]. \tag{2.21}$$

It has been shown that $E_{ex}(R, P_X, W_{Y|X})$ admits the following parametric form (e.g., [32])

$$E_{ex}(R, P_X, W_{Y|X}) = \sup_{\rho \ge 1} [E_x(\rho, P_X, W_{Y|X}) - \rho R] \tag{2.22}$$

where

$$E_x(\rho, P_X, W_{Y|X}) \triangleq -\rho \log_2 \sum_{x \in \mathcal{X}} \sum_{\tilde{x} \in \mathcal{X}} P_X(x) P_X(\tilde{x}) \left(\sum_{y \in \mathcal{Y}} \sqrt{W_{Y|X}(y|x) W_{Y|X}(y|\tilde{x})} \right)^{1/\rho}, \tag{2.23}$$

$\rho \ge 1$. Thus, the expurgated exponent can also be written by

$$E_{ex}(R, W_{Y|X}) = \sup_{\rho \ge 1} [E_x(\rho, W_{Y|X}) - \rho R] \tag{2.24}$$

where

$$E_x(\rho, W_{Y|X}) \triangleq \max_{P_X \in \mathcal{P}(\mathcal{X})} E_x(\rho, P_X, W_{Y|X}). \tag{2.25}$$

2.3.4 Sphere-Packing Exponent

For a given DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and any $R > 0$, define the sphere-packing exponent for $W_{Y|X}$ by

$$E_{sp}(R, W_{Y|X}) \triangleq \max_{P_X \in \mathcal{P}(\mathcal{X})} E_{sp}(R, P_X, W_{Y|X}) \tag{2.26}$$

where

$$E_{sp}(R, P_X, W_{Y|X}) \triangleq \min_{V_{Y|X}: I_{P_X V_{Y|X}}(X;Y) \le R} D(V_{Y|X} \,\|\, W_{Y|X} | P_X). \tag{2.27}$$

$E_{sp}(R, P_X, W_{Y|X})$ has a similar parametric form as $E_r(R, P_X, W_{Y|X})$ (e.g., [32])

$$E_{sp}(R, P_X, W_{Y|X}) = \max_{\rho \ge 0} [E_o(\rho, P_X, W_{Y|X}) - \rho R], \tag{2.28}$$

and hence $E_{sp}(R, W_{Y|X})$ can be rewritten by

$$E_{sp}(R, W_{Y|X}) = \max_{\rho \ge 0} [E_o(\rho, W_{Y|X}) - \rho R], \tag{2.29}$$

where $E_o(\rho, P_X, W_{Y|X})$ is defined by (2.17) and $E_o(\rho, W_{Y|X})$ is Gallager's channel function given by (2.19).

2.3.5 Straight-Line Bound and Relations Between Exponents

First of all, $E_r(R, W_{Y|X})$ and $E_{ex}(R, W_{Y|X})$ provide lower bounds for the channel error exponent, which are known as the random-coding lower bound and the expurgated lower bound, respectively, and $E_{sp}(R, W_{Y|X})$, known as the sphere-packing upper bound, is an upper bound to the channel error exponent, i.e.,

$$\max\{E_r(R, W_{Y|X}), E_{ex}(R, W_{Y|X})\} \leq E(R, W_{Y|X}) \leq E_{sp}(R, W_{Y|X}).$$

We point out that the sphere-packing bound $E_{sp}(R, W_{Y|X})$ is loose for small rates (as $R \downarrow 0$). In fact, it is shown in [85] that when R approaches 0, the expurgated bound becomes tight, i.e.,

$$\lim_{R \downarrow 0} E(R, W_{Y|X}) = E_{ex}(0, W_{Y|X}).$$

Of course this bound is nontrivial only if $E_{ex}(0, W_{Y|X}) < \infty$. Furthermore, the straight line connecting the point $(0, E_{ex}(0, W_{Y|X}))$ provided that $E_{ex}(0, W_{Y|X}) < \infty$ and any point on the sphere-packing exponent $(R, E_{sp}(R, W_{Y|X}))$ is an upper bound to $E(R, W_{Y|X})$. Thus, let l be a straight-line passing $(0, E_{ex}(0, W_{Y|X}))$ which is tangent on $E_{sp}(R, W_{Y|X})$ at R_l, then the exponent

$$E_{st}(R, W_{Y|X}) = \begin{cases} l & \text{if } 0 \leq R \leq R_l, \\ E_{sp}(R, W_{Y|X}) & \text{if } R \geq R_l \end{cases} \tag{2.30}$$

gives the smallest upper bound. By definition $E_{st}(R, W_{Y|X})$ is also a decreasing convex function of R and we call $E_{st}(R, W_{Y|X})$ the straight-line upper bound to $E(R, W_{Y|X})$.

The functions $E_r(R, P_X, W_{Y|X})$ and $E_{sp}(R, P_X, W_{Y|X})$ are equal if the maximizing $\rho \leq 1$ in (2.28) or equivalently, if $R \geq R_{cr}(P_X, W_{Y|X})$, where $R_{cr}(P_X, W_{Y|X})$ is the critical rate of the channel $W_{Y|X}$ under distribution P_X, defined by

$$R_{cr}(P_X, W_{Y|X}) \triangleq \left. \frac{\partial E_o(\rho, P_X, W_{Y|X})}{\partial \rho} \right|_{\rho=1}. \tag{2.31}$$

For all P_X, $E_r(R, P_X, W_{Y|X})$ and $E_{sp}(R, P_X, W_{Y|X})$ vanish for all $R \geq I_{P_X W_{Y|X}}(X; Y)$. Consequently, their maxima over P_X, $E_r(R, W_{Y|X})$ and $E_{sp}(R, W_{Y|X})$, vanish for $R \geq$

$C(W_{Y|X})$ and are equal on some interval $[R_{cr}(W_{Y|X}), C(W_{Y|X})]$ where $R_{cr}(W_{Y|X})$ is the critical rate of the channel and is defined by

$$R_{cr}(W_{Y|X}) \triangleq \inf\{R : E_r(R, W_{Y|X}) = E_{sp}(R, W_{Y|X})\}. \tag{2.32}$$

Furthermore, it is known that $E_{sp}(R, W_{Y|X})$ meets $E_r(R, W_{Y|X})$ on its supporting line of slope -1 [32, p. 171], which means that $E_r(R, W_{Y|X})$ is a straight line with slope -1 for $R \leq R_{cr}(W_{Y|X})$ and hence

$$E_r(R, W_{Y|X}) = E_o(1, W_{Y|X}) - R, \qquad R \leq R_{cr}(W_{Y|X}). \tag{2.33}$$

For all P_X, the function $E_{ex}(R, P_X, W_{Y|X})$ is a decreasing convex curve with a straight-line section of slope -1 for $R \geq R_{ex}(P_X, W_{Y|X})$, and $E_{ex}(R, P_X, W_{Y|X}) > E_r(R, P_X, W_{Y|X})$ for $R < R_{ex}(P_X, W_{Y|X})$, where $R_{ex}(P_X, W_{Y|X})$ is the "expurgated" rate of the channel $W_{Y|X}$ under distribution P_X, defined by

$$R_{ex}(P_X, W_{Y|X}) \triangleq \left. \frac{\partial E_x(\rho, P_X, W_{Y|X})}{\partial \rho} \right|_{\rho=1}. \tag{2.34}$$

Since the above are satisfied for all P_X, we then obtain the following relation between the two lower bounds: $E_r(R, W_{Y|X}) < E_{ex}(R, W_{Y|X})$ for $R < R_{ex}(W_{Y|X})$ and $E_r(R, W_{Y|X}) \geq E_{ex}(R, W_{Y|X})$ otherwise, where

$$R_{ex}(W_{Y|X}) \triangleq \inf\{R : E_r(R, W_{Y|X}) = E_{ex}(R, W_{Y|X})\} \tag{2.35}$$

is the expurgated rate of the channel.

Furthermore, it is known that $E_{ex}(R, W_{Y|X})$ and $E_r(R, W_{Y|X})$ meet their supporting line of slope -1 (according to the fact that $E_o(1, W_{Y|X}) = E_x(1, W_{Y|X})$) [42, p. 154]. This geometric relation implies that $R_{ex}(W_{Y|X}) \leq R_{cr}(W_{Y|X})$ and $E_r(R, W_{Y|X}) = E_{ex}(R, W_{Y|X})$ is a straight line in the region $[R_{ex}(W_{Y|X}), R_{cr}(W_{Y|X})]$.

2.3.6 Error Exponent for Continuous Channels with Cost Constraints

We assume in this section that the memoryless channel $W_{Y|X}$ has continuous alphabets $\mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}$, and $W_{Y|X}$ is a valid conditional pdf. Given an input cost function $g : \mathcal{X} \to$

$[0, \infty)$ such that $g(x) = 0$ if and only if $x = 0$, and a constraint $\mathcal{E} > 0$, the channel capacity of the continuous memoryless channel $W_{Y|X}$ is given by

$$C(W_{Y|X}, \mathcal{E}) = \sup_{P_X : \mathbb{E}g(X) \leq \mathcal{E}} I_{P_X W_{Y|X}}(X; Y). \tag{2.36}$$

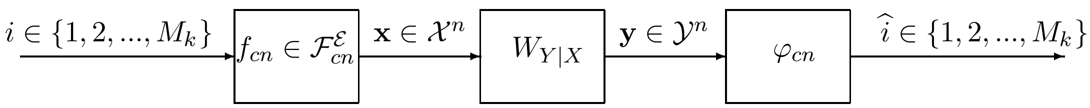

Figure 2.3: Memoryless channel coding system with cost constraint.

An (n, M_n) block channel code for a continuous memoryless channel $W_{Y|X}$ with an input cost constraint $\mathcal{E}$ is a pair of mappings (see Fig. 2.3):

$$f_{cn} : \{1, 2, ..., M_n\} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_{cn} : \mathcal{Y}^n \longrightarrow \{1, 2, ..., M_n\},$$

where f_{cn} is subject to an (arithmetic average) cost constraint:

$$f_{cn} \in \mathcal{F}_{cn}^{\mathcal{E}} \triangleq \left\{ f_{cn} : \frac{1}{n} \sum_{j=1}^{n} g(x_j) \leq \mathcal{E} \quad \text{for all} \quad \mathbf{x} = f_{cn}(i), \quad i \in \{1, 2, ..., M_n\} \right\}.$$

The code rate (measured in nats) is defined as

$$R_n \triangleq \frac{1}{n} \ln M_n \quad \text{nats/channel use.}$$

The (average) probability and the maximum probability of decoding error for the $(f_{cn}, \varphi_{cn}, \mathcal{E})$ code are respectively given by

$$P_{ec}^{(n)}(W_{Y|X}, R_n, \mathcal{E}) \triangleq \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j \neq i}^{M_n} \int_{\mathbf{y} : \varphi_{cn}(\mathbf{y}) = j} W_{Y|X}^{(n)}(\mathbf{y} | f_{cn}(i)) d\mathbf{y}. \tag{2.37}$$

and

$$P_{max,ec}^{(n)}(W_{Y|X}, R_n, \mathcal{E}) \triangleq \max_{1 \leq i \leq M_n} \sum_{j=1, j \neq i}^{M_n} \int_{\mathbf{y} : \varphi_{cn}(\mathbf{y}) = j} W_{Y|X}^{(n)}(\mathbf{y} | f_{cn}(i)) d\mathbf{y}. \tag{2.38}$$

Definition 2.4 For any $R > 0$, the channel error exponent $E(R, W_{Y|X}, \mathcal{E})$ of the channel $W_{Y|X}$ is defined as the supremum of the set of all numbers E for which there exists a sequence of (n, M_n) block codes $(f_{cn}, \varphi_{cn}, \mathcal{E})$ with

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{ec}^{(n)}(W_{Y|X}, R_n, \mathcal{E})$$

and

$$R \leq \liminf_{n\to\infty} R_n.$$

As in the channel coding case without cost constraint, the probability of error $P_{ec}^{(n)}$ and the maximal probability of error $P_{max,ec}^{(n)}$ yield the same channel error exponent (cf. [85, p. 416]). Thus, an equivalent definition for the channel error exponent follows if $P_{ec}^{(n)}(W_{Y|X}, R_n, \mathcal{E})$ is replaced by $P_{max,ec}^{(n)}(W_{Y|X}, R_n, \mathcal{E})$ in the above.

For the general continuous memoryless channel with an input cost constraint, only a lower bound for $E(R, W_{Y|X}, \mathcal{E})$ due to Gallager ([41], [42, Section 7.3]) is known, which we refer to as Gallager's random-coding lower bound for the channel error exponent $E(R, W_{Y|X}, \mathcal{E})$,

$$E(R, W_{Y|X}, \mathcal{E}) \geq E_r(R, W_{Y|X}, \mathcal{E}) \triangleq \max_{0\leq\rho\leq 1} [E_o(\rho, W_{Y|X}, \mathcal{E}) - \rho R], \tag{2.39}$$

where

$$E_0(W, \mathcal{E}, \rho) \triangleq \sup_{P_X : \mathbb{E}g(X)\leq\mathcal{E}, \mathbb{E}g(X)^3<\infty} \max_{r\geq 0} E_0(\rho, r, W, P_X) \tag{2.40}$$

is Gallager's constrained channel function with

$$E_o(\rho, r, P_X, W_{Y|X}, \mathcal{E}) \triangleq -\ln \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} P_X(x) e^{r(g(x)-\mathcal{E})} W_{Y|X}(y|x)^{\frac{1}{1+\rho}} dx \right]^{1+\rho} dy, \tag{2.41}$$

where the supremum in (2.40) is taken over all pdfs $P_X(x)$ defined on $\mathcal{X}$ subject to $\mathbb{E}g(X) \leq \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$. The constraints are satisfied, for example, when P_X is Gaussian distribution with mean 0 variance $\mathcal{E}$ and $g(x) = x^2$. The integrals should be replaced with summations if $W_{Y|X}$ is a memoryless channel with discrete alphabets under a cost constraint. Note that in general we do not have an explicit formula for this bound, because it is not known whether the supremum in (2.40) is achievable or not, and under what distribution it is achieved. In particular, when $W_{Y|X}$ is the memoryless Gaussian channel (MGC) with a power cost constraint ($g(x) = x^2$), we have an upper bound for $E(R, W_{Y|X}, \mathcal{E})$ due to Shannon [84]. We will discuss that bound in detail in the next section.

2.4 Reliability Functions for MGSs and MGCs

For the sake of convenience, all the logarithms and exponentials used in the section (for Gaussian systems) are in natural base. Consider an MGS with alphabet $\mathcal{S} = \mathbb{R}$, mean zero, variance σ_S^2, and pdf

$$Q_S(s) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left\{-\frac{s^2}{2\sigma_S^2}\right\}, \qquad s \in \mathcal{S},$$

denoted by $Q_S \sim \mathcal{N}(0, \sigma_S^2)$, and an MGC $W_{Y|X}$ with common input, output, and additive noise alphabets $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ and described by $Y_i = X_i + Z_i$, where Y_i, X_i and Z_i are the channel's output, input and noise symbols at time i. We assume that X_i and Z_i are independent of each other. The noise admits a zero-mean σ_Z^2-variance Gaussian pdf, denoted by $P_Z \sim \mathcal{N}(0, \sigma_Z^2)$ and thus the transition pdf of the channel is given by

$$W_{Y|X}(y|x) = P_Z(z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left\{-\frac{z^2}{2\sigma_Z^2}\right\}, \qquad z = y - x \in \mathcal{Z}.$$

We assume the squared-error distortion measure for the source given by $d(s, s') = (s - s')^2$ for any $s, s' \in \mathbb{R}$ and extended for k-tuples as

$$d^{(k)}(\mathbf{s}, \mathbf{s}') = \frac{1}{k}\sum_{i=1}^{k}(s_i - s_i')^2$$

for any $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^k$.

Let $\mathcal{S} = \mathcal{S}' \subseteq \mathbb{R}$ and let $P_{SS'}$ be a pdf on continuous alphabet $\mathcal{S} \times \mathcal{S}'$. Then the mutual information between RV's S and S' is given by (e.g., [29])

$$I_{P_{SS'}}(S; S') = \int_{\mathcal{S}\times\mathcal{S}'} P_{SS'}(s, s') \ln \frac{P_{SS'}(s, s')}{P_S(s)P_{S'}(s')} ds ds',$$

or simply written as $I(S; S')$ if $P_{SS'}$ is clear from the context.

Given a distortion threshold $\Delta > 0$, the rate-distortion function for MGS Q_S is given by (e.g., [42])

$$R(Q_S, \Delta) = \inf_{P_{S'|S}:\mathbb{E}d(S,S')\le\Delta} I(S; S') = \begin{cases} \frac{1}{2}\ln\frac{\sigma_S^2}{\Delta} & 0 < \Delta < \sigma_S^2, \\ 0 & \sigma_S^2 \le \Delta. \end{cases} \tag{2.42}$$

When $0 < \Delta < \sigma_S^2$, the infimum of (2.42) is achieved by a conditional Gaussian pdf [97]

$$P^*_{S'|S}(s'|s) = \frac{1}{\sqrt{2\pi \frac{\Delta(\sigma_S^2-\Delta)}{\sigma_S^2}}} \exp\left\{ -\frac{\left(s' - \frac{\sigma_S^2-\Delta}{\sigma_S^2} s\right)^2}{\frac{2\Delta(\sigma_S^2-\Delta)}{\sigma_S^2}} \right\}, \tag{2.43}$$

and hence the marginal pdf of s' under $P^*_{S'|S}(s'|s)$ is given by

$$P^*_{S'}(s') = \int Q_S(s) P^*_{S'|S}(s'|s) ds = \frac{1}{\sqrt{2\pi(\sigma_S^2 - \Delta)}} \exp\left\{ -\frac{s'^2}{2(\sigma_S^2 - \Delta)} \right\}. \tag{2.44}$$

When $\sigma_S^2 \le \Delta$, trivially, $R(Q_S, \Delta)$ can be achieved by setting $s' = 0$ and the resulting marginal distribution of S' is an identity function.

The excess distortion exponent $e_\Delta(R, Q_S)$ for the MGS under the squared-error distortion measure admits an explicit formula [54]: for the MGS $Q_S \sim \mathcal{N}(0, \sigma_S^2)$ and any $R > 0$, the excess distortion exponent $e_\Delta(R, Q_S)$ is determined exactly by $F(R, Q_S, \Delta)$ defined in (2.11), which admits the following parametric form

$$F(R, Q_S, \Delta) = \begin{cases} \frac{1}{2}\left(\frac{\Delta\beta}{\sigma_S^2} - \ln \frac{\Delta\beta}{\sigma_S^2} - 1 \right) & \text{if} \quad R > R(Q_S, \Delta), \\ 0 & \text{otherwise}, \end{cases} \tag{2.45}$$

where $\beta = e^{2R}$. Note that the function $F(R, Q_S, \Delta)$ might have a jump at $R = 0$ by definition. This is not good since we later need to deal with the source exponent as a convex and continuous function on $[0, \infty)$. Since $e_\Delta(R, Q_S)$ is not meaningful at $R = 0$, we define a new function $F_G(R, Q_S, \Delta)$ by

$$F_G(R, Q_S, \Delta) \triangleq F(R, Q_S, \Delta) \tag{2.46}$$

for $R > 0$, and

$$F_G(0, Q_S, \Delta) \triangleq \lim_{R \downarrow 0} F(R, Q_S, \Delta) = \begin{cases} \frac{1}{2}\left(\frac{\Delta}{\sigma_S^2} - \ln \frac{\Delta}{\sigma_S^2} - 1 \right) & \text{if} \quad R(Q_S, \Delta) = 0, \\ 0 & \text{if} \quad R(Q_S, \Delta) > 0, \end{cases} \tag{2.47}$$

Consequently, $F_G(R, Q_S, \Delta)$ is convex strictly increasing in $R \in [0, \infty)$ and we still have $e_\Delta(R, P_S) = F_G(R, Q_S, \Delta)$ for any $R > 0$. In the sequel, we will refer to $F_G(R, Q_S, \Delta)$ as the MGS exponent.

Given an input cost function $g(x) = x^2$ (power cost constraint) and a constraint $\mathcal{E} > 0$, the channel capacity of MGC $W_{Y|X}$ is given by

$$C(W_{Y|X}, \mathcal{E}) = \sup_{P_X : \mathbb{E}X^2 \le \mathcal{E}} I(X;Y) = \frac{1}{2} \ln (1 + \mathrm{SNR}), \tag{2.48}$$

where $\mathrm{SNR} \triangleq \mathcal{E}/\sigma_Z^2$ is the signal-to-noise ratio. It is known that the supremum in (2.48) is achieved by the Gaussian distribution [29, 42]

$$P_X^*(x) = \frac{1}{\sqrt{2\pi\mathcal{E}}} \exp\left\{ -\frac{x^2}{2\mathcal{E}} \right\}, \tag{2.49}$$

and the corresponding channel output has the pdf

$$P_Y^*(y) = \frac{1}{\sqrt{2\pi(\mathcal{E} + \sigma_Z^2)}} \exp\left\{ -\frac{y^2}{2(\mathcal{E} + \sigma_Z^2)} \right\}. \tag{2.50}$$

As mentioned before, the error exponent for the MGC $E(R, W_{Y|X}, \mathcal{E})$ is only partially known. In the last fifty years, the error exponent for the MGC was actively studied and several lower and upper bounds were established (see, e.g., [12, 42, 84]). The most familiar upper bound is obtained by Shannon [84], called the sphere-packing upper bound for the MGC and given by

$$\begin{aligned} E_{sp}(R, W_{Y|X}, \mathcal{E}) \triangleq\ & \frac{\mathrm{SNR}}{4\beta} \left[(\beta + 1) - (\beta - 1)\sqrt{1 + \frac{4\beta}{\mathrm{SNR}(\beta - 1)}} \right] \\ & + \frac{1}{2} \ln \left\{ \beta - \frac{\mathrm{SNR}(\beta - 1)}{2} \left[\sqrt{1 + \frac{4\beta}{\mathrm{SNR}(\beta - 1)}} - 1 \right] \right\}, \end{aligned} \tag{2.51}$$

where $\beta = e^{2R}$, $R \le C(W_{Y|X}, \mathcal{E})$. Note that $E_{sp}(R, W_{Y|X}, \mathcal{E})$ has the following important properties.[1]

Lemma 2.1 *$E_{sp}(R, W_{Y|X}, \mathcal{E})$ is convex strictly decreasing in $R \le C(W_{Y|X}, \mathcal{E})$ and vanishes for $R \ge C(W_{Y|X}, \mathcal{E})$. Furthermore, $E_{sp}(R, W_{Y|X}, \mathcal{E}) \to \infty$ as $R \downarrow 0$.*

[1] The properties of $E_{sp}(R, W_{Y|X}, \mathcal{E})$ were described in Gallager [42, Chapter 7] without a proof.

Proof: Since $E_{sp}(R, W_{Y|X}, \mathcal{E})$ is a differentiable function for $R > 0$, we have

$$\begin{aligned}
&\frac{\partial E_{sp}(R, W_{Y|X}, \mathcal{E})}{\partial R} \\
= \; & \frac{\beta\left[-\text{SNR}\beta^2 - 4\text{SNR}\beta + \text{SNR}^2 + (\text{SNR}+2)\Psi\right]}{\Psi\left[2\beta + \text{SNR}\beta - \text{SNR} - \Psi\right]} \\
= \; & \frac{\left[-\text{SNR}^2\beta - 4\text{SNR}\beta + \text{SNR}^2 + \Psi(\text{SNR}+2)\right](2\beta + \text{SNR}\beta - \text{SNR} + \Psi)}{4\beta\Psi} \\
= \; & \frac{2\text{SNR}^2 - 2\text{SNR}^2\beta - 8\text{SNR}\beta + (4\beta - 2\text{SNR})\Psi}{4\beta\Psi} \\
= \; & 1 - \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}\right), \qquad (2.52)
\end{aligned}$$

where $\beta = e^{2R}$ and

$$\Psi = \sqrt{(\text{SNR}\beta - \text{SNR} + 4\beta)\text{SNR}(\beta - 1)}.$$

Now solving

$$1 - \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}\right) \leq 0$$

yields

$$R \leq \frac{1}{2}\ln(1 + \text{SNR}) = C(W_{Y|X}, \mathcal{E}).$$

Particularly, we have

$$\lim_{R \to C(W_{Y|X}, \mathcal{E})} \frac{\partial E_{sp}(R, W_{Y|X}, \mathcal{E})}{\partial R} = 0 \quad \text{and} \quad \lim_{R \downarrow 0} \frac{\partial E_{sp}(R, W_{Y|X}, \mathcal{E})}{\partial R} = -\infty.$$

Hence, $E_{sp}(R, W_{Y|X}, \mathcal{E})$ is a strictly decreasing function in $R \in (0, C(W_{Y|X}, \mathcal{E})]$ with a slope ranging from $-\infty$ to 0.

It follows from (2.52) that for $R \in (0, C(W_{Y|X}, \mathcal{E})]$,

$$\begin{aligned}
\frac{\partial^2 E_{sp}(R, W_{Y|X}, \mathcal{E})}{\partial R^2} \; = \; & \frac{\text{SNR}}{\beta}\left[1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}\right] \\
& + \frac{2}{\text{SNR}^2(\beta-1)^2\sqrt{1 + \frac{4\beta}{\text{SNR}(\beta-1)}}} > 0. \qquad (2.53)
\end{aligned}$$

This demonstrates the (strict) convexity of $E_{sp}(R, W_{Y|X}, \mathcal{E})$. ■

For the lower bound, we specialize Gallager's random-coding lower bound (2.39) (note that the logarithm and exponential here are in natural base) for the MGC $W_{Y|X}$ as follows:

choosing the channel input distribution $P_X(x)$ as the Gaussian distribution $P_X^*(x)$ given in (2.49), and replacing $g(x)$ by our square cost function x^2 yield the following lower bound for $E_0(W_{Y|X}, \mathcal{E}, \rho)$

$$E_0(W_{Y|X}, \mathcal{E}, \rho) \geq \widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho) \triangleq \max_{r \geq 0} E_0(W_{Y|X}, \mathcal{E}, \rho, r, P_X^*)$$

$$= \max_{0 \leq r \leq 1/2\mathcal{E}} \left\{ r(1+\rho)\mathcal{E} + \frac{1}{2}\ln(1-2r\mathcal{E}) + \frac{\rho}{2}\ln\left[1 - 2r\mathcal{E} + \frac{\mathcal{E}}{(1+\rho)\sigma_Z^2}\right]\right\}. \quad (2.54)$$

$$(2.55)$$

We hereby call $\widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)$ Gallager's Gaussian-input channel function. Note also that

$$E_{sp}(R, W_{Y|X}, \mathcal{E}) = \max_{\rho \geq 0}[-\rho R + \widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)],$$

and the inner function is concave in ρ. Thus, the random-coding lower bound $E_r(R, W_{Y|X}, \mathcal{E})$ can be further lower bounded by [42, pp. 339–340]

$$E_{\dagger}(R, W_{Y|X}, \mathcal{E}) = \max_{0 \leq \rho \leq 1}[-\rho R + \widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)]$$

$$= \begin{cases} E_{sp}(R, W_{Y|X}, \mathcal{E}), \\ \qquad R_{cr}(W_{Y|X}) \leq R \leq C(W_{Y|X}, \mathcal{E}), \\ 1 - \gamma + \frac{\text{SNR}}{2} + \frac{1}{2}\ln\left(\gamma - \frac{\text{SNR}}{2}\right) + \frac{1}{2}\ln\gamma - R \\ \qquad 0 \leq R \leq R_{cr}(W_{Y|X}), \end{cases} \quad (2.56)$$

where

$$\gamma \triangleq \frac{1}{2}\left[1 + \frac{\text{SNR}}{2} + \sqrt{1 + \frac{\text{SNR}^2}{4}}\right],$$

and

$$R_{cr}(W_{Y|X}) \triangleq \frac{1}{2}\ln\left[\frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2}\sqrt{1 + \frac{\text{SNR}^2}{4}}\right]$$

is the critical rate of the MGC (obtained by solving for the R where the straight-line of slope -1 is tangent to $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$). It is easy to show that $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ is convex strictly decreasing in $0 < R \leq C(W_{Y|X}, \mathcal{E})$ with a straight-line section of slope -1 for $R \leq R_{cr}(W_{Y|X})$. It has to be pointed out [42] that $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ is not the real random-coding bound (as given in (2.39)) for $R < R_{cr}(W_{Y|X})$, but it admits a computable parametric form and it coincides with the upper bound $E_{sp}(R, W_{Y|X}, \mathcal{E})$ for $R \geq R_{cr}(W_{Y|X})$. Thus, we will refer to $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ as the Gaussian input random-coding exponent, and the channel coding error exponent $E(R, W_{Y|X}, \mathcal{E})$ is determined for high rates ($R \geq R_{cr}(W_{Y|X})$).[2]

2.5 Concluding Remarks

In this chapter, we presented the basic material on the source and channel coding reliability functions for memoryless systems that will be widely used in the thesis. In the fifty-year development of Shannon theory, the source reliability function has been exactly determined for DMS's (in both lossless and lossy coding cases), MGS's with the squared-error distortion, and certain memoryless sources with a metric distortion measure under some finiteness constraints, while the channel error exponent for DMC's and MGC's is only known for high rates. In Chapter 7 we derive a sphere-packing type upper bound for the error exponent for discrete additive channels with Markovian memory, and in Chapter 8, we determine the source excess distortion exponent for memoryless Laplacian sources with magnitude distortion measure.

We note that the the source error exponent for DMS can be expressed in two forms: the constrained divergence form (2.4) and the parametric form in terms of Gallager's source function (2.6). The random-coding exponent and the sphere-packing exponent can also be expressed in a form in terms of Gallager's channel function and alternately in a constrained divergence form. In Chapter 4, we re-examine these exponents, and a conjugacy relation between these exponents and the corresponding source/channel functions will be illustrated.

[2] In the recent work of [14], the lower bound $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ is improved and is shown to be tight in a interval slightly below the critical rate, i.e., it is shown that the error exponent of the MGC is determined by $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ for rates $R \geq R_1$ and R_1 can be less than $R_{cr}(W_{Y|X})$.

Chapter 3

Background and Fundamental Results on the Method of Types

This chapter provides a technical background which will be used to establish upper and lower bounds for the JSCC reliability function (error exponent and excess distortion exponent) in later chapters.

It is well known that the method of types is a very useful tool in information theory, particularly in Shannon theory, hypothesis testing and large deviation theory (e.g., [32,33]). For a DMS $Q_S \in \mathcal{P}(\mathcal{S})$, the type-$P$ class of k-length sequences $\mathbf{s} \triangleq (s_1 s_2 \cdots s_k) \in \mathcal{S}^k$ is the set of sequences that have single-symbol empirical distribution equal to P. Thus, by partitioning all sequences in $\mathcal{S}^k$ into type classes where the number of distinct classes grows sub-exponentially with k, the probability of a particular event (the probability of error, say) can be obtained by summing the probabilities of its intersections with the various type classes which decay exponentially as the sequence length k approaches infinity [32].

In Section 3.1, we go through the concept of types, joint types, and conditional types as well as the corresponding type classes for (memoryless) sequences with finite alphabets. For simplicity, we refer to these type classes of discrete sequences as discrete type classes. In Section 3.2, we develop a joint type packing lemma for a (2-dimensional) joint type setting which will be used (in Chapter 9) to establish a lower bound for the JSCC reliability function. The counterparts of the discrete type class in $\mathbb{R}^k$ are next investigated in Section 3.3. We first introduce the notion of a Gaussian-type class proposed in [6]. We also propose a Laplacian-type class. The properties of the two continuous type classes are discussed. In Section 3.4, we introduce another fundamental lemma on the method of types, the type covering lemma, for the discrete type class, Gaussian-type class and Laplacian-type class, respectively. We finally conclude in Section 3.5.

3.1 Types, Joint Types, and Conditional Types

The following notation and conventions are adopted from [30,32]. The type of an n-length sequence $\mathbf{x} \in \mathcal{X}^n$ is the empirical probability distribution $P_{\mathbf{x}} \in \mathcal{P}(\mathcal{X})$ defined by

$$P_{\mathbf{x}}(a) \triangleq \frac{1}{n} N(a|\mathbf{x}), \quad a \in \mathcal{X},$$

where $N(a|\mathbf{x})$ is the number of occurrences of a in $\mathbf{x}$. Let $\mathcal{P}_n(\mathcal{X}) \subseteq \mathcal{P}(\mathcal{X})$ be the collection of all types of sequences in $\mathcal{X}^n$. For any $P_X \in \mathcal{P}_n(\mathcal{X})$, the set of all $\mathbf{x} \in \mathcal{X}^n$ with type P_X is denoted by $\mathbb{T}_{P_X}$, or simply by $\mathbb{T}_X$ if P_X is understood. We also call $\mathbb{T}_{P_X}$ or $\mathbb{T}_X$ a type class.

Similarly, the joint type of n-length sequences $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$ is the empirical joint probability distribution $P_{\mathbf{xy}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ defined by

$$P_{\mathbf{xy}}(a, b) \triangleq \frac{1}{n} N(a, b|\mathbf{x}, \mathbf{y}), \quad (a, b) \in \mathcal{X} \times \mathcal{Y}.$$

Let $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \subseteq \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be the collection of all joint types of sequences in $\mathcal{X}^n \times \mathcal{Y}^n$. The set of all $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$ with joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ is denoted by $\mathbb{T}_{P_{XY}}$, or simply by $\mathbb{T}_{XY}$.

The conditional type of $\mathbf{y} \in \mathcal{Y}^n$ given $\mathbf{x} \in \mathbb{T}_{P_X}$ is the empirical conditional probability distribution $P_{\mathbf{y}|\mathbf{x}} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ defined by

$$P_{\mathbf{y}|\mathbf{x}}(b|a) = \frac{N(a, b|\mathbf{x}, \mathbf{y})}{N(a|\mathbf{x})},$$

whenever $N(a|\mathbf{x}) > 0$; otherwise (if $N(a|\mathbf{x}) = 0$) define $P_{\mathbf{y}|\mathbf{x}}(b|a) = 0$, $(a, b) \in \mathcal{X} \times \mathcal{Y}$.

Let $\mathcal{P}_n(\mathcal{Y}|P_X)$ be the collection of all conditional distributions $V_{Y|X}$ which are conditional types of $\mathbf{y} \in \mathcal{Y}^n$ given an $\mathbf{x} \in \mathbb{T}_{P_X}$. For any conditional type $V_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|P_X)$, the set of all $\mathbf{y} \in \mathcal{Y}^n$ for a given $\mathbf{x} \in \mathbb{T}_{P_X}$ satisfying $P_{\mathbf{y}|\mathbf{x}} = V_{Y|X}$ is denoted by $\mathbb{T}_{V_{Y|X}}(\mathbf{x})$, or simply by $\mathbb{T}_{Y|X}(\mathbf{x})$, which is also called a conditional type class (V-shell) with respect to $\mathbf{x}$.

Note that for a given joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, for any $\mathbf{x} \in \mathbb{T}_{P_X}$, $\mathbb{T}_{P_{Y|X}}(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \mathbb{T}_{P_{XY}}\}$. Note also that

$$\left\{P_X V_{Y|X} : P_X \in \mathcal{P}_n(\mathcal{X}), V_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|P_X)\right\} = \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}).$$

In addition, we denote

$$\mathcal{P}_n(\mathcal{Y}|\mathcal{X}) \triangleq \bigcup_{P_X \in \mathcal{P}_n(\mathcal{X})} \mathcal{P}_n(\mathcal{Y}|P_X) \subseteq \mathcal{P}(\mathcal{Y}|\mathcal{X}).$$

To distinguish between different distributions (or types) defined on the same alphabet, we use sub-subscripts, say, i, j, in P_{X_i}, $P_{X_iY_j}$, $\mathbb{T}_{X_iY_j}$, and so on. For example, $\mathbb{T}_{X_iY_j}$ is the type class of the joint type $P_{X_iY_j} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$. The following facts will be frequently used throughout the thesis.

Lemma 3.1 [32]

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

(b) For any $P_X, Q_X \in \mathcal{P}_n(\mathcal{X})$, *we have*

$$(n+1)^{-|\mathcal{X}|} 2^{nH_{P_X}(X)} \leq |\mathbb{T}_{P_X}| \leq 2^{nH_{P_X}(X)},$$

$$Q_X^{(n)}(\boldsymbol{x}) = 2^{-n[D(P_X\|Q_X) + H_{P_X}(X)]} \quad \text{if } \boldsymbol{x} \in \mathbb{T}_{P_X},$$

and hence

$$(n+1)^{-|\mathcal{X}|} 2^{-nD(P_X\|Q_X)} \leq Q_X^{(n)}(\mathbb{T}_{P_X}) \leq 2^{-nD(P_X\|Q_X)}.$$

(c) For any $\boldsymbol{x} \in \mathbb{T}_{P_X}$, $\boldsymbol{y} \in \mathbb{T}_{V_{Y|X}}(\boldsymbol{x})$ *and* $W_{Y|X}, V_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|P_X)$, *we have*

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} 2^{nH_{P_X V_{Y|X}}(Y|X)} \leq |\mathbb{T}_{V_{Y|X}}(\boldsymbol{x})| \leq 2^{nH_{P_X V_{Y|X}}(Y|X)},$$

$$W_{Y|X}^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = 2^{-n[D(V_{Y|X}\|W_{Y|X}|P_X) + H_{P_X V_{Y|X}}(Y|X)]},$$

and hence

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|}2^{-nD(V_{Y|X}\|W_{Y|X}|P_X)} \le W^{(n)}_{Y|X}(\mathbb{T}_{V_{Y|X}}(\boldsymbol{x})|\boldsymbol{x}) \le 2^{-nD(V_{Y|X}\|W_{Y|X}|P_X)}.$$

Proof: (a) is trivial. For (b), we only need to prove the bounds for $|\mathbb{T}_{P_X}|$. For $P_X \in \mathcal{P}_n(\mathcal{X})$, and an n-length sequence $\mathbf{x} \in \mathbb{T}_{P_X}$, by the definition of type,

$$P^{(n)}_X(\mathbf{x}) = \prod_{a\in\mathcal{X}} P_X(a)^{N(a|\mathbf{x})} = 2^{-nH_{P_X}(X)}.$$

Since each sequence in $\mathbb{T}_{P_X}$ are equiprobable, the size of the type class is

$$|\mathbb{T}_{P_X}| = P^{(n)}_X(\mathbb{T}_{P_X})2^{nH_{P_X}(X)}.$$

The upper bound for $|\mathbb{T}_{P_X}|$ is obvious since $P^{(n)}_X(\mathbb{T}_{P_X}) \le 1$. If we can show

$$P^{(n)}_X(\mathbb{T}_{P_X}) \ge P^{(n)}_X(\mathbb{T}_{\widehat{P}_X}) \tag{3.1}$$

for any $\widehat{P}_X \in \mathcal{P}_n(\mathcal{X})$ which implies that $\mathbb{T}_{P_X}$ has the largest probability among the $|\mathcal{P}_n(\mathcal{X})|$ types, then by (a),

$$P^{(n)}_X(\mathbb{T}_{P_X}) \ge \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \ge (n+1)^{-|\mathcal{X}|}$$

and thus the lower bound for $|\mathbb{T}_{P_X}|$ is proved. Now we show (3.1). For any $\widehat{P}_X \in \mathcal{P}_n(\mathcal{X})$, the probability of the type class $\mathbb{T}_{\widehat{P}_X}$ under the distribution P_X is given by

$$P^{(n)}_X(\mathbb{T}_{\widehat{P}_X}) = |\mathbb{T}_{\widehat{P}_X}| \prod_{a\in\mathcal{X}} P_X(a)^{n\widehat{P}_X(a)} = \frac{n!}{\prod_{a\in\mathcal{X}}(n\widehat{P}_X(a))!} \prod_{a\in\mathcal{X}} P_X(a)^{n\widehat{P}_X(a)}.$$

It follows that

$$\frac{P^{(n)}_X(\mathbb{T}_{\widehat{P}_X})}{P^{(n)}_X(\mathbb{T}_{P_X})} = \prod_{a\in\mathcal{X}} \frac{\prod_{a\in\mathcal{X}}(nP_X(a))!}{\prod_{a\in\mathcal{X}}(n\widehat{P}_X(a))!} P_X(a)^{n(\widehat{P}_X(a)-P_X(a))}.$$

Applying the inequality $n!/m! \le n^{n-m}$ yields

$$\frac{P^{(n)}_X(\mathbb{T}_{\widehat{P}_X})}{P^{(n)}_X(\mathbb{T}_{P_X})} \le \prod_{a\in\mathcal{X}} n^{n(P_X(a)-\widehat{P}_X(a))} = n^{n(\sum_{a\in\mathcal{X}} P_X(a) - \sum_{a\in\mathcal{X}} \widehat{P}_X(a))} = 1.$$

(c) is proved similarly as (b). ∎

3.2 A Joint Type Packing Lemma

We generalize Csiszár's type packing lemma [30, Theorem 5] from a (1-dimensional) single-letter type setting to a (2-dimensional) joint type setting. This packing lemma will play a key role in deriving an exponentially achievable upper bound for the probability of erroneous transmission for the DMC in Proposition 6.1 and the asymmetric 2-user channel in Proposition 9.1.

Lemma 3.2 (Joint Type Packing Lemma) *Given finite sets $\mathcal{A}$ and $\mathcal{B}$, a sequence of positive integers $\{m_n\}$, and a sequence of positive integers $\{m'_{in}\}$ associated with every $i = 1, 2, ..., m_n$, for arbitrary (not necessarily distinct) types $P_{A_i} \in \mathcal{P}_n(\mathcal{A})$ and conditional types $P_{B_j|A_i} \in \mathcal{P}_n(\mathcal{B}|P_{A_i})$, and positive integers N_i and M_{ij}, $i = 1, 2, ..., m_n$ and $j = j(i) = 1, 2, ..., m'_{in}$ with*

$$\frac{1}{n}\log_2 N_i < H_{P_{A_i}}(A) - \delta, \tag{3.2}$$

and

$$\frac{1}{n}\log_2 M_{ij} < H_{P_{A_i}P_{B_j|A_i}}(B|A) - \delta, \tag{3.3}$$

where

$$\delta \triangleq \frac{2}{n}\left[|\mathcal{A}|^2|\mathcal{B}|^2\log_2(n+1) + \log_2 m_n + \log_2(\max_i m'_{in}) + \log_2 12\right],$$

there exist m_n disjoint subsets

$$\Omega_i = \left\{\boldsymbol{a}_p^{(i)}\right\}_{p=1}^{N_i} \subseteq \mathbb{T}_{A_i} \triangleq \mathbb{T}_{P_{A_i}}$$

such that

$$|\mathbb{T}_{V_{A'|A}}(\boldsymbol{a}_p^{(i)}) \bigcap \Omega_k| \leq N_k 2^{-n\left[I_{P_{A_i}V_{A'|A}}(A;A')-\delta\right]}, \tag{3.4}$$

for every i, k, p and $V_{A'|A} \in \mathcal{P}_n(\mathcal{A}|\mathcal{A})$, with the exception of the case when both $i = k$ and $V_{A'|A}$ is the conditional distribution such that $V_{A'|A}(a'|a)$ is 1 if $a' = a$ and 0 otherwise; furthermore, for every $\boldsymbol{u}_p^{(i)} \in \Omega_i$ and every i, there exist m'_{in} disjoint subsets

$$\Omega_{ij}(\boldsymbol{a}_p^{(i)}) = \left\{(\boldsymbol{a}_p^{(i)}, \boldsymbol{b}_{p,q}^{(j)})\right\}_{q=1}^{M_{ij}}$$

such that $\boldsymbol{b}_{p,q}^{(j)} \in \mathbb{T}_{B_j|A_i}(\boldsymbol{a}_p^{(i)}) \triangleq \mathbb{T}_{P_{B_j|A_i}}(\boldsymbol{a}_p^{(i)})$ *and*

$$\left| \mathbb{T}_{V_{A'B'|AB}}(\boldsymbol{a}_p^{(i)}, \boldsymbol{b}_{p,q}^{(j)}) \bigcap \bigcup_{p'=1}^{N_k} \Omega_{kl}(\boldsymbol{a}_{p'}^{(k)}) \right| \leq N_k M_{kl} 2^{-n\left[I_{P_{A_iB_j}V_{A'B'|AB}}(A,B;A',B')-\delta\right]}, \tag{3.5}$$

$$\left| \mathbb{T}_{V_{A'B'|AB}}(\boldsymbol{a}_p^{(i)}, \boldsymbol{b}_{p,q}^{(j)}) \bigcap \bigcup_{p'=1}^{N_i} \Omega_{il}(\boldsymbol{a}_{p'}^{(i)}) \right| \leq M_{il} 2^{-n\left[I_{P_{A_iB_j}V_{A'B'|AB}}(B;B'|A)-\delta\right]}, \tag{3.6}$$

for any i, j, k, l, p, q *and* $V_{A'B'|AB} \in \mathcal{P}_n(\mathcal{A} \times \mathcal{B}|\mathcal{A} \times \mathcal{B})$*, with the exception of the case when both* $i = k$*,* $j = l$ *and* $V_{A'B'|AB}$ *is the conditional distribution such that* $V_{A'B'|AB}(a', b'|a, b)$ *is 1 if* $(a', b') = (a, b)$ *and 0 otherwise.*

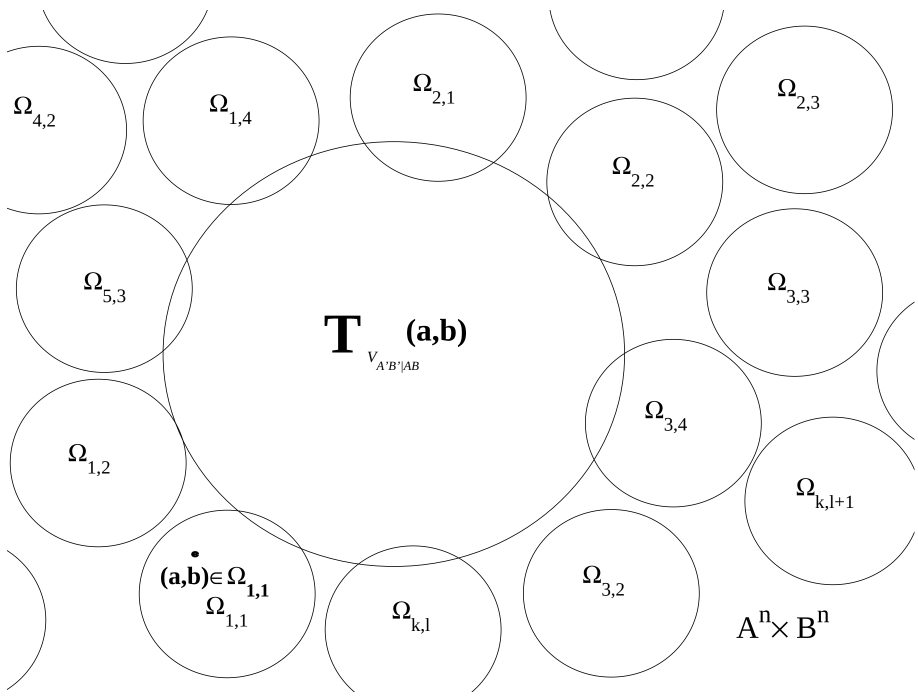

Figure 3.1: A graphical illustration of the (2-dimensional) joint type packing lemma. There exist disjoint subsets Ω_{ij}'s with bounded cardinalities in the "2-dimensional" space $\mathcal{A}^n \times \mathcal{B}^n$ such that for any $(\mathbf{a}, \mathbf{b}) \in \Omega_{ij}$ (say, $(\mathbf{a}, \mathbf{b}) \in \Omega_{1,1}$), the conditional type class $\mathbb{T}_{V_{A'B'|AB}}(\mathbf{a}, \mathbf{b})$ is "almost disjoint" with these subsets $\Omega_{k,l}$'s, provided that $V_{A'B'|AB}$ is not the conditional distribution such that $V_{A'B'|AB}(a', b'|a, b) = 1$ if $(a', b') = (a, b)$ and 0 otherwise.

We remark that the assertion of (3.4) is Csiszár's type packing lemma [30, Theorem 5] for a single-letter type setting. Roughly and intuitively, if $(\mathbf{a}, \mathbf{b})$ is a pair of transmitted codewords, then the possible sequences decoded as $(\mathbf{a}, \mathbf{b})$ can be seen as elements in the "sphere" $\mathbb{T}_{V_{A'B'|AB}}(\mathbf{a}, \mathbf{b})$ "centered" at $(\mathbf{a}, \mathbf{b})$ for some $V_{A'B'|AB}$. The packing lemma states that there exist disjoint sets $\Omega_{kl} = \bigcup_{p'=1}^{N_k} \Omega_{kl}(\mathbf{a}_{p'}^{(k)})$ with bounded cardinalities such that the size of the intersection between the sphere $\mathbb{T}_{V_{A'B'|AB}}(\mathbf{a}, \mathbf{b})$ for every $(\mathbf{a}, \mathbf{b}) \in \Omega_{ij}$ and every set Ω_{kl} is "exponentially small" compared with the size of each Ω_{kl}; see Fig. 3.1. So the packing lemma can be used to prove the existence of good codes that have an exponentially small probability of error.

Note also that the above extended packing lemma is analogous to, but different from the one introduced by Körner and Sgarro [60], which is used to prove a lower bound for the asymmetric broadcast channel coding exponent. Lemma 3.2 here is used for the JSCC problem.

Proof of Lemma 3.2: Although the result (3.4) of Lemma 3.2 was already shown in [30], we include its proof here since we need to show that (3.4) holds simultaneously with (3.5) and (3.6). We employ a random selection argument as used in [30]. For each $i = 1, 2, ..., m_n$, we randomly generate a set of $2N_i$ sequences (according to a uniform distribution) from the type class $\mathbb{T}_{A_i} = \mathbb{T}_{P_{A_i}}$, $\mathcal{C}_i \triangleq \left\{\mathbf{a}_1^{(i)}, \mathbf{a}_2^{(i)}, ..., \mathbf{a}_{2N_i}^{(i)}\right\} \subseteq \mathbb{T}_{A_i}$, i.e., each $\mathbf{a}_p^{(i)}$ is randomly drawn from the type class $\mathbb{T}_{A_i}$ with probability $1/|\mathbb{T}_{A_i}|$, $p = 1, 2, ..., 2N_i$. Each sets has $2N_i$ elements rather than N_i because an expurgate operation will be performed later. Also, we denote the set $\mathcal{C}_i^p \triangleq \mathcal{C}_i / \left\{\mathbf{a}_p^{(i)}\right\}$.

Now for each i with associated $j = j(i) = 1, 2, ..., m'_{in}$, we randomly generate $4N_iM_{ij}$ sequences (according to a uniform distribution)

$$\left\{\mathbf{b}_{11}^{(j)}, \mathbf{b}_{12}^{(j)}, ..., \mathbf{b}_{1,2M_{ij}}^{(j)}, \mathbf{b}_{21}^{(j)}, \mathbf{b}_{22}^{(j)}, ..., \mathbf{b}_{2,2M_{ij}}^{(j)}, \cdots, \mathbf{b}_{2N_i,1}^{(j)}, \mathbf{b}_{2N_i,2}^{(j)}, ..., \mathbf{b}_{2N_i,2M_{ij}}^{(j)}\right\}$$

such that the set

$$\begin{aligned}\mathcal{C}_{ij} \triangleq & \left\{\left(\mathbf{a}_1^{(i)}, \mathbf{b}_{11}^{(j)}\right), \left(\mathbf{a}_1^{(i)}, \mathbf{b}_{12}^{(j)}\right), ..., \left(\mathbf{a}_1^{(i)}, \mathbf{b}_{1,2M_{ij}}^{(j)}\right),\right. \\ & \left(\mathbf{a}_2^{(i)}, \mathbf{b}_{21}^{(j)}\right), \left(\mathbf{a}_2^{(i)}, \mathbf{b}_{22}^{(j)}\right), ..., \left(\mathbf{a}_2^{(i)}, \mathbf{b}_{2,2M_{ij}}^{(j)}\right),\end{aligned}$$

$$\cdots\cdots$$

$$\left(\mathbf{a}_{2N_i}^{(i)}, \mathbf{b}_{2N_i,1}^{(j)}\right), \left(\mathbf{a}_{2N_i}^{(i)}, \mathbf{b}_{2N_i,2}^{(j)}\right), ..., \left(\mathbf{a}_{2N_i}^{(i)}, \mathbf{b}_{2N_i,2M_{ij}}^{(j)}\right)\right\} \subseteq \mathbb{T}_{A_iB_j} = \mathbb{T}_{P_{A_i}P_{B_j|A_i}}.$$

In other words, each $\mathbf{b}_{p,q}^{(j)}$ is drawn from $\mathbb{T}_{B_j|A_i}\left(\mathbf{a}_p^{(i)}\right)$ with probability $1/\left|\mathbb{T}_{B_j|A_i}\left(\mathbf{a}_p^{(i)}\right)\right|$, $q = 1, 2, ..., M_{ij}$, and hence each pair $\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)$ is drawn from $\mathbb{T}_{A_iB_j}$ with probability $1/\left|\mathbb{T}_{A_iB_j}\right|$. Furthermore, we denote the set $\mathcal{C}_{ij}^{pq} \triangleq \mathcal{C}_{ij}/\left\{\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right\}$. For any $1 \le i, k \le m_n$, $1 \le j \le m'_{in}$ and $1 \le l \le m'_{kn}$, define

$$\mathcal{V}_{i,k} \triangleq \left\{V_{A'|A} \in \mathcal{P}_n\left(\mathcal{A}|P_{A_i}\right) : \sum_{a\in\mathcal{A}} P_{A_i}(a)V_{A'|A}(a'|a) = P_{A_k}(a')\right\}$$

and

$$\mathcal{V}_{ij,kl} \triangleq \Big\{V_{A'B'|AB} \in \mathcal{P}_n\left(\mathcal{A}\times\mathcal{B}|P_{A_iB_j}\right) : \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_{A_iB_j}(a,b)V_{A'B'|AB}(a',b'|a,b) = P_{A_kB_l}(a',b')\Big\}.$$

Based on the above set-up, the following inequalities hold.

1. For any $(i,j) \neq (k,l)$ and any $V_{A'B'|AB} \in \mathcal{V}_{ij,kl}$,

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\bigcap \mathcal{C}_{kl}\right| \\
\le\ & \mathbb{E}\left|\left\{(p',q') : \left(\mathbf{a}_{p'}^{(k)}, \mathbf{b}_{p',q'}^{(l)}\right) \in \mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right\}\right| \\
=\ & 4N_kM_{kl}\Pr\left\{\left(\mathbf{a}_1^{(k)}, \mathbf{b}_{1,1}^{(l)}\right) \in \mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right\} \\
=\ & 4N_kM_{kl}\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right|}{\left|\mathbb{T}_{A_kB_l}\right|} \\
\le\ & 4N_kM_{kl}(n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-nI_{P_{A_iB_j}V_{A'B'|AB}}(A',B';A,B)},
\end{aligned} \tag{3.7}$$

where the above expectation and probability are taken over the uniform distribution

$$\widehat{P}_{k,l}\left(\mathbf{a}_{p'}^{(k)}, \mathbf{b}_{p',q'}^{(l)}\right) \triangleq \frac{1}{\left|\mathbb{T}_{A_kB_l}\right|}$$
$$\forall \quad 1 \le k \le m_n, \quad 1 \le l \le m'_{kn}, \quad 1 \le p' \le N_k, \quad 1 \le q' \le M_{kl}, \tag{3.8}$$

and (3.7) follows from the basic facts (Lemma 3.1) that

$$\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right| \le 2^{nH_{P_{A_iB_j}V_{A'B'|AB}}(A',B'|A,B)}$$

and that

$$|\mathbb{T}_{A_k B_l}| \geq (n+1)^{-|\mathcal{A}||\mathcal{B}|} 2^{nH_{P_{A_k B_l}}(A',B')},$$

noting that the marginal distribution of $P_{A_i B_j} V_{A'B'|AB}$ for RV's (A', B') is $P_{A_k B_l}$.

2. For any $(i,j) = (k,l)$ and any $V_{A'B'|AB} \in \mathcal{V}_{ij,ij}$, likewise,

$$\mathbb{E} \left| \mathbb{T}_{V_{A'B'|AB}} \left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)} \right) \bigcap \mathcal{C}_{ij}^{pq} \right| \leq 4 N_i M_{ij} (n+1)^{|\mathcal{A}||\mathcal{B}|} 2^{-nI_{P_{A_i B_j} V_{A'B'|AB}}(A',B';A,B)}, \tag{3.9}$$

where the expectation is taken over the uniform distribution $\widehat{P}_{i,j}$ defined by (3.8).

3. For any i and $j \neq l$, and any $V_{AB'|AB} \in \mathcal{V}_{ij,il}$, similarly we have

$$\mathbb{E} \left| \mathbb{T}_{V_{A'B'|AB}} \left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)} \right) \bigcap \mathcal{C}_{il} \right| \leq 4 N_i M_{il} (n+1)^{|\mathcal{A}||\mathcal{B}|} 2^{-nI_{P_{A_i B_j} V_{AB'|AB}}(A,B';A,B)}.$$

Using the identity

$$I_{P_{A_i B_j} V_{AB'|AB}}(A, B'; A, B) = H_{P_{A_i}}(A) + I_{P_{A_i B_j} V_{AB'|AB}}(B'; B|A)$$

and assumption (3.2)

$$\frac{1}{n} \log_2 N_i < H_{P_{A_i}}(A) - \delta,$$

we obtain another bound

$$\mathbb{E} \left| \mathbb{T}_{V_{A'B'|AB}} \left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)} \right) \bigcap \mathcal{C}_{il} \right| \leq 4 M_{il} (n+1)^{|\mathcal{A}||\mathcal{B}|} 2^{-nI_{P_{A_i B_j} V_{A'B'|AB}}(B';B|A)}, \tag{3.10}$$

where the expectation is taken over the uniform distribution $\widehat{P}_{i,l}$.

4. For any i and $j = l$, and any $V_{A'B'|AB} \in \mathcal{V}_{ij,il}$, likewise,

$$\mathbb{E} \left| \mathbb{T}_{V_{A'B'|AB}} \left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)} \right) \bigcap \mathcal{C}_{ij}^{pq} \right| \leq 4 M_{ij} (n+1)^{|\mathcal{A}||\mathcal{B}|} 2^{-nI_{P_{A_i B_j} V_{A'B'|AB}}(B';B|A)}, \tag{3.11}$$

where the expectation is taken over the uniform distribution $\widehat{P}_{i,j}$.

5. For any $i \neq k$ and any $V_{A'|A} \in \mathcal{V}_{i,k}$,

$$\begin{aligned}
\mathbb{E} \left| \mathbb{T}_{V_{A'|A}} \left(\mathbf{a}_p^{(i)} \right) \bigcap \mathcal{C}_k \right| &\leq \mathbb{E} \left| \left\{ p' : \mathbf{a}_{p'}^{(k)} \in \mathbb{T}_{V_{A'|A}} \left(\mathbf{a}_p^{(i)} \right) \right\} \right| \\
&= 2N_k \Pr \left\{ \mathbf{a}_1^{(i)} \in \mathbb{T}_{V_{A'|A}} \left(\mathbf{a}_p^{(i)} \right) \right\} \\
&= 2N_k \frac{\left| \mathbb{T}_{V_{A'|A}} \left(\mathbf{a}_p^{(i)} \right) \right|}{|\mathbb{T}_{A_k}|} \\
&\leq 2N_k (n+1)^{-|\mathcal{A}|} 2^{-nI_{P_{A_i} V_{A'|A}}(A';A)},
\end{aligned} \tag{3.12}$$

where the above expectation and probability are taken over the uniform distribution

$$\widetilde{P}_k(\mathbf{a}_{p'}^{(k)}) \triangleq \frac{1}{|\mathbb{T}_{A_k}|}, \quad \forall \quad 1 \leq k \leq m_n, \quad 1 \leq p' \leq N_k, \tag{3.13}$$

and (3.12) follows from the basic facts (Lemma 3.1) that

$$\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_1^{(i)}\right)\right| \leq 2^{nH_{P_{A_i}V_{A'|A}}(A'|A)}$$

and that

$$|\mathbb{T}_{A_k}| \geq (n+1)^{|\mathcal{A}|} 2^{nH_{P_{A_k}}(A')},$$

noting that the marginal distribution of $P_{A_i}V_{A'|A}$ for the RV A' is P_{A_k}.

6. For any $i = k$ and any $V_{A'|A} \in \mathcal{V}_{i,k}$, likewise,

$$\mathbb{E}\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right)\bigcap \mathcal{C}_i^p\right| \leq 2N_k(n+1)^{-|\mathcal{A}|}2^{-nI_{P_{A_i}V_{A'|A}}(A';A)}, \tag{3.14}$$

where the expectation is taken over the uniform distribution $\widetilde{P}_i$ defined in (3.13).

Note also if $V_{A'B'|AB} \notin \mathcal{V}_{ij,kl}$

$$\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\bigcap \mathcal{C}_{kl}\right| = 0,$$

and if $V_{A'B'|AB} \notin \mathcal{V}_{ij,ij}$

$$\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\bigcap \mathcal{C}_{ij}^{pq}\right| = 0.$$

Therefore, it follows from (3.7) and (3.9) that for any $V_{A'B'|AB} \in \mathcal{P}_n(\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B})$,

$$\begin{aligned} &\frac{\mathbb{E}\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{x}_{p,q}^{(j)}\right)\cap \mathcal{C}_{ij}^{pq}\right|}{4N_iM_{ij}} + \sum_{(k,l)\neq(i,j)} \frac{\mathbb{E}\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\cap \mathcal{C}_{kl}\right|}{4N_kM_{kl}} \\ \leq\ & m_n(\max_i m'_{in})(n+1)^{|\mathcal{A}||\mathcal{B}|}2^{-nI_{P_{A_iB_j}V_{A'B'|AB}}(A',B';A,B)}. \end{aligned} \tag{3.15}$$

Taking the sum over all $V_{A'B'|AB} \in \mathcal{P}_n(\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B})$, and using the fact (Lemma 3.1)

$$|\mathcal{P}_n(\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B})| \leq (n+1)^{|\mathcal{A}|^2|\mathcal{B}|^2}$$

and $|\mathcal{A}|^2|\mathcal{B}|^2 + |\mathcal{A}||\mathcal{B}| \leq 2|\mathcal{A}|^2|\mathcal{B}|^2$, we obtain

$$\mathbb{E}S_{ij}^{pq} \leq (n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2} m_n(\max_i m'_{in})$$

where

$$S_{ij}^{pq} \triangleq \sum_{V_{A'B'|AB}\in\mathcal{P}_n(\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B})} 2^{nI_{P_{A_iB_j}V_{A'B'|AB}}(A',B';A,B)} \times \left[\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\mathcal{C}_{ij}^{pq}\right|}{4N_iM_{ij}} + \sum_{(k,l)\neq(i,j)}\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\mathcal{C}_{kl}\right|}{4N_kM_{kl}}\right].$$

Immediately, normalizing by $4N_iM_{ij}$ and taking the sum over $1\le i\le m_n$, $1\le j\le m'_{in}$, $1\le p\le N_i$,$1\le q\le M_{ij}$ yields

$$\mathbb{E}\sum_{i=1}^{m_n}\sum_{j=1}^{m'_{in}}\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}S_{ij}^{pq}\le(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2. \tag{3.16}$$

Similarly, it follows from (3.10) and (3.11) that

$$\mathbb{E}\sum_{i=1}^{m_n}\sum_{j=1}^{m'_{in}}\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}K_{ij}^{pq}\le(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n(\max_i m'_{in})^2\le(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2, \tag{3.17}$$

where

$$K_{ij}^{pq} \triangleq \sum_{V_{A'B'|AB}\in\mathcal{P}_n(\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B})} 2^{nI_{P_{A_iB_j}V_{A'B'|AB}}(B';B|A)} \times \left[\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\mathcal{C}_{ij}^{pq}\right|}{4M_{ij}} + \sum_{l\neq j}\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\mathcal{C}_{il}\right|}{4M_{il}}\right],$$

and it follows from (3.12) and (3.14) that

$$\mathbb{E}\sum_{i=1}^{m_n}\sum_{j=1}^{m'_{in}}\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}L_{ij}^{pq}\le(n+1)^{2|\mathcal{A}|^2}m_n^2(\max_i m'_{in})\le(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n(\max_i m'_{in})^2, \tag{3.18}$$

where L_{ij}^{pq} is actually independent of j and q and is given by

$$L_{ij}^{pq} = L_i^p \triangleq \sum_{V_{A'|A}\in\mathcal{P}_n(\mathcal{A}|\mathcal{A})} 2^{nI_{P_{A_i}V_{A'|A}}(A';A)} \times \left[\frac{\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right)\bigcap\mathcal{C}_i^p\right|}{2N_i} + \sum_{k\neq i}\frac{\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right)\bigcap\mathcal{C}_k\right|}{2N_k}\right].$$

Summing (3.16), (3.17) and (3.18) together, we obtain

$$\mathbb{E}\sum_{i=1}^{m_n}\sum_{j=1}^{m'_{in}}\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{pq}+K_{ij}^{pq}+L_{ij}^{pq}\right)\le 3(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2. \tag{3.19}$$

Therefore, there exists at least a selection of these sets $\{\widehat{\mathcal{C}}_i\}_{i=1}^{m_n}$ and $\{\widehat{\mathcal{C}}_{ij}\}_{i=1,j=1}^{i=m_n,j=m'_{in}}$ such that

$$\sum_{i=1}^{m_n}\sum_{j=1}^{m'_{in}}\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{pq}+K_{ij}^{pq}+L_{ij}^{pq}\right)\leq 3(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2,$$

which implies that for all $i=1,2,...,m_n$ and $j=1,2,...,m'_{in}$ the following is satisfied

$$\frac{1}{4N_iM_{ij}}\sum_{p=1}^{2N_i}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{pq}+K_{ij}^{pq}+L_{ij}^{pq}\right)\leq 3(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2. \tag{3.20}$$

We next proceed the proof with an expurgation argument. Without loss of generality, we assume

$$\begin{aligned}\frac{1}{2M_{ij}}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{1q}+K_{ij}^{1q}+L_{ij}^{1q}\right) &\leq \frac{1}{2M_{ij}}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{2q}+K_{ij}^{2q}+L_{ij}^{2q}\right)\leq\cdots\\ &\leq \frac{1}{2M_{ij}}\sum_{q=1}^{2M_{ij}}\left(S_{ij}^{2N_i,q}+K_{ij}^{2N_i,q}+L_{ij}^{2N_i,q}\right),\end{aligned}$$

then we must have, for every $1\leq p\leq N_i$,

$$\frac{1}{2M_{ij}}\sum_{q=1}^{2M_{ij}}S_{ij}^{pq}+K_{ij}^{pq}+L_{ij}^{pq}\leq 6(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2.$$

Similarly, suppose for each $p=1,2,...,N_i$,

$$S_{ij}^{p1}+K_{ij}^{p1}+L_{ij}^{p1}\leq S_{ij}^{p2}+K_{ij}^{p2}+L_{ij}^{p2}\leq\cdots\leq S_{ij}^{p,2M_{ij}}+K_{ij}^{p,2M_{ij}}+L_{ij}^{p,2M_{ij}},$$

the above implies that for each $p=1,2,...,N_i$ and each $q=1,2,...,M_{ij}$,

$$S_{ij}^{pq}+K_{ij}^{pq}+L_{ij}^{pq}\leq 12(n+1)^{2|\mathcal{A}|^2|\mathcal{B}|^2}m_n^2(\max_i m'_{in})^2. \tag{3.21}$$

We now let for $i=1,2,...,m_n$, $p=1,2,,,.,N_i$, $\Omega_i\triangleq\left\{\mathbf{a}_1^{(i)},\mathbf{a}_2^{(i)},...,\mathbf{a}_{N_i}^{(i)}\right\}\subseteq\widehat{\mathcal{C}}_i$, $\Omega_i^p\triangleq\Omega_i/\left\{\mathbf{a}_p^{(i)}\right\}\subseteq\widehat{\mathcal{C}}_i^p$ and for $j=1,2,...,m'_{in}$, $q=1,2,...,M_{ij}$, let $\Omega_{ij}(\mathbf{a}_p^{(i)})=\left\{(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)})\right\}_{q=1}^{M_{ij}}$ such that

$$\begin{aligned}\Omega_{ij}\triangleq\bigcup_{p=1}^{N_i}\Omega_{ij}(\mathbf{a}_p^{(i)}) = &\left\{\left(\mathbf{a}_1^{(i)},\mathbf{b}_{11}^{(j)}\right),\left(\mathbf{a}_1^{(i)},\mathbf{b}_{12}^{(j)}\right),...,\left(\mathbf{a}_1^{(i)},\mathbf{b}_{1,M_{ij}}^{(j)}\right),\right.\\ &\left(\mathbf{a}_2^{(i)},\mathbf{b}_{21}^{(j)}\right),\left(\mathbf{a}_2^{(i)},\mathbf{b}_{22}^{(j)}\right),...,\left(\mathbf{a}_2^{(i)},\mathbf{b}_{2,M_{ij}}^{(j)}\right),\\ &\cdots\cdots\\ &\left.\left(\mathbf{a}_{N_i}^{(i)},\mathbf{b}_{N_i,1}^{(j)}\right),\left(\mathbf{a}_{N_i}^{(i)},\mathbf{b}_{N_i,2}^{(j)}\right),...,\left(\mathbf{a}_{N_i}^{(i)},\mathbf{b}_{N_i,M_{ij}}^{(j)}\right)\right\}\subseteq\widehat{\mathcal{C}}_{ij},\end{aligned}$$

and denote also $\Omega_{ij}^{pq} \triangleq \Omega_{ij}/\left\{\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right)\right\} \subseteq \widehat{\mathcal{C}}_{ij}^{pq}$. Immediately, it follows from (3.21) that for every $i = 1, 2, ..., m_n$, $j = 1, 2, ..., m'_{in}$, $k = 1, 2, ..., m_n$, $l = 1, 2, ..., m'_{kn}$, $p = 1, 2, ..., N_i$, $q = 1, 2, ..., M_{ij}$, and every $V_{A'B'|AB} \in \mathcal{P}_n(\mathcal{A} \times \mathcal{B}|\mathcal{A} \times \mathcal{B})$ and $V_{A'|A} \in \mathcal{P}_n(\mathcal{A}|\mathcal{A})$

$$\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right) \bigcap \Omega_{kl}\right|}{N_k M_{kl}} \leq 2^{-n\left[I_{P_{A_i B_j} V_{A'B'|AB}}(A',B';A,B)-\delta\right]}, \quad (k,l) \neq (i,j), \tag{3.22}$$

$$\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right) \bigcap \Omega_{ij}^{pq}\right|}{N_i M_{ij}} \leq 2^{-n\left[I_{P_{A_i B_j} V_{A'B'|AB}}(A',B';A,B)-\delta\right]}, \tag{3.23}$$

$$\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right) \bigcap \Omega_{il}\right|}{M_{il}} \leq 2^{-n\left[I_{P_{A_i B_j} V_{A'B'|AB}}(B';B|A)-\delta\right]}, \quad l \neq j, \tag{3.24}$$

$$\frac{\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)}, \mathbf{b}_{p,q}^{(j)}\right) \bigcap \Omega_{ij}^{pq}\right|}{M_{ij}} \leq 2^{-n\left[I_{P_{A_i B_j} V_{A'B'|AB}}(B';B|A)-\delta\right]}, \tag{3.25}$$

$$\frac{\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right) \bigcap \Omega_{k}\right|}{N_k} \leq 2^{-n\left[I_{P_{A_i} V_{A'|A}}(A';A)-\delta\right]}, \quad k \neq i, \tag{3.26}$$

$$\frac{\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right) \bigcap \Omega_{i}^{p}\right|}{N_i} \leq 2^{-n\left[I_{P_{A_i} V_{A'|A}}(A';A)-\delta\right]}, \tag{3.27}$$

where

$$\delta = \frac{2}{n}\left[|\mathcal{A}|^2|\mathcal{B}|^2 \log_2(n+1) + \log_2 m_n + \log_2(\max_i m'_{in}) + \log_2 12\right].$$

Now we proved the existence of the sets Ω_i and Ω_{ij} with elements selected uniformly from each $\mathbb{T}_{A_i}$ and $\mathbb{T}_{A_i B_j}$ satisfying the inequalities (3.22)–(3.27) for any $V_{A'|A}$ and $V_{A'B'|AB}$. It remains to show that these sets are disjoint and have distinct elements provided assumptions (3.2) and (3.3). Indeed, since (3.26) and (3.27) hold for every $V_{A'|A} \in \mathcal{P}_n(\mathcal{A}|\mathcal{A})$, they of course hold when $V_{A'|A}$ is a conditional distribution such that $V^*_{A'|A}(a'|a)$ is 1 if $a' = a$ and 0 otherwise. It then follows from (3.2)

$$\frac{1}{n}\log_2 N_i < H_{P_{A_i}}(A) - \delta = I_{P_{A_i} V^*_{A'|A}}(A';A) - \delta$$

that $\left|\mathbb{T}_{V^*_{A'|A}}\left(\mathbf{a}_p^{(i)}\right) \bigcap \Omega_k\right| = \left|\left\{\mathbf{a}_p^{(i)}\right\} \bigcap \Omega_k\right| < 1$ or equivalently, $\left|\left\{\mathbf{a}_p^{(i)}\right\} \bigcap \Omega_k\right| = 0$, which means any elements in Ω_i does not belong to Ω_k for $i \neq k$, i.e., Ω_i and Ω_k are disjoint. Likewise, using assumption (3.2) in (3.27), we see that

$$\left|\mathbb{T}_{V^*_{A'|A}}\left(\mathbf{a}_p^{(i)}\right) \bigcap \Omega_i^p\right| = \left|\left\{\mathbf{a}_p^{(i)}\right\} \bigcap \Omega_i^p\right| = 0,$$

which means that Ω_i has N_i disjoint elements. Similarly, setting $V_{A'B'|AB}$ be the conditional distribution such that $V^*_{A'B'|AB}(a',b'|a,b)$ is 1 if $a'=a$, $b'=b$ and 0 otherwise, and using (3.3)

$$\frac{1}{n}\log_2 M_{ij} < H_{P_{A_i}P_{B_j|A_i}}(B|A)-\delta,$$

we see that for any $\mathbf{a}_p^{(i)}\in\Omega_i$, $\Omega_{ij}(\mathbf{a}_p^{(i)})$'s are disjoint and the elements in $\Omega_{ij}(\mathbf{a}_p^{(i)})$ are all distinct, i.e., $|\Omega_{ij}(\mathbf{a}_p^{(i)})| = M_{ij}$ for every $\mathbf{a}^{(i)}\in\Omega_i$. Finally, when $V_{A'|A}$ is not the conditional distribution such that $V_{A'|A}(a'|a)$ is 1 if $a'=a$ and 0 otherwise, we can write (3.26) and (3.27) in the same way as (3.4), and when $V_{A'B'|AB}$ is not the conditional distribution such that $V_{A'B'|AB}(a',b'|a,b)$ is 1 if $a'=a$, $b'=b$ and 0 otherwise, we can write (3.22)–(3.23) as (3.5), and write (3.24)–(3.25) as (3.6), since

$$\begin{aligned}
\left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right)\bigcap\Omega_i^p\right| &= \left|\mathbb{T}_{V_{A'|A}}\left(\mathbf{a}_p^{(i)}\right)\bigcap\Omega_i\right|,\\
\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\Omega_{ij}^{pq}\right| &= \left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\Omega_{ij}\right|,\\
\left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\Omega_{ij}^{pq}\right| &= \left|\mathbb{T}_{V_{A'B'|AB}}\left(\mathbf{a}_p^{(i)},\mathbf{b}_{p,q}^{(j)}\right)\bigcap\Omega_{ij}\right|.
\end{aligned}$$

■

3.3 Type Classes for Sequences with Continuous Alphabets

Partitioning the sequence space into disjoint type classes (type sets) and analyzing the probability of a particular event for each single type class is the essential idea of the method of types. The type and type class defined in the sense of a common composition of single-symbol frequencies, however, cannot be implemented to sequences with a continuous alphabet. In order to apply the method of types to sequences in a continuous space, say the k-dimensional Euclidean space, we need to find a counterpart to the type classes which partition the whole space $\mathbb{R}^k$, while keeping an exponentially small probability with respect to k.

3.3.1 Gaussian-Type Classes

In [6, Sec. VI. A], a continuous-alphabet analog to the method of types was studied for the MGS by introducing the notion of Gaussian-type classes. Given $\sigma^2 > 0$ and $\epsilon \in (0, \sigma^2)$, the Gaussian-type class, denoted by $\mathbb{T}^{\epsilon}(\sigma^2)$, is the set of all k-length sequences $\mathbf{s} \in \mathbb{R}^k$ such that

$$|\mathbf{s}^t\mathbf{s} - k\sigma^2| \leq k\epsilon, \tag{3.28}$$

where t is the transpose operation. A typical 2-dimensional Gaussian-type class with $\sigma^2 = 5$ and $\epsilon = 1$ is a ring and is shown in Fig. 3.2 (a). Based on a sequence of (appropriate) positive parameters $\{\sigma_i^2\}_{i=1}^{\infty}$, say $\sigma_1^2 = 5$ and $\sigma_i^2 = \sigma_{i-1}^2 + 2$ for $i \geq 2$, the Euclidean space $\mathbb{R}^k$, say $k = 2$, can be partitioned using the sequence of rings given by (3.28) together with the disc $\{\mathbf{s} : \mathbf{s}^t\mathbf{s} \leq k\epsilon\}$; see Fig. 3.2 (b). Like the discrete type classes, the size (which is volume here) of a particular Gaussian-type class grows exponentially with the dimension k, and the probability of each type class defined by (3.28) under a zero-mean Gaussian distribution decays exponentially in k; see the following lemma.

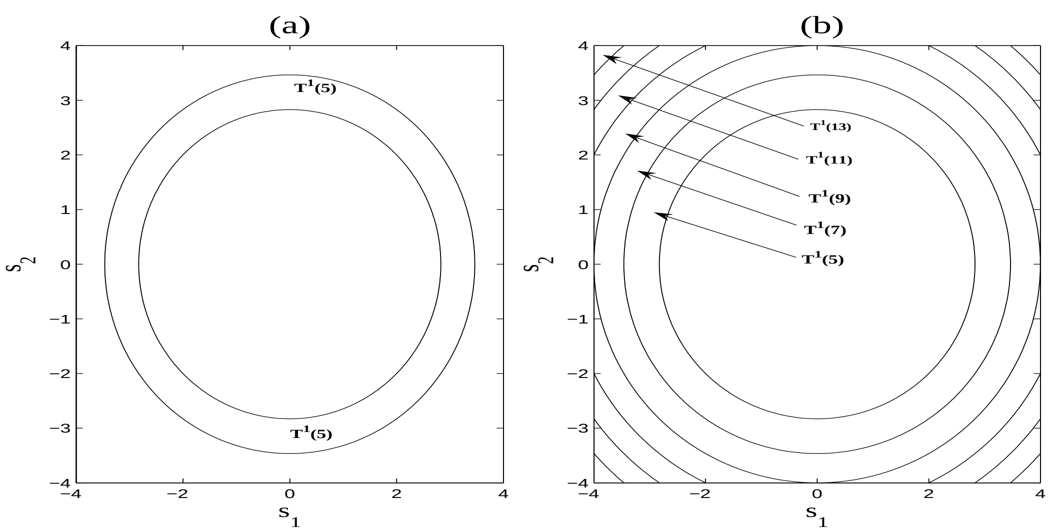

Figure 3.2: (a) A typical Gaussian-type class with $k = 2$, $\sigma^2 = 5$, $\epsilon = 1$; (b) $\mathbb{R}^2$ is partitioned by a sequence of Gaussian-type classes.

Lemma 3.3 [6] *Let* $Vol\{A\} \triangleq \int_{s \in A} d\boldsymbol{s}$ *be the volume of set A. For any* $\widehat{\sigma}_S^2 > \epsilon > 0$*, we have*

$$\left(1 - \frac{2\widehat{\sigma}_S^4}{k\epsilon^2}\right)\left[2\pi\widehat{\sigma}_S^2 e^{1-\frac{\epsilon}{\widehat{\sigma}_S^2}}\right]^{k/2} \leq Vol\left\{\mathbb{T}^\epsilon(\widehat{\sigma}^2)\right\} \leq [2\pi e(\widehat{\sigma}^2 + \epsilon)]^{k/2}. \tag{3.29}$$

Furthermore, let $Q_S \sim \mathcal{N}(0, \sigma_S^2)$ *be a Gaussian distribution with mean zero and variance* σ_S^2*. Then*

$$\begin{aligned}
&\left(1 - \frac{2\widehat{\sigma}_S^4}{k\epsilon^2}\right)\exp\left\{-k\left(D(\widehat{Q}_S \parallel Q_S) + \zeta_0(\epsilon)\right)\right\} \\
&\leq Q_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right) \\
&\leq \exp\left\{-k\left(D(\widehat{Q}_S \parallel Q_S) + \zeta_1(\epsilon)\right)\right\},
\end{aligned} \tag{3.30}$$

where $\widehat{Q}_S \backsim \mathcal{N}(0, \widehat{\sigma}_S^2)$*,*

$$D(\widehat{Q}_S \parallel Q_S) = \frac{1}{2}\left(\frac{\widehat{\sigma}_S^2}{\sigma_S^2} - \ln\frac{\widehat{\sigma}_S^2}{\sigma_S^2} - 1\right) \tag{3.31}$$

is the Kullback-Leibler divergence between the two Gaussian distributions $\widehat{Q}_S$ *and* Q_S*,*

$$\zeta_0(\epsilon) = \frac{1}{2}\left(\frac{\epsilon}{\sigma_S^2} + \frac{\epsilon}{\widehat{\sigma}_S^2}\right),$$

and

$$\zeta_1(\epsilon) = -\frac{\epsilon}{\sigma_S^2} - \ln\left(1 + \frac{\epsilon}{\widehat{\sigma}_S^2}\right).$$

Remark 3.1 Since the exponentially vanishing probability $Q_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right)$ does not hold in general, the Gaussian-type class can only be applied to MGS's.

Proof: Consider an auxiliary zero-mean Gaussian distribution with variance $\widehat{\sigma}_S^2 + \epsilon$. Then the upper bound of Vol $\left\{\mathbb{T}^\epsilon(\widehat{\sigma}^2)\right\}$ follows from

$$\begin{aligned}
1 &\geq \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s} - k\widehat{\sigma}_S^2| \leq k\epsilon\}} \frac{1}{\left[\sqrt{2\pi(\widehat{\sigma}_S^2 + \epsilon)}\right]^k} e^{-\frac{\mathbf{s}^t\mathbf{s}}{2(\widehat{\sigma}_S^2+\epsilon)}} d\mathbf{s} \\
&\geq \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s} - k\widehat{\sigma}_S^2| \leq k\epsilon\}} \frac{1}{\left[\sqrt{2\pi(\widehat{\sigma}_S^2 + \epsilon)}\right]^k} e^{-\frac{k((\widehat{\sigma}_S^2+\epsilon))}{2(\widehat{\sigma}_S^2+\epsilon)}} d\mathbf{s} \\
&= \frac{1}{\left[2\pi e(\widehat{\sigma}_S^2 + \epsilon)\right]^{k/2}} \text{Vol}\left\{\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right\}.
\end{aligned}$$

To lower bound the volume, consider an auxiliary MGS $\widetilde{Q}_S \backsim \mathcal{N}(0, \widehat{\sigma}_S^2)$ with k-tuple distribution

$$\widetilde{Q}_S^{(k)}(\mathbf{s}) = \frac{1}{\left[2\pi\widehat{\sigma}_S^2\right]^{k/2}} e^{-\frac{\mathbf{s}^t\mathbf{s}}{2\widehat{\sigma}_S^2}}.$$

Based on $\widetilde{Q}_S$, we can bound the probability

$$\begin{aligned}
\widetilde{Q}_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right) &= \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[2\pi\widehat{\sigma}_S^2\right]^{k/2}} e^{-\frac{\mathbf{s}^t\mathbf{s}}{2\widehat{\sigma}_S^2}} d\mathbf{s} \\
&\le \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[2\pi\widehat{\sigma}_S^2\right]^{k/2}} e^{-\frac{k(\widehat{\sigma}_S^2-\epsilon)}{2\widehat{\sigma}_S^2}} d\mathbf{s} \\
&= \left[\frac{1}{2\pi\widehat{\sigma}_S^2 e^{1-\frac{\epsilon}{\widehat{\sigma}_S^2}}}\right]^{k/2} \mathrm{Vol}\{\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\}. \qquad (3.32)
\end{aligned}$$

On the other hand, we can upper bound

$$\begin{aligned}
1 - \widetilde{Q}_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right) &= \widetilde{Q}_S^{(k)}\left(\left|\frac{1}{k}\mathbf{s}^t\mathbf{s} - \widehat{\sigma}_S^2\right| > \epsilon\right) \\
&\le \frac{2\widehat{\sigma}_S^4}{k\epsilon^2}, \qquad (3.33)
\end{aligned}$$

where the inequality follows from the Chebychev inequality by noting that $\mathbb{E}[s_i^2] = \widetilde{\sigma}_S^2$. Plugging (3.33) into (3.32) yields the left inequality in (3.29). Similarly, (3.30) follows from

$$\begin{aligned}
Q_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right) &= \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[\sqrt{2\pi\sigma_S^2}\right]^{k}} e^{-\frac{\mathbf{s}^t\mathbf{s}}{2\sigma_S^2}} d\mathbf{s} \\
&\le \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[\sqrt{2\pi\sigma_S^2}\right]^{k}} e^{-\frac{k(\widehat{\sigma}_S^2-\epsilon)}{2\sigma_S^2}} d\mathbf{s} \\
&\le e^{-\frac{k(\widehat{\sigma}_S^2-\epsilon)}{2\sigma_S^2}} \left[\frac{e(\widehat{\sigma}_S^2+\epsilon)}{\sigma_S^2}\right]^{k/2} \qquad (3.34)
\end{aligned}$$

and

$$\begin{aligned}
Q_S^{(k)}\left(\mathbb{T}^\epsilon(\widehat{\sigma}_S^2)\right) &= \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[\sqrt{2\pi\sigma_S^2}\right]^{k}} e^{-\frac{\mathbf{s}^t\mathbf{s}}{2\sigma_S^2}} d\mathbf{s} \\
&\ge \int_{\{\mathbf{s}:|\mathbf{s}^t\mathbf{s}-k\widehat{\sigma}_S^2|\le k\epsilon\}} \frac{1}{\left[\sqrt{2\pi\sigma_S^2}\right]^{k}} e^{-\frac{k(\widehat{\sigma}_S^2+\epsilon)}{2\sigma_S^2}} d\mathbf{s} \\
&\ge e^{-\frac{k(\widehat{\sigma}_S^2+\epsilon)}{2\sigma_S^2}} \left(1 - \frac{2\widehat{\sigma}_S^4}{k\epsilon^2}\right) \left[\frac{\widehat{\sigma}_S^2}{\sigma_S^2} e^{1-\frac{\epsilon}{\widehat{\sigma}_S^2}}\right]^{k/2} \qquad (3.35)
\end{aligned}$$

where (3.34) and (3.35) follow from (3.29). ■

3.3.2 Laplacian-Type Classes

Of course, in addition to (3.28), there are other ways to partition the whole Euclidean space $\mathbb{R}^k$. For given $\alpha > 0$ and $0 < \epsilon < \alpha$, we define a Laplacian-type class $\mathbb{T}^\epsilon(\alpha)$ by the set of all k-vectors $\mathbf{s} \in \mathbb{R}^k$ such that $\left|\sum_{i=1}^k |s_i| - k\alpha\right| \leq k\epsilon$, i.e.,

$$\mathbb{T}^\epsilon(\alpha) \triangleq \left\{ \mathbf{s} : \left| \sum_{i=1}^k |s_i| - k\alpha \right| \leq k\epsilon \right\}. \tag{3.36}$$

A typical shape of a 2-dimensional Laplacian-type class with $\alpha = 5$ and $\epsilon = 1$ is shown in Fig. 3.3 (a). Based on a sequence of (appropriate) positive parameters $\{\alpha_i\}_{i=1}^\infty$, say $\alpha_1 = 5$ and $\alpha_i = \alpha_{i-1} + 2$ for $i \geq 2$, the Euclidean space $\mathbb{R}^k$, say $k = 2$, can be partitioned using the sequence of $\mathbb{T}^\epsilon(\alpha)$ together with the rhombus $\{\mathbf{s} : \sum_{i=1}^k |s_i| \leq k\epsilon\}$; see Fig. 3.3 (b). We can bound the volume of the Laplacian-type class as for the Gaussian-type class. It turns out that the probability of a Laplacian-type class vanishes exponentially under a zero-mean Laplacian distribution. Thus, analogously to the Gaussian-type class, the Laplacian-type class can be used to deal with memoryless Laplacian sources (MLSs).

Lemma 3.4 *For any $\widetilde{\alpha} > \epsilon > 0$, we have*

$$\left[1 - \frac{2\widetilde{\alpha}^2}{k\epsilon^2}\right] \left(2\widetilde{\alpha} e^{1-\frac{\epsilon}{\widetilde{\alpha}}}\right)^k \leq Vol\{\mathbb{T}^\epsilon(\widetilde{\alpha})\} \leq [2e(\widetilde{\alpha} + \varepsilon)]^k. \tag{3.37}$$

Furthermore, let $Q_S \sim \mathcal{L}(0, \alpha)$ be a Laplacian distribution with mean zero and first moment α (or variance $2\alpha^2$ equivalently). Then

$$\begin{aligned} &\left[1 - \frac{2\widetilde{\alpha}^2}{k\epsilon^2}\right] \exp\left\{-k\left(D(\widetilde{Q}_S \parallel Q_S) + \widehat{\zeta}_0(\epsilon)\right)\right\} \\ \leq\ & Q_S^{(k)}(\mathbb{T}^\epsilon(\widetilde{\alpha})) \\ \leq\ & \exp\left\{-k\left(D(\widetilde{Q}_S \parallel Q_S) + \zeta_2(\epsilon)\right)\right\} \end{aligned} \tag{3.38}$$

where

$$D(\widetilde{Q}_S \parallel Q_S) = \frac{\widetilde{\alpha}}{\alpha} - \ln\frac{\widetilde{\alpha}}{\alpha} - 1 \tag{3.39}$$

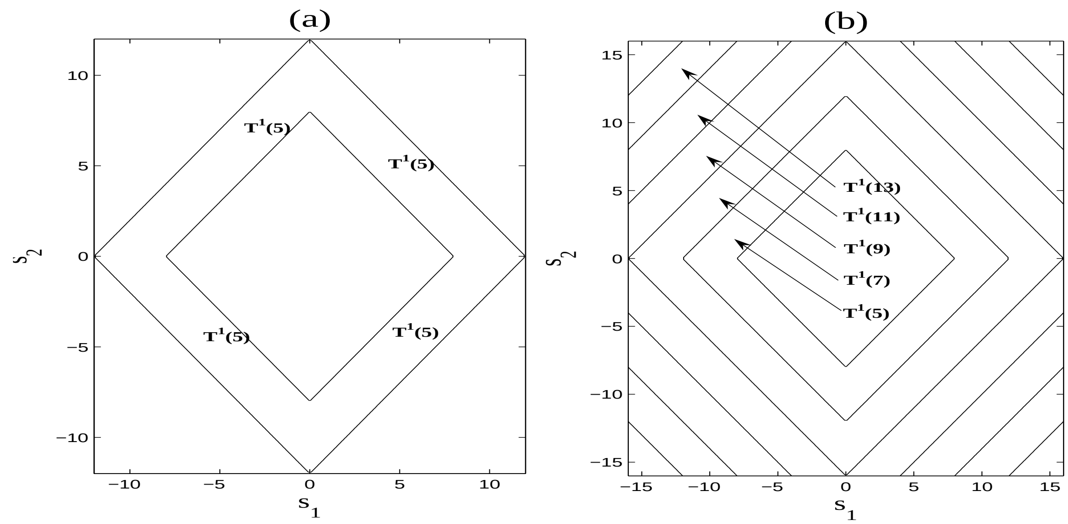

Figure 3.3: (a) A typical Laplacian-type class with $k = 2$, $\alpha = 5$, $\epsilon = 1$; (b) $\mathbb{R}^2$ is partitioned by a sequence of Laplacian-type classes.

is the Kullback-Leibler divergence between the two Laplaxian distributions $\widetilde{Q}_S$ *and* Q_S*,*

$$\widehat{\zeta}_0(\epsilon) = \frac{\epsilon}{\alpha} + \frac{\epsilon}{\widetilde{\alpha}},$$

and

$$\zeta_2(\epsilon) = -\frac{\epsilon}{\alpha} - \ln\left(1 + \frac{\epsilon}{\widetilde{\alpha}}\right).$$

The proof is similar to the last one and is omitted. ■

3.4 Type Covering Lemmas for Discrete and Continuous Type Classes

When we transmit an information source over a noisy or noiseless channel with a fidelity criterion, we usually expect that the resulting distortion between the original source sequence and the recovered sequence is less than some distortion threshold, say Δ, with a high probability. To design a code with a small probability of exceeding the distortion Δ, we can employ the "Δ-admissible" quantization approach proposed by Berger [15]. Specifically, we first perform vector quantization on each source sequence with quantization error less than Δ, then we transmit these quantized discrete sequences (i.e., quantization cells) and estimate them at the receiver losslessly. The coding scheme will be described in detail in Chapter 8, where we apply it to obtain an upper bound for the probability of excess distortion.

In this section, we address the following quantization problem: how many codewords are needed to ensure that for every source sequence, say $\mathbf{s}$ in $\mathcal{S}^k$, there exists a codeword $\mathbf{c}$ with distortion less than Δ? In other words, from a geometric point of view, how many k-dimensional balls of size Δ do we need to at least entirely cover the whole source space $\mathcal{S}^k$? When $\mathcal{S}$ is a finite alphabet, this problem is tackled in [15] by the method of types (also cf. [32]). Recalling that the source space $\mathcal{S}^k$ can be partitioned by a polynomial number of different type classes (with respect to k), the following type covering lemma states that, for each particular type, exponential number of Δ-size balls (with respect to k) are required to cover the type class (see Fig. 3.4).

Lemma 3.5 (**Covering Lemma for Discrete Type Classes** [15, 32]) *Given $\mu > 0$, for each sufficiently large k depending only on the distortion measure $d(\cdot, \cdot)$ and μ, for every type class there exists a set $\mathcal{C}_{P_S} \subset \mathcal{S}^k$ of size*

$$|\mathcal{C}_{P_S}| \leq \exp\{k[R(P_S, \Delta) + \mu]\}$$

such that every sequence $\boldsymbol{s} \in \mathbb{T}_{P_S}$ is contained, for some $\boldsymbol{c}_{P_S} \in \mathcal{C}_{P_S}$, in the ball of size Δ

$$B(\boldsymbol{c}_{P_S}, \Delta) \triangleq \left\{ \boldsymbol{s} : d^{(k)}(\boldsymbol{s}, \boldsymbol{c}_{P_S}) \leq \Delta \right\},$$

where $R(P_S, \Delta)$ is the rate-distortion function of the DMS P_S.

When $\mathcal{S} = \mathbb{R}$, we have similar results for the Gaussian-type and Laplacian-type classes. The type covering lemma for Gaussian-type classes is proved in [6]. We only give the proof for the type covering lemma for Laplacian-type classes.

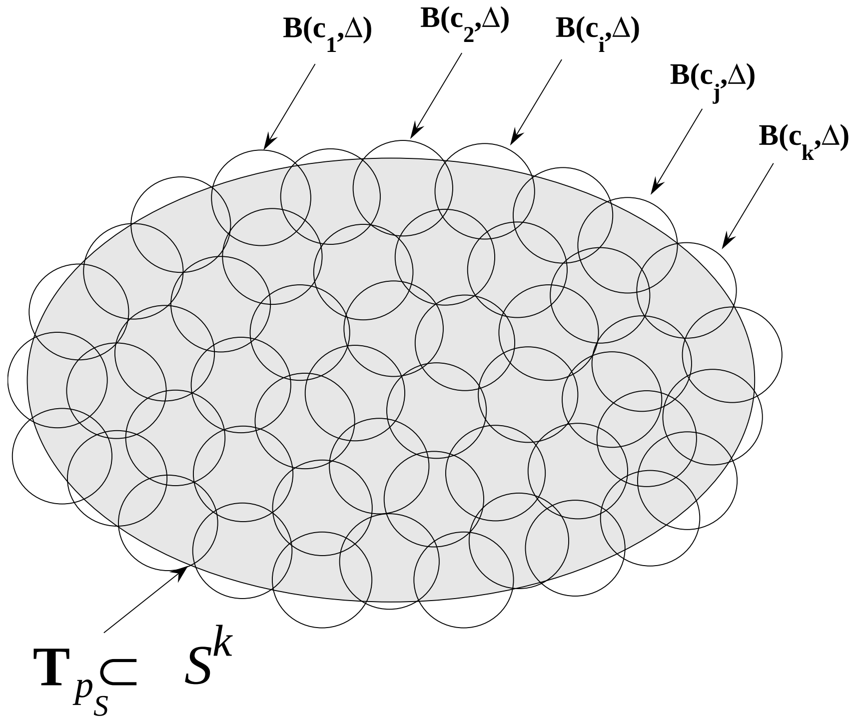

Figure 3.4: A graphical illustration of the type covering lemma. We need at least $\approx \exp\{kR(P_S,\Delta)\}$ k-dimensional balls of size Δ to entirely cover the type class $\mathbb{T}_{P_S}$.

Lemma 3.6 (**Covering Lemma for Gaussian-Type Classes** [6]) *Given* $\sigma_S^2 > \Delta$ *and* $\mu > 0$*, for sufficiently small* ϵ *and for sufficiently large* k*, there exists a set* $\mathcal{C} \subset \mathbb{R}^k$ *of size*

$$|\mathcal{C}| \leq \exp\{k[R(P_S,\Delta) + \zeta_3(\epsilon)] + \mu\}$$

with

$$\zeta_3(\epsilon) = \frac{1}{2}\ln\frac{\Delta}{(\sqrt{\Delta}-\epsilon)^2 - \epsilon\Delta\left(1+4\sqrt{\frac{\Delta}{\sigma_S^2}}\right)} + 2\epsilon + 2\ln\left[1+\epsilon\left(1+4\sqrt{\frac{\Delta}{\sigma_S^2-\Delta}}\right)\right]$$

such that every sequence $\boldsymbol{s} \in \mathbb{T}^{\epsilon}(\sigma_S^2)$ *is contained, for some* $\boldsymbol{c} \in \mathcal{C}$*, in the ball of size* Δ

$$B(\boldsymbol{c},\Delta) \triangleq \left\{\boldsymbol{s} : \frac{1}{k}\sum_{i=1}^{k}(s_i - c_i)^2 \leq \Delta\right\},$$

where $R(P_S,\Delta)$ *is the rate-distortion function of MGS* $P_S \sim \mathcal{N}(0,\sigma_S^2)$*.*

Lemma 3.7 (**Covering Lemma for Laplacian-Type Classes**) *Given* $\alpha > \Delta$ *and* $\mu > 0$*, for sufficiently small* ϵ *and for sufficiently large* k*, there exists a set* $\mathcal{C} \subset \mathbb{R}^k$ *of size* $|\mathcal{C}| \leq \exp\{k[R(P_S, \Delta) + \zeta_4(\epsilon)] + \mu\}$ *with*

$$\zeta_4(\epsilon) = \ln \frac{\Delta}{\Delta - \epsilon} + \ln\left(1 + \frac{\epsilon}{\alpha - \Delta + \epsilon}\right) + \frac{2\alpha\epsilon}{(\alpha - \Delta + \epsilon)(\Delta - \epsilon)}$$

such that every sequence in $\mathbb{T}^\epsilon(\alpha)$ *is contained, for some* $\boldsymbol{c} \in \mathcal{C}$*, in the ball (cube)*

$$B(\boldsymbol{c}, \Delta) \triangleq \left\{\boldsymbol{s} : \frac{1}{k}\sum_{i=1}^{k} |s_i - c_i| \leq \Delta\right\}$$

of size Δ*, where* $R(P_S, \Delta)$ *is the rate distortion function of Laplacian source* $P_S \backsim L(0, \alpha)$*.*

Proof: Before we proceed with the proof, we introduce a "shifted" Laplacian-type class. Given a sequence $\mathbf{s}^* = (s_1^* \cdots s_k^*)$, denote

$$\mathbb{T}^\epsilon(\alpha|\mathbf{s}^*) \triangleq \left\{\mathbf{s} : \left|\sum_{i=1}^{k} |s_i - s_i^*| - k\alpha\right| \leq k\epsilon\right\}.$$

Clearly, $\mathbb{T}^\epsilon(\alpha|\mathbf{s}^*)$ is a shifted set generated from $\mathbb{T}^\epsilon(\alpha)$ and by Lemma 3.4

$$\mathrm{Vol}\{\mathbb{T}^\epsilon(\alpha|\mathbf{s}^*)\} = \mathrm{Vol}\{\mathbb{T}^\epsilon(\alpha)\} \geq \left[1 - \frac{\alpha^2}{k\epsilon^2}\right]\left(2\alpha e^{1-\frac{\epsilon}{\alpha}}\right)^k.$$

We start by assuming that $\alpha \geq \Delta$ since when $\alpha < \Delta$ the type class $\mathbb{T}^\epsilon(\alpha)$ is covered by the ball $B(\mathbf{0}, \Delta)$ for ϵ sufficiently small ($\epsilon < \Delta - \alpha$), i.e., for $\alpha < \Delta$ and for $\epsilon < \Delta - \alpha$ there exists a code with size $|\mathcal{C}| = 1$ that covers $\mathbb{T}^\epsilon(\alpha)$.

Construct a grid $X(\delta)$ of all vectors in the Euclidean space $\mathbb{R}^k$ whose components are integer multiples of δ for some small $0 < \delta < \Delta$ (we set $\delta = \epsilon$ in the following) and consider the k-dimensional balls of size δ, centered at the grid points. Now we randomly (independently) choose M vectors $\mathbf{c}^{(1)}, ..., \mathbf{c}^{(M)}$ in the set $\mathbb{T}^\xi(\alpha - (\Delta - \delta))$ according to the uniform pdf $P(\mathbf{c}) \triangleq 1/\mathrm{Vol}\left\{\mathbb{T}^\xi(\alpha - (\Delta - \delta))\right\}$, where $\xi \triangleq \left[1 + \left(1 - \frac{\Delta - \delta}{\alpha}\right)^2\right]\epsilon$ and

$$\exp\left\{k[R(P_S, \Delta) + \zeta_4(\epsilon)] + \frac{\mu}{2}\right\} \leq M \leq \exp\left\{k[R(P_S, \Delta) + \zeta_4(\epsilon)] + \mu\right\}. \tag{3.40}$$

Define the set $U(\Delta)$ by

$$U(\Delta) = \left\{\mathbf{s} \in \mathbb{T}^\epsilon(\alpha) \bigcap X(\delta) : \frac{1}{k}\sum_{j=1}^{k} |s_j - c_j^{(i)}| > \Delta - \delta, \quad \text{for all } i = 1, 2, ..., M\right\}.$$

Clearly, $U(\Delta)$ is a set of all grid points in set $\mathbb{T}^{\epsilon}(\alpha)$ which are not covered by the codewords in $\mathcal{C} \triangleq \{\mathbf{c}^{(1)}, ..., \mathbf{c}^{(M)}\}$ of size M within distortion threshold $\Delta - \delta$. Now if we can show that $\mathbb{E}_P|U(\Delta)| < 1$, where the expectation is taken under the uniform distribution $P(\mathbf{c})$, then there must exist a code for which $U(\Delta)$ is empty. For such code $\mathcal{C}$, $U(\Delta)$ is covered by the union of cubes $B\left(\mathbf{c}^{(i)}, \Delta - \delta\right)$, $i = 1, 2, ..., M$, which then implies that, $\mathbb{T}^{\epsilon}(\alpha)$ is entirely covered by the union of cubes $B\left(\mathbf{c}^{(i)}, \Delta\right)$. According to the above random selection assumption,

$$\begin{aligned} \mathbb{E}_P|U(\Delta)| &= \mathbb{E}_P\left\{\sum_{\mathbf{s}\in\mathbb{T}^{\epsilon}(\alpha)\bigcap X(\delta)} \prod_{i=1}^{M} \mathbb{1}\left(\frac{1}{k}\sum_{j=1}^{k}|s_j - c_j^{(i)}| > \Delta - \delta\right)\right\} \\ &= \sum_{\mathbf{s}\in\mathbb{T}^{\epsilon}(\alpha)\bigcap X(\delta)} \prod_{i=1}^{M}\left[1 - P\left\{\mathbf{c}^{(i)} : \frac{1}{k}\sum_{j=1}^{k}|s_j - c_j^{(i)}| \le \Delta - \delta\right\}\right]. \end{aligned} \tag{3.41}$$

Now for each $\mathbf{s} \in \mathbb{T}^{\epsilon}(\alpha)$, we consider the auxiliary set $\mathbb{T}^{\epsilon}\left(D - \frac{D^2}{\alpha}\middle|\left(1 - \frac{D}{\alpha}\right)^2\mathbf{s}\right)$ where $D \triangleq \Delta - \delta < \alpha$. It is seen that $\mathbb{T}^{\epsilon}\left(D - \frac{D^2}{\alpha}\middle|\left(1 - \frac{D}{\alpha}\right)^2\mathbf{s}\right) \subseteq \mathbb{T}^{\xi}(\alpha - D)$ since

$$k\left(1 - \frac{D}{\alpha}\right)^2(\alpha + \epsilon) + k\left(D - \frac{D^2}{\alpha}\right) + k\epsilon = k(\alpha - D + \xi)$$

and

$$k\left(1 - \frac{D}{\alpha}\right)^2(\alpha - \epsilon) - k\left(D - \frac{D^2}{\alpha}\right) - k\epsilon = k(\alpha - D - \xi),$$

and similarly it can be readily checked that for any $\mathbf{s} \in \mathbb{T}^{\epsilon}(\alpha)$

$$\mathbb{T}^{\epsilon}\left(D - \frac{D^2}{\alpha}\middle|\left(1 - \frac{D}{\alpha}\right)^2\mathbf{s}\right) \subseteq \left\{\mathbf{c}^{(i)} : \frac{1}{k}\sum_{j=1}^{k}|s_j - c_j^{(i)}| \le D\right\}.$$

Since the codewords are distributed uniformly in $\mathbb{T}^{\xi}(\alpha - D)$, applying Lemma 3.4 and recalling that $\delta = \epsilon$ we have

$$\begin{aligned} P\left\{\mathbf{c}^{(i)} : \frac{1}{k}\sum_{j=1}^{k}|s_j - c_j^{(i)}| \le D\right\} &\ge \frac{\mathrm{Vol}\left\{\mathbb{T}^{\epsilon}\left(D - \frac{D^2}{\alpha}\middle|\left(1 - \frac{D}{\alpha}\right)^2\mathbf{s}\right)\right\}}{\mathrm{Vol}\left\{\mathbb{T}^{\xi}(\alpha - D)\right\}} \\ &\ge \exp\left\{-k\left[\ln\frac{\alpha}{\Delta} + \widetilde{\zeta}_4(\epsilon)\right] + o(k)\right\}, \end{aligned} \tag{3.42}$$

where

$$\widetilde{\zeta}_4(\epsilon) = -\ln\frac{\Delta - \epsilon}{\Delta} + \ln\left(1 + \frac{\epsilon}{\alpha - D}\right) + \frac{\xi\alpha}{D(\alpha - D)} = \zeta_4(\epsilon)$$

and $o(k) = \ln\left[1 - \frac{D - D^2/\alpha}{k\xi^2}\right]$. Substituting (3.42) into (3.41) yields

$$\mathbb{E}_P|U(\Delta)| \leq |\mathbb{T}^\epsilon(\alpha)\bigcap X(\delta)|\left[1 - \exp\left\{-k\left[\ln\frac{\alpha}{\Delta} + \zeta_4(\epsilon)\right] + o(k)\right\}\right]^M \quad (3.43)$$

$$\leq \left[\frac{2e(\alpha+\epsilon)}{\delta}\right]^k \exp\left\{-M\exp\left\{-k\left[\ln\frac{\alpha}{\Delta} + \zeta_4(\epsilon)\right] + o(k)\right\}\right\}, \quad (3.44)$$

where (3.43) holds since each codeword is independently selected and (3.44) follows from the inequality $(1-x)^M \leq e^{-Mx}$ and the fact that the number of balls in $\mathbb{T}^\epsilon(\alpha)$ is bounded by the ratio between the volumes of $\mathbb{T}^\epsilon(\alpha)$ and of a ball δ^k. From (3.44) we note that for sufficiently small ϵ ($\epsilon < D - D^2/\alpha$), $\delta = \epsilon$, and any given $\mu > 0$, there exists a set of codewords with size M of exponential order $\exp\{k[\ln(\alpha/\Delta) + \zeta_4(\epsilon)] + \mu\}$ (see (3.40)) such that $|U(\Delta)| = 0$ as k goes to infinity, which means that there exists a code of such exponential size covering $\mathbb{T}^\epsilon(\alpha)$ within distortion Δ for sufficiently large k. ■

3.5 Concluding Remarks

In this chapter we established some necessary background on the method of types. We reviewed the properties of classical discrete types and type classes, and we introduced two continuous type classes: the Gaussian-type class and the Laplacian type class. The important feature of the Gaussian-type class (Laplacian-type class) is that the volume of each type class is exponentially large, while the probability of each type class under a Gaussian (Laplacian) product distribution is exponentially small. A generalized joint type packing lemma was developed, which will be used to upper bound the probability of error for coding DMS's over DMC's and 2-user asymmetric channels in Chapters 6 and 9, respectively. We also summarized different versions of type covering lemmas, which will be used to bound the size of a quantization codebook in a "Δ-admissible" coding scheme in Chapter 8. We remark that except for the packing lemma and the covering lemma, the method of types will be frequently used throughout Chapters 6–9. Markov types will be further discussed in Chapter 7.

Chapter 4

Conjugate Functions: Fenchel Transforms

In this chapter, we introduce conjugate convex/concave functions and study their applications to source and channel reliability functions. An important duality result, called Fenchel duality theorem, is presented.

In Sections 4.1, we introduce conjugate convex/concave functions (which are also called convex/concave Fenchel transforms, or convex/concave Fenchel-Legendre transforms in the literature) and we present the Fenchel duality theorem regarding the optimization of the sum of two convex (not necessarily differentiable) functions, which will be widely used in the following chapters. In Section 4.2, we recast the source and channel reliability functions introduced in Chapter 2 and derive their conjugate functions. Several pairs of Fenchel transforms are revealed. We finally state concluding remarks in Section 4.3.

4.1 Conjugate Functions and Fenchel Duality Theorem

For any function f defined on $F \subset \mathbb{R}$, define its convex Fenchel transform (conjugate function, Legendre transform 4.1) f^* by

$$f^*(y) \triangleq \sup_{x \in F}[xy - f(x)]$$

and let F^* be the set $\{y : f^*(y) < \infty\}$. It is easy to see from its definition that f^* is a convex function on F^*. Moreover, if f is convex and continuous, then $(f^*)^* = f$. More generally, $f^{**} \leq f$ and f^{**} is the convex hull of f, *i.e.* the largest convex function that is bounded above by f [79, Section 3], [17, Section 7.1].

Similarly, for any function g defined on $G \subset \mathbb{R}$, define its concave Fenchel transform g_* by

$$g_*(y) \triangleq \inf_{x \in G} [xy - g(x)]$$

and let G_* be the set $\{y : g_*(y) > -\infty\}$. It is easy to see from its definition that g_* is a concave function on G_*. Moreover, if g is concave and continuous, then $(g_*)_* = g$. More generally, $g_{**} \geq g$ and g_{**} is the concave hull of g, *i.e.* the smallest concave function that is bounded below by g.

Fenchel Duality Theorem [65, p. 201] *Assume that f and g are, respectively, convex and concave functions on the non-empty intervals F and G in $\mathbb{R}$ and assume that $F \cap G$ has interior points. Suppose further that $\mu = \inf_{x \in F \cap G}[f(x) - g(x)]$ is finite. Then*

$$\mu = \inf_{x \in F \cap G} [f(x) - g(x)] = \max_{y \in F^* \cap G_*} [g_*(y) - f^*(y)], \tag{4.1}$$

where the maximum on the right is achieved by some $y_0 \in F^* \cap G_*$. If the infimum on the left is achieved by some $x_0 \in F \cap G$, then

$$\max_{x \in F} [xy_0 - f(x)] = x_0 y_0 - f(x_0) \tag{4.2}$$

and

$$\min_{x \in G} [xy_0 - g(x)] = x_0 y_0 - g(x_0). \tag{4.3}$$

The Fenchel duality theorem will be widely used in the thesis to obtain dual (equivalent) forms of the lower and upper bounds for the JSCC reliability function. As will become more clear in the next chapter, Fenchel duality plays an important role in studying the JSCC reliability function. First, it facilitates the computation of the JSCC reliability function; second, it is a tool to evaluate the tightness of the bounds and to establish the lower/upper bound for the JSCC reliability function.

We remark that all the concepts and results can be easily extended to vector functions (on subsets of $\mathbb{R}^k$).

4.2 Applications: Source and Channel Reliability Functions Revisited

In this section we revisit the source and channel reliability functions and theirs bounds presented in Chapter 2. The conjugacy between the source and channel reliability functions and the corresponding source and channel functions are studied.

Lemma 4.1 *The source function $E_s(\rho, Q_S)$ defined by (2.7) is a strictly convex function of ρ if Q_S is not a uniform distribution; otherwise $E_s(\rho, Q_S)$ is a linear function of ρ. Thus, the DMS error exponent $e(R, Q_S)$ given in (2.4) and $E_s(\rho, Q_S)$ are a pair of convex Fenchel transforms, i.e.,*

$$e(R, Q_S) = (E_s(\rho, Q_S))^*, \quad R \in [H_{Q_S}(S), \log_2 |\mathcal{S}|]$$

and

$$E_s(\rho, Q_S) = (e(R, Q_S))^*, \quad \rho \in [0, +\infty).$$

Proof: Since $E_s(\rho, Q_S)$ is differentiable in ρ, it can be easily verified that the second derivative is nonnegative, and strictly positive if Q_S is not the uniform distribution (also see (5.3) and Lemma 5.1). It then follows from the parametric form of the source error exponent (2.6) that $e(R, Q_S)$ and $E_s(\rho, Q_S)$ are a pair of Fenchel transforms. ■

The relation between Gallager's channel function $E_o(\rho, W_{Y|X})$ and the random-coding and sphere-packing bounds is more complicated. First of all, recall that for each $P_X \in \mathcal{P}(\mathcal{S})$, $E_r(R, P_X, W_{Y|X})$ as defined in (2.15) is a convex non-increasing function for all R, and is a linear function of R with slope -1 for $R \le R_{cr}(P_X, W_{Y|X})$ [42, p. 143]. It will be convenient to regard this linear function as defining $E_r(R, P_X, W_{Y|X})$ for all negative R. The random-coding bound $E_r(R, W_{Y|X})$, which is the maximum of this family of convex functions, is a convex strictly decreasing function of R for $R < C(W_{Y|X})$, and is a linear function of R with slope -1 for all R below the critical rate $R_{cr}(W_{Y|X})$. For $R \geq C(W_{Y|X})$, $E_r(R, W_{Y|X}) = 0$. Since $E_r(R, W_{Y|X})$ is convex, then $-E_r(R, W_{Y|X})$ is concave. Let $T_r(\rho, W_{Y|X})$ be the concave transform of $-E_r(R, W_{Y|X})$, i.e.,

$$T_r(\rho, W_{Y|X}) \triangleq \inf_{R \in \mathbb{R}} [\rho R + E_r(R, W_{Y|X})]. \tag{4.4}$$

It follows from the properties of $E_r(R, W_{Y|X})$ noted above that $T_r(\rho, W_{Y|X}) = -\infty$ for $\rho < 0$ and $\rho > 1$ and that $T_r(\rho, W_{Y|X})$ is finite for $\rho \in [0, 1]$.

Lemma 4.2 *The function $T_r(\rho, W_{Y|X})$ defined by (4.4) is the concave hull on the interval $[0, 1]$ of the channel function $E_o(\rho, W_{Y|X})$ defined in (2.19). Thus, $E_o(\rho, W_{Y|X}) \leq T_r(\rho, W_{Y|X})$ for $0 \leq \rho \leq 1$.*

Proof: We form the concave transform of $E_o(R, W_{Y|X})$ on the interval $[0, 1]$ to get

$$\big(E_o(\rho, W_{Y|X})\big)_* = \inf_{0 \leq \rho \leq 1} [\rho R - E_o(\rho, W_{Y|X})] = - \sup_{0 \leq \rho \leq 1} [E_o(\rho, W_{Y|X}) - \rho R].$$

Now use, in succession, (2.19), (2.16), and (2.18) to get

$$\begin{aligned}
\big(E_o(\rho, W_{Y|X})\big)_* &= - \sup_{0 \leq \rho \leq 1} \max_{P_X} [E_o(\rho, P_X, W_{Y|X}) - \rho R] \\
&= - \max_{P_X} \sup_{0 \leq \rho \leq 1} [E_o(\rho, P_X, W_{Y|X}) - \rho R] \\
&= - \max_{P_X} E_r(R, P_X, W_{Y|X}) \\
&= -E_r(R, W_{Y|X}).
\end{aligned}$$

Since $T_r(\rho, W_{Y|X})$ is the concave transform of the concave function, $-E_r(R, W_{Y|X})$, we have that

$$\big(-E_r(R, W_{Y|X})\big)_* = T_r(\rho, W_{Y|X}) \quad \text{and so} \quad \big(E_o(\rho, W_{Y|X})\big)_{**} = T_r(\rho, W_{Y|X}).$$

Hence, $T_r(\rho, W_{Y|X})$ is the concave hull on $[0, 1]$ of $E_o(\rho, R)$. ■

Similarly to the above, recall that $E_{sp}(R, W_{Y|X})$, defined in (2.26) is convex, zero for $R \geq C(W_{Y|X})$, positive for $R < C(W_{Y|X})$, and finite if $R > R_\infty(W_{Y|X})$ [32, 42], where $R_\infty(W_{Y|X})$ is given by

$$R_\infty(W_{Y|X}) \triangleq \lim_{\rho \to \infty} \frac{E_o(\rho, W_{Y|X})}{\rho}. \tag{4.5}$$

A computable expression for $R_\infty(W_{Y|X})$ is given in [42, p. 158]. The normal situation is $R_\infty(W_{Y|X}) = 0$. (As shown by Gallager, $R_\infty(W_{Y|X}) = 0$ unless each channel output symbol is unreachable from at least one input. In the latter case, $R_\infty(W_{Y|X}) > 0$.) We now let $T_{sp}(\rho, W_{Y|X})$ be the concave transform of the concave function $-E_{sp}(R, W_{Y|X})$, i.e.,

$$T_{sp}(\rho, W_{Y|X}) \triangleq \inf_{R_\infty(W_{Y|X})<R<\infty} [\rho R + E_{sp}(R, W_{Y|X})]. \tag{4.6}$$

It follows that $T_{sp}(\rho, W_{Y|X}) = -\infty$ for $\rho < 0$ and that $0 \le T_{sp}(\rho, W_{Y|X}) < \infty$ for $\rho \ge 0$.

Lemma 4.3 *The function $T_{sp}(\rho, W_{Y|X})$ defined by (4.6) is the concave hull on $[0, \infty)$ of the channel function $E_o(\rho, W_{Y|X})$ defined in (2.19).*

Proof: We now form the concave transform of $E_o(\rho, W_{Y|X})$ on the interval $[0, \infty)$ to get

$$\big(E_o(\rho, W_{Y|X})\big)_* = \inf_{0\le\rho<\infty} [\rho R - E_o(\rho, W_{Y|X})] = -\sup_{0\le\rho<\infty} [E_o(\rho, W_{Y|X}) - \rho R].$$

Now use (2.19), (2.28), and (2.29) to get

$$\begin{aligned}
\big(E_o(\rho, W_{Y|X})\big)_* &= -\sup_{0\le\rho<\infty} \max_{P_X} [\tilde{E}_0(\rho, P_X, W_{Y|X}) - \rho R] \\
&= -\max_{P_X} \sup_{0\le\rho<\infty} [\tilde{E}_0(\rho, P_X, W_{Y|X}) - \rho R] \\
&= -\max_{P_X} \tilde{E}_{sp}(R, P_X, W_{Y|X}) \\
&= -E_{sp}(R, W_{Y|X}).
\end{aligned}$$

As in the previous proof, $\big(E_o(\rho, W_{Y|X})\big)_{**} = T_{sp}(\rho, W_{Y|X})$. Hence, $T_{sp}(\rho, W_{Y|X})$ is the concave hull on $[0, \infty)$ of $E_o(\rho, R)$. ■

Observation 4.1 Note that the function $E_o(\rho, P_X, W_{Y|X})$ is concave in ρ for each P_X [42, p. 142]. Hence, if the maximizing P_X in (2.19) is *independent* of ρ, $E_o(\rho, W_{Y|X})$ is concave and thus $T_r(\rho, W_{Y|X})$ and $T_{sp}(\rho, W_{Y|X})$ are equal to $E_o(\rho, W_{Y|X})$. This situation holds if the channel is symmetric in the sense of Gallager [42, p. 94] (also see Example 5.3.3). For this case, the maximizing distribution is the uniform distribution $P_X(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$. However, there are channels for which $E_o(\rho, W_{Y|X})$ is not concave. One example of such a channel is provided by Gallager [42, Fig. 5.6.5]. For this particular (6-ary input, 4-ary output) channel, we plot $E_o(\rho, W_{Y|X})$ against ρ in Fig. 4.1. It is noted that the derivative of $E_o(\rho, W_{Y|X})$ has a positive jump increase at around $\rho = 0.51$ (see [42, Fig. 5.6.5]), and its concave hull $T_r(\rho, W_{Y|X})$ is strictly larger than $E_o(\rho, W_{Y|X})$ in the interval $\rho \in (0.41, 0.62)$.

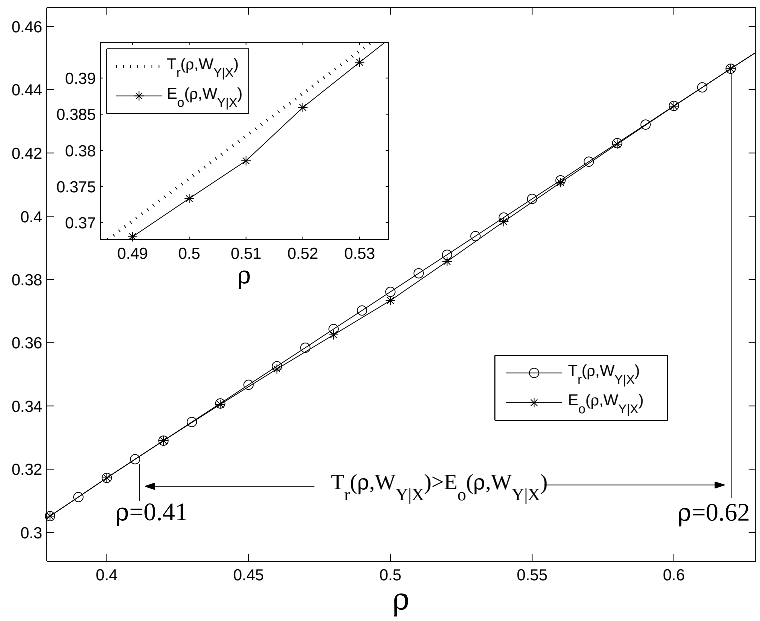

Figure 4.1: Example of a 6-ary input, 4-ary output DMC (see [42, Fig. 5.6.5]) for which $E_o(\rho, W_{Y|X})$ is not concave.

We next prove another two pairs of Fenchel transforms for the memoryless Gaussian system, which will be used in Chapter 8.

Given $\rho \geq 0$, for the continuous memoryless source Q_S, $\mathcal{S} = \mathbb{R}$, define source function

$$E(Q_S, \Delta, \rho) \triangleq \sup_{P_S} [\rho R(P_S, \Delta) - D(P_S \parallel Q_S)], \tag{4.7}$$

where the supremum is taken over all the probability distributions P_S's defined on $\mathcal{S}$ such that $R(P_S, \Delta)$ and $D(P_S \parallel Q_S)$ are well-defined and finite. We remark that (4.7) is equal to the guessing exponent for MGS's [6] under the squared-error distortion measure and admits an explicit form

$$E(Q_S, \Delta, \rho) = \max\left\{0, \frac{1}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]\right\}. \tag{4.8}$$

Lemma 4.4 *$E(Q_S, \Delta, \rho)$ and the MGS exponent $F_G(R, Q_S, \Delta)$ defined by (2.46) and (2.47) are a pair of convex Fenchel transforms $\rho \geq 0$ and $R \geq 0$, i.e.,*

$$E(Q_S, \Delta, \rho) = F_G(R, Q_S, \Delta)^* \qquad \textit{for all } \rho \geq 0$$

and

$$F_G(R, Q_S, \Delta) = E(Q_S, \Delta, \rho)^* \qquad \textit{for all } R \geq 0.$$

Proof: By definition,

$$F_G(R, Q_S, \Delta)^* = \sup_{R \geq 0}\left[\rho R - F(R, Q_S, \Delta)\right] = \sup_{R \geq R(Q_S, \Delta)} f(R)$$

where

$$f(R) = \rho R - \frac{1}{2}\left(\frac{\Delta e^{2R}}{\sigma_S^2} - \ln \frac{\Delta e^{2R}}{\sigma_S^2} - 1\right).$$

Since

$$\frac{\partial f(R)}{\partial R} = 1 + \rho - \frac{\Delta e^{2R}}{\sigma_S^2},$$

it is seen that $f(R)$ is concave and

$$\sup_{R \geq R(Q_S, \Delta)} f(R) = f\left(\frac{1}{2}\ln \frac{\sigma_S^2(1+\rho)}{\Delta}\right) = \frac{1}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right] > 0$$

if $\frac{\Delta}{\sigma_S^2} \leq 1 + \rho$, and $f(R)$ is concave decreasing with

$$\sup_{R \geq R(Q_S, \Delta)} f(R) = \sup_{R \geq 0} f(R) = f(0) = 0 > \frac{1}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]$$

if $\frac{\Delta}{\sigma_S^2} > 1+\rho$, which implies that $E(Q_S, \Delta, \rho)$ is the convex Fenchel transform of $F_G(R, Q_S, \Delta)$, i.e.,

$$F_G(R, Q_S, \Delta)^* = E(Q_S, \Delta, \rho) = \max\left\{0, \frac{1}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]\right\}.$$

Finally, $F_G(R, Q_S, \Delta)$ is also the convex Fenchel transform of $E(Q_S, \Delta, \rho)$ since $F_G(R, Q_S, \Delta)$ is convex. ■

Lemma 4.5 *The negative MGC sphere-packing exponent* $-E_{sp}(R, W_{Y|X}, \mathcal{E})$ *given in (2.51) and the Gaussian input Gaussian-input channel function* $\widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)$ *given in (2.54) are a pair of concave Fenchel transforms for* $\rho \geq 0$ *and* $R > 0$*, i.e.,*

$$-E_{sp}(R, W_{Y|X}, \mathcal{E}) = \widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho)_* \qquad \textit{for all } R > 0$$

and

$$\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho) = (-E_{sp}(R, W_{Y|X}, \mathcal{E}))_* \qquad \textit{for all } \rho \geq 0.$$

Proof: Note that

$$E_{sp}(R, W, \mathcal{E}) = \max_{\rho \geq 0}[-\rho R + \widetilde{E}_0(W, \mathcal{E}, \rho)] = -\inf_{\rho \geq 0}[\rho R - \widetilde{E}_0(W, \mathcal{E}, \rho)],$$

which implies that $-E_{sp}(R, W, \mathcal{E})$ is the concave transform of $\widetilde{E}_0(W, \mathcal{E}, \rho)$ on

$$\{R : -E_{sp}(R, W, \mathcal{E}) > -\infty\} = \mathbb{R}^+.$$

Thus, the transform

$$(-E_{sp}(R, W, \mathcal{E}))_* = \inf_{R \in \mathbb{R}^+} [\rho R + E_{sp}(R, W, \mathcal{E})]$$

is the concave hull of $\widetilde{E}_0(W, \mathcal{E}, \rho)$ in $\rho \in [0, \infty)$. We next show $(-E_{sp}(R, W, \mathcal{E}))_* = \widetilde{E}_0(W, \mathcal{E}, \rho)$ by definition. Now if we set

$$\frac{\partial}{\partial R}[\rho R + E_{sp}(R, W, \mathcal{E})] = 0,$$

we have (refer to Lemma 2.1)

$$\sqrt{1 + \frac{4\beta}{\mathrm{SNR}(\beta - 1)}} = \frac{2\beta}{\mathrm{SNR}}(1 + \rho) - 1, \tag{4.9}$$

where $\beta = e^{2R}$. Substituting (4.9) back into $(-E_{sp}(R, W, \mathcal{E}))_*$ and using (2.51) yield

$$(-E_{sp}(R, W, \mathcal{E}))_* = \frac{1}{2}\left[\rho \ln \beta^* + (1 - \beta^*)(1 + \rho) + \mathrm{SNR} + \ln\left(\beta^* - \frac{\mathrm{SNR}}{1 + \rho}\right)\right], \tag{4.10}$$

where β^* is determined by (4.9), which can be equivalently written by

$$-(1 + \rho) + \frac{1 + \rho}{\beta(1 + \rho) - \mathrm{SNR}} + \frac{\rho}{\beta} = 0, \tag{4.11}$$

subject to $\beta > \text{SNR}/(1+\rho)$ according to (4.10). In this range the left-hand side of (4.11) is decreasing in β and ranges from $+\infty$ to the negative number $-(1+\rho)$, which means there is a unique β^* satisfying (4.11). Solving the function (4.11) for the stationary point β^* we obtain

$$\beta^* = \frac{1}{2}\left(1+\frac{\text{SNR}}{1+\rho}\right)\left[1+\sqrt{1-\frac{4\text{SNR}\rho}{(1+\rho+\text{SNR})^2}}\right]. \tag{4.12}$$

On the other hand, we can replace

$$\widehat{\beta} = 1 - 2r\mathcal{E} + \frac{\text{SNR}}{1+\rho}$$

in the expression of $\widetilde{E}_o(W,\mathcal{E},\rho)$ given by (2.54) and obtain

$$\widetilde{E}_o(W,\mathcal{E},\rho) = \max_{\frac{\text{SNR}}{1+\rho}<\widehat{\beta}<1+\frac{\text{SNR}}{1+\rho}} \frac{1}{2}\left[\rho\ln\widehat{\beta} + (1-\widehat{\beta})(1+\rho) + \text{SNR} + \ln\left(\widehat{\beta} - \frac{\text{SNR}}{1+\rho}\right)\right].$$

Maximizing the above over $\widehat{\beta}$ (see [42, p. 339] for details), we see that $\widetilde{E}_o(W,\mathcal{E},\rho)$ has the same parametric form as (4.10), which implies

$$(-E_{sp}(R,W,\mathcal{E}))_* = \widetilde{E}_o(W,\mathcal{E},\rho),$$

and hence $\widetilde{E}_o(W,\mathcal{E},\rho)$ is the concave transform of $-E_{sp}(R,W,\mathcal{E})$. ■

Lemma 4.6 *The negative Gaussian input random-coding exponent $-E_\dagger(R,W_{Y|X},\mathcal{E})$ given in (2.56) and $\widetilde{E}_o(W_{Y|X},\mathcal{E},\rho)$ are a pair of concave Fenchel transforms for $0\le\rho\le 1$ and $R\ge 0$, i.e.,*

$$-E_\dagger(R,W_{Y|X},\mathcal{E}) = \widetilde{E}_0(W_{Y|X},\mathcal{E},\rho)_* \qquad \textit{for all } R>0$$

and

$$\widetilde{E}_0(W_{Y|X},\mathcal{E},\rho) = (-E_\dagger(R,W_{Y|X},\mathcal{E}))_* \qquad \textit{for all } 0\le\rho\le 1.$$

Proof: Recall that by Gallager [42, Chapter 7]

$$E_\dagger(R,W,\mathcal{E}) = \max_{0\le\rho\le 1}[-\rho R + \widetilde{E}_0(W,\mathcal{E},\rho)] = -\inf_{0\le\rho\le 1}[\rho R - \widetilde{E}_0(W,\mathcal{E},\rho)],$$

which means that $-E_\dagger(R,W,\mathcal{E})$ is the concave transform of $\widetilde{E}_0(W,\mathcal{E},\rho)$ on

$$\{R : -E_\dagger(R,W,\mathcal{E}) > -\infty\} = \mathbb{R}^+.$$

Thus, the transform

$$(-E_{\dagger}(R,W,\mathcal{E}))_* = \inf_{R\in\mathbb{R}^+}[\rho R + E_{\dagger}(R,W,\mathcal{E})]$$

is the concave hull of $\widetilde{E}_0(W,\mathcal{E},\rho)$ in $\rho \in [0,1]$. Lemma 4.5 implies that $\widetilde{E}_0(W,\mathcal{E},\rho)$ is concave in $[0,\infty)$. Thus we have $(-E_{\dagger}(R,W,\mathcal{E}))_* = \widetilde{E}_0(W,\mathcal{E},\rho)$ for all $\rho \in [0,1]$. ■

We close this section by summarizing these Fenchel transform pairs in the following table.

System	Conjugacy	Source/Channel Exponents	Source/Channel Functions
DMS	Convex	$e(R,Q_S)$	$E_s(\rho,Q_S)$
MGS	Convex	$F_G(R,Q_S,\Delta)$	$E(Q_S,\Delta,\rho)$
DMC	Concave	$-E_r(R,W_{Y\|X})$	$T_r(\rho,W_{Y\|X})$
DMC	Concave	$-E_{sp}(R,W_{Y\|X})$	$T_{sp}(\rho,W_{Y\|X})$
MGC	Concave	$-E_{\dagger}(R,W,\mathcal{E})$	$\widetilde{E}_0(W_{Y\|X},\mathcal{E},\rho)$
MGC	Concave	$-E_{sp}(R,W_{Y\|X})$	$\widetilde{E}_0(W_{Y\|X},\mathcal{E},\rho)$

Table 4.1: Useful Fenchel transform pairs.

4.3 Concluding Remarks

In this chapter, we introduced one-dimensional conjugate convex and concave functions and the Fenchel duality theorem. Note that similar results can be easily carried out for vector functions; readers may refer to [17], [65], [79]. Consequently, we applied these properties of conjugacy to the source and channel reliability functions introduced in Chapter 2. The Fenchel transforms of these functions are summarized in Table 4.1.

Chapter 5

JSCC Reliability Function for Discrete Memoryless Systems

In this chapter, we study the JSCC reliability function for discrete memoryless systems. In particular, we investigate the computation of Csiszár's bounds for the JSCC error exponent, E_J, of a communication system consisting of a DMS Q_S and a DMC $W_{Y|X}$.

We first formally describe the system and define the JSCC error exponent in Section 5.1. In Section 5.2, we investigate the analytical computation of Csiszár's random-coding lower bound and sphere-packing upper bound for the JSCC error exponent. By applying the Fenchel duality theorem introduced in the last chapter, we provide equivalent expressions for these bounds which involve a maximization over a non-negative parameter of the difference between the concave hull of Gallager's channel function and Gallager's source function; hence, they can be readily computed for arbitrary source-channel pairs by applying Arimoto's algorithm [9]. When the channel's distribution is symmetric, our bounds admit closed-form parametric expressions.

In Section 5.3, we provide formulas of the rates for which the bounds are attained and establish explicit computable sufficient and necessary conditions in terms of Q_S and $W_{Y|X}$ under which the upper and lower bounds coincide; in this case, E_J can be determined exactly. A byproduct of our results is the observation that Csiszár's JSCC random-coding lower bound can be larger than Gallager's earlier lower bound obtained in [42]. Using a similar approach, we obtain in Section 5.4 the equivalent expression of Csiszár's expurgated lower bound [31] and establish the condition when the random-coding lower bound can be improved by the expurgated bound. As an example, we give closed-form parametric expressions of the improved lower bound and the corresponding condition for DMSs and equidistant DMCs.

In Section 5.5, we partially address the computation of Csiszár's lower and upper bounds for the (lossy) JSCC excess distortion exponent with distortion threshold Δ, E_J^{Δ}. Under the case of the Hamming distortion measure, and for a binary DMS and an arbitrary DMC, we express the bounds for E_J^{Δ} and the rates for which the bounds are attained as in the lossless case. Finally, we state our conclusions in Section 5.6.

5.1 Definitions and System Description

5.1.1 JSCC System and JSCC Error Exponent

We consider throughout this chapter a communication system consisting of a DMS Q_S with finite alphabet $\mathcal{S}$ and distribution Q_S, and a DMC $W_{Y|X}$ with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition probability $W_{Y|X}$. Without loss of generality we assume that $Q_S(s) > 0$ for each $s \in \mathcal{S}$. Also, if the source distribution is uniform, optimal (lossless) JSCC amounts to optimal channel coding which is already well-studied. Therefore, we assume throughout this chapter that Q_S is not the uniform distribution on $\mathcal{S}$ except in Section 5.5 where we deal with JSCC under a fidelity criterion.

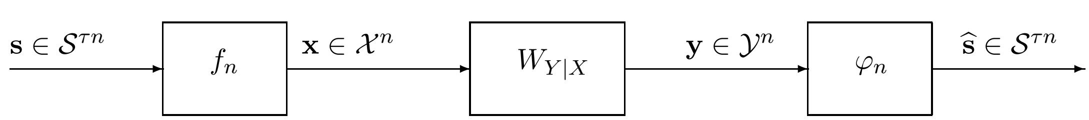

Figure 5.1: JSCC system consisting of a DMS and a DMC.

A joint source-channel (JSC) code with blocklength n and transmission rate $\tau > 0$

(measured in source symbols/channel use) is a pair of mappings

$$f_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n}.$$

That is, blocks $\mathbf{s} \triangleq (s_1, s_2, ..., s_{\tau n})$ of source symbols of length τn are encoded as blocks $\mathbf{x} \triangleq (x_1, x_2, ..., x_n) = f_n(\mathbf{s})$ of symbols from $\mathcal{X}$ of length n, transmitted, received as blocks $\mathbf{y} \triangleq (y_1, y_2, ..., y_n)$ of symbols from $\mathcal{Y}$ of length n and decoded as blocks of source symbols $\varphi_n(\mathbf{y})$ of length τn; see Fig. 5.1. The probability of erroneously decoding the block is

$$P_e^{(n)}(Q_S, W_{Y|X}, \tau) \triangleq \sum_{\{(\mathbf{s},\mathbf{y}):\varphi_n(\mathbf{y})\neq \mathbf{s}\}} Q_S^{(\tau n)}(\mathbf{s}) W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))\,. \tag{5.1}$$

Definition 5.1 The JSCC error exponent $E_J(Q_S, W_{Y|X}, \tau)$ is defined as the supremum of the set of all numbers E for which there exists a sequence of JSC codes (f_n, φ_n) with transmission rate τ and blocklength n such that

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_e^{(n)}(Q_S, W_{Y|X}, \tau).$$

When there is no possibility of confusion, $E_J(Q_S, W_{Y|X}, \tau)$ will be written as E_J. We know from the JSCC theorem (e.g., [29, p. 218], [42]) that E_J can be positive if and only if $\tau H_{Q_S}(S) < C(W_{Y|X})$.

5.1.2 Tilted Distributions

We associate with the source distribution Q_S a family of tilted distributions $Q_S^{(\rho)}$ defined by

$$Q_S^{(\rho)}(s) \triangleq \frac{Q_S^{\frac{1}{1+\rho}}(s)}{\sum_{s'\in\mathcal{S}} Q_S^{\frac{1}{1+\rho}}(s')}, \qquad s \in \mathcal{S}, \qquad \rho \geq 0. \tag{5.2}$$

Lemma 5.1 [32, p. 44] *The entropy $H_{Q_S^{(\rho)}}(S)$ is a strictly increasing function of ρ except in the case that $Q_S(s) = 1/|\mathcal{S}|$ for all $s \in \mathcal{S}$. Moreover, for $H_{Q_S}(S) \leq R \leq \log_2|\mathcal{S}|$, the equation $H_{Q_S^{(\rho)}}(S) = R$ is satisfied by a unique value ρ^* (where we define $\rho^* \triangleq \infty$ if $R = \log_2|\mathcal{S}|$ and define $H_{Q_S^{(\infty)}}(S) \triangleq \log_2|\mathcal{S}|$).*

The proof that $H_{Q_S^{(\rho)}}(S)$ is increasing follows easily from differentiation with respect to ρ and a use of the Cauchy-Schwarz inequality. The remainder of the proof follows from the facts that $H_{Q_S^{(0)}}(S) = H_{Q_S}(S)$, $\lim_{\rho\to\infty} H_{Q_S^{(\rho)}}(S) = \log_2 |\mathcal{S}|$ and that $H_{Q_S^{(\rho)}}(S)$ is a continuous function of ρ.

It is easily seen that

$$H_{Q_S^{(\rho)}}(S) = \frac{\partial E_s(\rho, Q_S)}{\partial \rho}, \tag{5.3}$$

where $E_s(\rho, Q)$ is defined by (2.7). From this we see that for $R \geq H_{Q_S}(S)$ the supremum in (2.6) is achieved at ρ^*.

5.2 Csiszár's Random-Coding and Sphere-Packing Bounds

In [30], Csiszár establishes a lower and an upper bound for the JSCC error exponent E_J in terms of the source and channel error exponents. Given a DMS Q_S, a DMC $W_{Y|X}$, and the transmission rate τ (source symbol/channel use), he proved that the JSCC error exponent in definition 5.1 satisfies

$$\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) \leq E_J(Q_S, W_{Y|X}, \tau) \leq \inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X}) \right], \tag{5.4}$$

where $e(R, Q_S)$ is the source error exponent, $E(R, W_{Y|X})$ is the channel error exponent, and

$$\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) \triangleq \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X}) \right] \tag{5.5}$$

is called Csiszár's source-channel random-coding lower bound since it contains $E_r(R, W_{Y|X})$ in its expression. It is shown that if the zero-error capacity[1] of the channel $W_{Y|X}$ is positive, i.e., if $E(R, W_{Y|X}) = \infty$ for some R positive, then the upper bound given by (5.4) is tight. The proof of the upper bound in (5.4) follows from a simple type counting argument. We will recast the proof in Observation 6.1. The lower bound is proved by employing a generalized minimum mutual information decoding rule. We will recast the proof for $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ in Observation 6.2.

[1]The zero-error capacity $C_0(W_{Y|X})$ is defined by $C_0(W_{Y|X}) \triangleq \sup\{R : E(R, W_{Y|X}) = \infty\}$.

Note that the channel error exponent is only partially known, and hence the upper bound in (5.4), although looking elegant, is uncomputable in general. We then further upper bound it by replacing the channel error exponent $E(R, W_{Y|X})$ by the sphere-packing upper bound, i.e.,

$$E_J(Q_S, W_{Y|X}, \tau) \leq \inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X}) \right] \leq \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$$

where

$$\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) \triangleq \inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right] \tag{5.6}$$

is called Csiszár's source-channel sphere-packing upper bound. In the following we address the computation of the lower and upper bounds $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$.

Applying Fenchel duality theorem, these bounds $\underline{E}_{Jr}$ and $\overline{E}_{Jsp}$ can be expressed in a form more adapted to calculation as follows.

Theorem 5.1 *Let $\tau H_{Q_S}(S) < C(W_{Y|X})$ and let $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. Then*

$$\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) = \max_{0 \leq \rho \leq 1} [T_r(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)] \tag{5.7}$$

and

$$\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) = \max_{0 \leq \rho < \infty} [T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)] \tag{5.8}$$

where $T_r(\rho, W_{Y|X})$ and $T_{sp}(\rho, W_{Y|X})$ are the concave hulls of $E_o(\rho, W_{Y|X})$ on $[0,1]$ and $[0,\infty)$ defined in (4.4) and (4.6), respectively, and R_∞ is defined in (4.5). If the maximizing P_X in (2.19) is independent of ρ, $T_r(\rho, W_{Y|X})$ and $T_{sp}(\rho, W_{Y|X})$ can be replaced by $E_o(\rho, W_{Y|X})$.

Remark 5.1 When $\tau H_{Q_S}(S) \geq C(W_{Y|X})$, $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) = \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) = 0$.

Observation 5.1 According to Lemma 4.2, $E_o(\rho, W_{Y|X}) \leq T_r(\rho, W_{Y|X})$. Thus the lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ can be replaced by the *possibly looser* lower bound

$$\max_{0 \leq \rho \leq 1} [E_o(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \tag{5.9}$$

This is the lower bound implied by Gallager's work [42, p. 535]. As noted earlier, if the maximizing P_X in (2.19) is independent of ρ (e.g., for symmetric channels, see Section 5.3.3), the two lower bounds are identical.

Proof of Theorem 5.1: We first apply Fenchel's Duality Theorem (4.1) to the lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, t)$. From Lemma 4.1, (2.6), and (2.5), $\tau e(R/\tau, Q_S)$ is convex on $(-\infty, \tau \log |\mathcal{S}|]$ and has convex transform $\tau E_s(\rho, Q_S)$ on the set $[0, \infty)$. Also, from the discussion preceding Lemma 4.2, $-E_r(R, W_{Y|X})$ is concave on $\mathbb{R}$ and has concave transform $T_r(\rho, W_{Y|X})$ which is bounded on $[0, 1]$. Thus, by Fenchel's Duality Theorem,

$$\inf_{-\infty \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X}) \right] = \max_{0 \leq \rho \leq 1} [T_r(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \quad (5.10)$$

Now the convex function $\tau e(R/\tau, Q_S) + E_r(R, W_{Y|X})$ is non-increasing for $R \leq \tau H_{Q_S}(S)$ since $\tau e(R/\tau, Q_S) = 0$ in this region. This implies that the infimum on the left side of (5.10) can be restricted to the interval $\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|$. Since this is now the infimum of a continuous function on a finite interval this will be a minimum. Hence, (5.7) is an equivalent representation of $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$.

Similarly, for the upper bound, recall from the discussion preceding Lemma 4.3 that $-E_{sp}(R, W_{Y|X})$ is concave and finite for $R > R_\infty(W_{Y|X})$ and has a concave transform $T_{sp}(\rho, W_{Y|X})$, which is finite on $0 \leq \rho < \infty$. Thus, by Fenchel's Duality Theorem,

$$\inf_{R_\infty(W_{Y|X}) < R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right] = \max_{0 \leq \rho < \infty} [T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \quad (5.11)$$

The assumption $R_\infty(W_{Y|X}) < \tau \log_2 |\mathcal{S}|$ ensures that the infimum on the left of (5.11) is taken over a set with interior points. If $R_\infty(W_{Y|X}) < \tau H_{Q_S}(S)$, the infimum can be replaced by a minimum on the interval $\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|$ by the same argument as for the lower bound. If $R_\infty(W_{Y|X}) \geq \tau H_{Q_S}(S)$, we no longer form the infimum of a continuous function, but it can still be shown that there is a minimum point which lies in the interval $\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|$. Hence, (5.11) is an equivalent representation of $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$. ■

Observation 5.2 The parametric form of the lower and upper bounds (5.7) and (5.8) indeed facilitates the computation of Csiszár's bounds. In order to compute the bounds for general non-symmetric channels (when $\tau H_{Q_S}(S) < C(W_{Y|X})$ and $\tau \log_2 |\mathcal{S}| > R_\infty$), one could employ Arimoto's algorithm [9] to find the maximizing distribution and thus $E_o(\rho, W_{Y|X})$. We then can immediately obtain the concave hulls $T_r(\rho, W_{Y|X})$ and $T_{sp}(\rho, W_{Y|X})$ numerically (e.g., using Matlab) and thus the maxima of $T_r(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$ and $T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$. This significantly reduces the computation complexity since to compute (5.5) and (5.6), we need to first compute $E_r(R, W_{Y|X})$ and $E_{sp}(R, W_{Y|X})$ for *each* R, which requires almost the same complexity as above, and then we need to find the minima by searching over *all* R's. For symmetric channels, (5.7) and (5.8) are analytically solved; see Section 5.3.3.

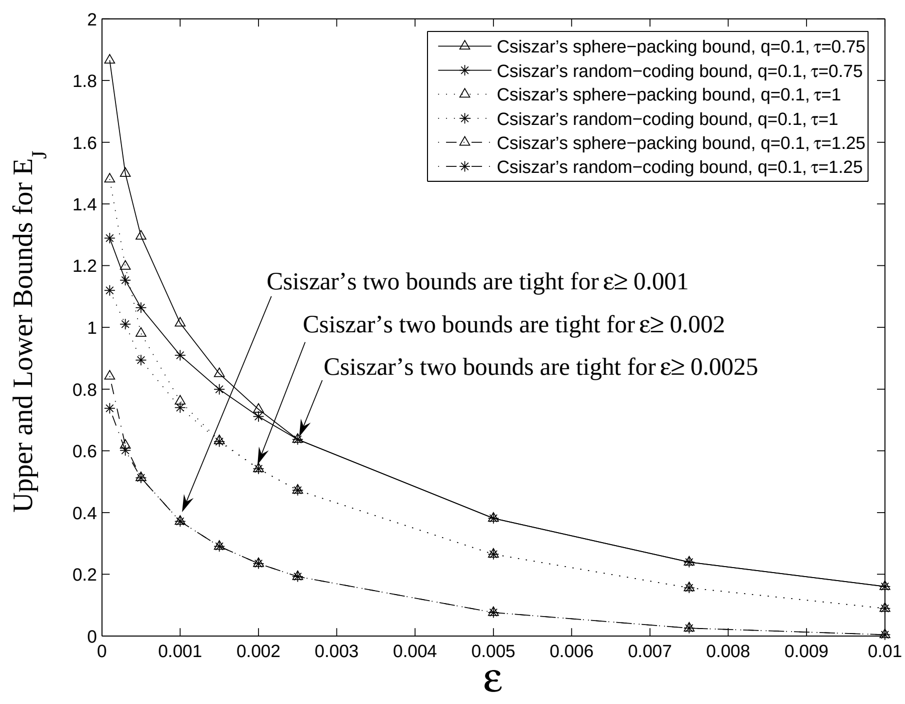

Figure 5.2: Csiszár's random-coding and sphere-packing bounds for the system of Example 5.1.

Example 5.1 Consider a communication system with a binary DMS with distribution $Q_S = \{q, 1-q\}$ and a DMC with $|\mathcal{X}| = 6$, $|\mathcal{Y}| = 4$, and transition probability matrix

$$W_{Y|X} = \begin{bmatrix} 1-18\varepsilon & 6\varepsilon & 6\varepsilon & 6\varepsilon \\ 6\varepsilon & 1-18\varepsilon & 6\varepsilon & 6\varepsilon \\ 6\varepsilon & 6\varepsilon & 1-18\varepsilon & 6\varepsilon \\ 6\varepsilon & 6\varepsilon & 6\varepsilon & 1-18\varepsilon \\ 0.5-\varepsilon & 0.5-\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0.5-\varepsilon & 0.5-\varepsilon \end{bmatrix}, \qquad 0 \le \varepsilon \le \frac{1}{18}.$$

We then compute Csiszár's random-coding and sphere-packing bounds, $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$. For fixed Q_S and transmission rate τ, we plot these bounds in terms of ε in Fig. 5.2. Our numerical results show that E_J could be determined exactly for a large class of $(q,\ \varepsilon,\ \tau)$ triplets: when source $Q_S = \{0.1, 0.9\}$ and rate $\tau = 0.75$, E_J is exactly known for $\varepsilon \ge 0.0025$; when $Q_S = \{0.1, 0.9\}$ and $\tau = 1$, E_J is known for $\varepsilon \ge 0.002$; and when $Q_S = \{0.2, 0.8\}$ and $\tau = 1.25$, E_J is known for $\varepsilon \ge 0.001$. Since for this channel $E_o(\rho, W_{Y|X})$ might not be concave (e.g., when $\varepsilon = 0.01$, $W_{Y|X}$ reduces to the DMC discussed in Observation 4.1 at the end of Section 4.2), our results indicate that Csiszár's lower bound is slightly but strictly larger (by ≈ 0.0001) than Gallager's lower bound (5.9) for $q = 0.1$, $\tau = 1$, and ε around 0.02. This is illustrated in Fig. 5.3.

5.3 Tightness of the Upper and Lower Bounds

One important objective in investigating the bounds for the JSCC error exponent E_J is to ascertain when the bounds $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ are tight so that the exact value of E_J is obtained. According to Csiszár's result (5.4), we note that if the minimum in the expressions of $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ or $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ is attained for a rate (strictly) larger than the critical rate $R_{cr}(W_{Y|X})$, then the two bounds coincide and thus E_J is determined exactly. This raises the following question: how can we check whether the minimum in $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ or $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ is attained for a rate

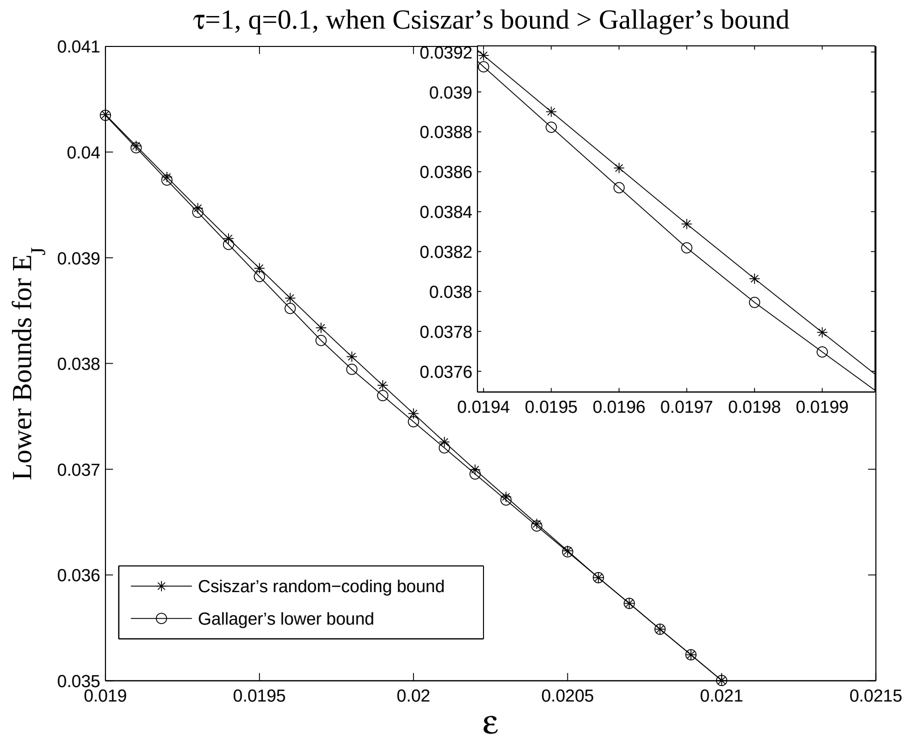

Figure 5.3: Csiszár's random-coding bound vs Gallager's lower bound for the system of Example 5.1.

larger than $R_{cr}(W_{Y|X})$? One may indeed wonder if there exist explicit conditions for which $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) = \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$. The answer is affirmative; furthermore, we can verify whether the two bounds are tight in two ways: one is to compare $\tau H_{Q_S^{(1)}}(S)$ with $R_{cr}(W_{Y|X})$, and the other is to compare the minimizer of $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ in (5.8), $\overline{\rho}^*$ say, with 1.

5.3.1 A Sufficient and Necessary Condition

Before we present these conditions, we first define the following quantities which achieve the bounds $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ under the assumptions $\tau H_{Q_S}(S) < C(W_{Y|X})$ and $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$:

$$\underline{R}_m \triangleq \arg \min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X}) \right], \tag{5.12}$$

$$\overline{R}_m \triangleq \arg \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right], \tag{5.13}$$

$$\underline{\rho}^* \triangleq \arg \max_{0 \leq \rho \leq 1} [T_r(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)], \tag{5.14}$$

$$\overline{\rho}^* \triangleq \arg \max_{0 \leq \rho < \infty} [T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \tag{5.15}$$

Since the functions between brackets to be minimized (or maximized) in (5.12)-(5.15) are strictly convex (or concave) functions of R (or ρ), $\underline{R}_m$, $\overline{R}_m$, $\underline{\rho}^*$ and $\overline{\rho}^*$ are well-defined and unique. We then have the following relations.

Lemma 5.2 *Let $\tau H_{Q_S}(S) < C(W_{Y|X})$ and let $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. Then:*

(a) $\overline{\rho}^*$ and $\underline{\rho}^*$ are positive and finite.

(b) $\overline{R}_m = \tau H_{Q_S^{(\overline{\rho}^*)}}(S)$.

(c) $\underline{R}_m = \tau H_{Q_S^{(\underline{\rho}^*)}}(S)$ if $\underline{\rho}^* < 1$; $\underline{R}_m \geq \tau H_{Q_S^{(1)}}(S)$ if $\underline{\rho}^* = 1$.

Proof: We first prove (a). Since $T_{sp}(\rho, W_{Y|X})$ is the concave hull of $E_o(\rho, W_{Y|X})$, we have the following relation

$$\lim_{\rho \downarrow 0} \frac{T_{sp}(\rho, W_{Y|X})}{\rho} \geq \lim_{\rho \downarrow 0} \frac{E_0(\rho, W_{Y|X})}{\rho} = C(W_{Y|X})$$

where the last equality follows from [8, Lemma 2]. Since $\lim_{\rho \downarrow 0} E_s(\rho, Q_S)/\rho = H(Q_S)$ by (5.3) and Lemma 5.1, we have

$$\lim_{\rho \downarrow 0} \frac{T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)}{\rho} \geq C(W_{Y|X}) - \tau H_{Q_S}(S) > 0.$$

Note that the right-derivative of $T_{sp}(\rho, W_{Y|X})$ (at $\rho = 0$) must exist due to its concavity [80, pp. 113–114], and hence $\lim_{\rho \downarrow 0} T_{sp}(\rho, W_{Y|X})/\rho$ exists. Next we denote $\varepsilon = \tau \log_2 |\mathcal{S}| - R_\infty(W_{Y|X}) > 0$. It follows from the definition of $T_{sp}(\rho, W_{Y|X})$ that

$$\begin{aligned} \lim_{\rho \to \infty} \frac{T_{sp}(\rho, W_{Y|X})}{\rho} &\leq \lim_{\rho \to \infty} \frac{\rho(R_\infty(W_{Y|X}) + \varepsilon/2) + E_{sp}(R_\infty(W_{Y|X}) + \varepsilon/2, W_{Y|X})}{\rho} \\ &= R_\infty(W_{Y|X}) + \varepsilon/2 \end{aligned}$$

because of the finiteness of $E_{sp}(R, W_{Y|X})$ for $R > R_\infty(W_{Y|X})$. This together with

$$\lim_{\rho \to \infty} \frac{E_s(\rho, Q_S)}{\rho} = \log_2 |\mathcal{S}|$$

implies

$$\lim_{\rho\to\infty} \frac{T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)}{\rho} \le R_\infty(W_{Y|X}) + \varepsilon/2 - \tau \log_2 |\mathcal{S}| < 0.$$

Since $T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$ is 0 and has a positive right slope at $\rho = 0$ and is negative for ρ sufficiently large, by the strict concavity of $T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$, the maximum in (5.15) must be achieved by a positive finite $\overline{\rho}^*$. The positivity of $\underline{\rho}^*$ can be shown in the same way and $\underline{\rho}^*$ is finite by its definition.

We next prove (b). If we now regard $\tau e(R/\tau, Q_S)$ as $f^*(y)$ and $\tau E_s(\rho, Q_S)$ as $f(x)$ (by noting that $f^{**} = f$), then according to (4.2) in Fenchel's Duality Theorem,

$$\max_{0\le\rho<\infty} [\rho \overline{R}_m - \tau E_s(\rho, Q_S)] = \overline{\rho}^* \overline{R}_m - \tau E_s(\overline{\rho}^*, Q_S).$$

Setting the derivative of $\rho \overline{R}_m - \tau E_s(\rho, Q_S)$ equal to 0, we can solve for the stationary point[2] $\overline{\rho}^*$, which gives $\overline{R}_m = \tau H_{Q_S^{(\overline{\rho}^*)}}(S)$.

For the lower bound, using a similar argument, we obtain the relation

$$\max_{0\le\rho\le 1} [\rho \underline{R}_m - \tau E_s(\rho, Q_S)] = \underline{\rho}^* \underline{R}_m - \tau E_s(\underline{\rho}^*, Q_S).$$

Recalling that the function between the brackets to be maximized is strictly concave, if the above maximum is achieved by $\underline{\rho}^* \in (0,1)$, then we can solve for the stationary point as above and obtain $\underline{R}_m = \tau H_{Q_S^{(\underline{\rho}^*)}}(S)$. If the maximum is achieved at $\underline{\rho}^* = 1$, then the stationary point is beyond (at least equal to) 1, and hence $\underline{R}_m \ge \tau H_{Q_S^{(1)}}(S)$. Thus (c) follows. ■

In order to summarize the explicit conditions for the calculation of E_J it is convenient to define a critical rate for the source by

$$R_{cr}^{(s)}(Q_S) \triangleq \left.\frac{\partial E_s(\rho, Q_S)}{\partial \rho}\right|_{\rho=1} = H_{Q_S^{(1)}}(S), \tag{5.16}$$

recalling that $Q_S^{(1)}(s) = \sqrt{Q_S(s)}/(\sum_{s'\in\mathcal{S}} \sqrt{Q_S(s')})$, $s \in \mathcal{S}$.

Theorem 5.2 *Let* $\tau H_{Q_S}(S) < C(W_{Y|X})$ *and let* $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. *Then*

[2] The stationary points of a differentiable function $f(x)$ are the solutions of $f'(x) = 0$.

- $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X}) \Longleftrightarrow \overline{\rho}^* \leq 1 \Longleftrightarrow \tau R_{cr}^{(s)}(Q_S) \geq \overline{R}_m = \underline{R}_m \geq R_{cr}(W_{Y|X})$. *In this case,*

$$E_J(Q_S, W_{Y|X}, \tau) = T_{sp}(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S).$$

- $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X}) \Longleftrightarrow \overline{\rho}^* > 1 \Longleftrightarrow R_{cr}(W_{Y|X}) \geq \overline{R}_m > \underline{R}_m = \tau R_{cr}^{(s)}(Q_S)$. *In this case,*

$$E_o(1, W_{Y|X}) - \tau E_s(1, Q_S) \leq E_J(Q_S, W_{Y|X}, \tau) \leq T_{sp}(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S).$$

Remark 5.2 Under the condition $\tau R_{cr}^{(s)}(Q_S) > R_{cr}(W_{Y|X})$, $\overline{\rho}^* = 1$ is possible. However, if $\tau R_{cr}^{(s)}(Q_S) = R_{cr}(W_{Y|X})$, then we definitely have $\overline{\rho}^* = 1$ and $\tau R_{cr}^{(s)}(Q_S) = \overline{R}_m = \underline{R}_m = R_{cr}(W_{Y|X})$.

Remark 5.3 It can be shown that $T_{sp}(1, W_{Y|X}) = E_o(1, W_{Y|X})$ and thus when $\overline{\rho}^* = 1$, the JSCC error exponent is determined by

$$E_J(Q_S, W_{Y|X}, t) = E_o(1, W_{Y|X}) - \tau E_s(1, Q_S).$$

Corollary 5.1 *Let* $\tau H_{Q_S}(S) < C(W_{Y|X})$ *and let* $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. *Then* $\underline{\rho}^* = \min\{1, \overline{\rho}^*\}$ *and* $\underline{R}_m = \tau H_{Q_S^{(\underline{\rho}^*)}}(S)$.

The proof of Theorem 5.2 and Corollary 5.1 is deferred to the next subsection.

Corollary 5.2 *If* $\underline{R}_m \geq R_{cr}(W_{Y|X})$ *or* $\overline{R}_m > R_{cr}(W_{Y|X})$, *then* $\tau R_{cr}^{(s)}(Q_S) \geq \underline{R}_m = \overline{R}_m \geq R_{cr}(W_{Y|X})$, *and the other equivalent conditions in Theorem 5.2 hold.*

Proof: If $\underline{R}_m \geq R_{cr}(W_{Y|X})$ or $\overline{R}_m > R_{cr}(W_{Y|X})$, then $\underline{R}_m = \overline{R}_m$ by Lemma 5.4. $\tau R_{cr}^{(s)}(Q_S) \geq \underline{R}_m$ immediately follows from Corollary 5.1. ■

Remark 5.4 Corollary 5.2 states that if $\underline{R}_m \geq R_{cr}(W_{Y|X})$ or $\overline{R}_m > R_{cr}(W_{Y|X})$, then E_J is determined exactly. Note that when $\overline{R}_m = R_{cr}(W_{Y|X})$, the upper and lower bounds of E_J may not be tight. In that case $\underline{R}_m < R_{cr}(W_{Y|X}) = \overline{R}_m$ is possible. The relation between $\underline{R}_m$ and $\overline{R}_m$ is summarized in Lemma 5.4.

We point out that, in both the computation and analysis aspects, the above conditions play an important role in verifying whether E_J can be determined exactly or not. For the class of symmetric DMCs, we can use the conditions $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ and $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$ to derive explicit formulas for E_J, see Section 5.3.3. In Section 10.2, we apply Theorem 5.2 to establish the conditions for which the JSCC exponent is larger than the tandem coding exponent. Note that when $\tau R_{cr}^{(s)}(Q_S) \leq R_{cr}(W_{Y|X})$, the source-channel random-coding bound admits a simple expression

$$\underline{E}_r(Q_S, W_{Y|X}, t) = E_o(1, W_{Y|X}) - \tau E_s(1, Q_S). \tag{5.17}$$

Consequently, we have the following statement.

Corollary 5.3 *If $\tau R_{cr}^{(s)}(Q_S) \leq R_{cr}(W_{Y|X})$, then Csiszár's random-coding bound and Gallager's lower bound (5.9) are identical.*

Proof: Recall Gallager's lower bound to E_J given by (5.9)

$$\max_{0 \leq \rho \leq 1} [E_o(\rho, W_{Y|X}) - tE_s(\rho, Q_S)] \geq E_o(1, W_{Y|X}) - tE_s(1, Q_S).$$

Since in general Gallager's lower bound cannot be larger than Csiszár's random-coding bound, they must be equal when $\tau R_{cr}^{(s)}(Q_S) \leq R_{cr}(W_{Y|X})$. ■

5.3.2 Proof of Theorem 5.2 and Corollary 5.1

Theorem 5.2 is shown by a left- and right- derivatives argument combined with the results of Lemma 5.2. Let $s_l(R)$ and $s_r(R)$ be the left and right slopes (or left- and right- derivatives) of $E_{sp}(R, W_{Y|X})$ at each $R > R_\infty(W_{Y|X})$. Let $r_l(R)$ and $r_r(R)$ be the left and right slopes of $E_r(R, W_{Y|X})$ at each $R \geq 0$. Let $\rho(R)$ be the slope of $\tau e(R/\tau, Q_S)$ for any $R \in [\tau H_{Q_S}(S), \tau \log_2 |\mathcal{S}|]$. It is easy to verify that these slopes have the following properties (cf. [19], [42], [80]):

(a) $s_l(R)$ and $s_r(R)$ exist for every $R > R_\infty(W_{Y|X})$ and are nondecreasing in R.

(b) $r_l(R)$ and $r_r(R)$ exist for every $R \geq 0$ and are nondecreasing in R.

(c) $s_l(R) \le s_r(R) < -1$ for $R < R_{cr}(W_{Y|X})$, $-1 \le s_l(R) \le s_r(R) \le 0$ for $R_{cr}(W_{Y|X}) < R < C(W_{Y|X})$, and $s_l(R) = s_r(R) = 0$ for $R > C(W_{Y|X})$. $s_l(R_{cr}(W_{Y|X})) \le -1 \le s_r(R_{cr}(W_{Y|X}))$ and $s_l(C(W_{Y|X})) \le 0 = s_r(C(W_{Y|X}))$.

(d) $r_l(R) = r_r(R) = -1$ for $R < R_{cr}(W_{Y|X})$, $r_l(R) = s_l(R)$ for $R > R_{cr}(W_{Y|X})$, and $r_r(R) = s_r(R)$ for $R \ge R_{cr}(W_{Y|X})$. $r_l(R_{cr}(W_{Y|X})) = -1 \le r_r(R_{cr}(W_{Y|X}))$.

(e) $\rho(R)$ is a strictly increasing function of R and is determined by $R = tH\left(Q_S^{(\rho(R))}\right)$ for $\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|$. Specifically, $\rho(\tau H_{Q_S}(S)) = 0$ and $\rho(\tau \log_2 |\mathcal{S}|) = \infty$.

(f) $\overline{\rho}^* = \rho(\overline{R}_m)$, where $\overline{\rho}^*$ and $\overline{R}_m$ are defined in (5.13) and (5.15), respectively.

Proof of (a)–(f): (a) and (b) follows from the convexity of $E_{sp}(R, W_{Y|X})$ for $R > R_\infty(W_{Y|X})$ and $E_r(R, W_{Y|X})$ for $R \ge 0$, see [80, pp. 113–114]. Recalling that $E_r(R, W_{Y|X})$ involves a straight-line section with slope -1 for $R \in [0, R_{cr}(W_{Y|X})]$ and $E_r(R, W_{Y|X}) = E_{sp}(R, W_{Y|X})$ only for $R \ge R_{cr}(W_{Y|X})$, where they both are equal to 0 for $R \ge C(W_{Y|X})$, we obtain (c) and (d) from (a) and (b). From (2.5) and (5.3), we know that $\tau e(R/\tau, Q_S) = tD\left(Q_S^{(\rho^*)} \parallel Q_S\right)$ for $\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|$, where ρ^* is the unique root of $\tau H_{Q_S^{(\rho)}}(S) = R$. Also, it is easy to verify [19] that such ρ^* is exactly the slope of $\tau e(R/\tau, Q_S)$ at R, i.e.,

$$\frac{\partial \tau e(R/\tau, Q_S)}{\partial R} = \rho^*.$$

Thus (e) follows. Recalling also that in Lemma 5.2 we have shown the relation $\overline{R}_m = \tau H_{Q_S^{(\overline{\rho}^*)}}(S)$, since there is unique ρ satisfying this equation, we obtain (f). ■

Based on the above setup, the following lemma illustrates the geometric conditions for which $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ are attained.

Lemma 5.3 Let $\tau H_{Q_S}(S) < C(W_{Y|X})$ and let $R_\infty(W_{Y|X}) < \tau \log_2 |\mathcal{S}|$. The minimum in (5.6) is attained at $\overline{R}_m$ if and only if $-s_l(\overline{R}_m) \ge \rho(\overline{R}_m) \ge -s_r(\overline{R}_m)$, and the minimum in (5.5) is attained at $\underline{R}_m$ if and only if $-r_l(\underline{R}_m) \ge \rho(\underline{R}_m) \ge -r_r(\underline{R}_m)$.

Proof:

1. *Forward part*: We only show the case for the upper bound $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$, since

the case for the lower bound can be shown in a similar manner. We first show that a rate $R_1 \in [\tau H_{Q_S}(S), \tau \log_2 |\mathcal{S}|]$ satisfying $-s_l(R_1) \geq \rho(R_1) \geq -s_r(R_1)$ must achieve the minimum in $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$. Define functions

$$f_1(R) \triangleq \begin{cases} E_{sp}(R, W_{Y|X}) & \text{if} \quad R \leq R_1, \\ E_{sp}(R_1, W_{Y|X}) - \frac{|s_l(R_1)|+|\rho(R_1)|}{2}(R - R_1) & \text{if} \quad R \geq R_1. \end{cases}$$

and

$$g_1(R) \triangleq \begin{cases} \tau e\left(\frac{R}{\tau}, Q_S\right) & \text{if} \quad R \leq R_1, \\ \tau e\left(\frac{R_1}{\tau}, Q_S\right) + \frac{|\rho(R_1)|+|s_l(R_1)|}{2}(R - R_1) & \text{if} \quad R \geq R_1. \end{cases}$$

Since $-s_l(R_1) \geq \rho(R_1)$ implies $s_l(R_1) \leq -(|s_l(R_1)| + |\rho(R_1)|)/2$ and $\rho(R_1) \leq (|\rho(R_1)| + |s_l(R_1)|)/2$, we claim that $f_1(R)$ and $g_1(R)$ are both convex functions and hence their sum is convex,

$$f_1(R) + g_1(R) = \begin{cases} \tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) & \text{if} \quad R \leq R_1, \\ \tau e\left(\frac{R_1}{\tau}, Q_S\right) + E_{sp}(R_1, W_{Y|X}) & \text{if} \quad R \geq R_1. \end{cases}$$

Since the convex function $f_1(R) + g_1(R)$ is constant for $R \geq R_1$ (noting that the convexity is strict in the interval $[\tau H_{Q_S}(S), R_1]$), we may write

$$\min_{\tau H_{Q_S}(S) \leq R \leq R_1} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X})\right] = \tau e\left(\frac{R_1}{\tau}, Q_S\right) + E_{sp}(R_1, W_{Y|X}).$$

Similarly, using the relation $\rho(R_1) \geq -s_r(R_1)$ we can construct convex functions

$$f_2(R) \triangleq \begin{cases} E_{sp}(R, W_{Y|X}) & \text{if} \quad R \geq R_1, \\ E_{sp}(R_1, W_{Y|X}) + \frac{s_r(R_1)-\rho(R_1)}{2}(R - R_1) & \text{if} \quad R \leq R_1. \end{cases}$$

and

$$g_2(R) \triangleq \begin{cases} \tau e\left(\frac{R}{\tau}, Q_S\right) & \text{if} \quad R \geq R_1, \\ \tau e\left(\frac{R_1}{\tau}, Q_S\right) + \frac{\rho(R_1)-s_r(R_1)}{2}(R - R_1) & \text{if} \quad R \leq R_1, \end{cases}$$

and use them to show that the minimum

$$\min_{R_1 \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X})\right]$$

is attained at R_1. Thus, R_1 is the minimizer of $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$, i.e.,

$$\min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right] = \tau e\left(\frac{R_1}{\tau}, Q_S\right) + E_{sp}(R_1, W_{Y|X}).$$

2. *Converse part:* We assume $\overline{R}_m \in (R_\infty(W_{Y|X}), \tau \log_2 |\mathcal{S}|)$ achieves the minimum in (5.6) but $\rho(\overline{R}_m) < -s_r(\overline{R}_m)$. Note that $\rho(\tau \log_2 |\mathcal{S}|) = \infty > -s_r(\tau \log_2 |\mathcal{S}|)$ provided that $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. Now let R_1 be the smallest rate in $[R_\infty(W_{Y|X}), \tau \log_2 |\mathcal{S}|]$ satisfying $\rho(R_1) \ge -s_r(R_1)$. According to our assumption together with (a) and (e), $R_1 > \overline{R}_m$. However, using our previous method, we can construct two convex functions $f_1(R)$ and $g_1(R)$ associated with R_1 to show

$$\min_{\tau H_{Q_S}(S) \le R \le R_1} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right] = \tau e\left(\frac{R_1}{\tau}, Q_S\right) + E_{sp}(R_1, W_{Y|X}).$$

This is clearly contradicted with the assumption that the minimum is attained at $\overline{R}_m$, a rate smaller than R_1, since there is a unique minimum due to the strict convexity. Thus, at $\overline{R}_m$ we must have $\rho(\overline{R}_m) \ge -s_r(\overline{R}_m)$. Consequently, we can show in a similar manner that $\rho(\overline{R}_m) \le -s_l(\overline{R}_m)$. ■

The following facts immediately follow from Lemma 5.3.

Lemma 5.4 We have the following relations between $\overline{R}_m$ and $\underline{R}_m$:

(1). If $\overline{R}_m > R_{cr}(W_{Y|X})$ or $\underline{R}_m \ge R_{cr}(W_{Y|X})$, then $\underline{R}_m = \overline{R}_m > R_{cr}(W_{Y|X})$ and $\overline{E}_{Jsp}(Q_S, W, t) = \underline{E}_{Jr}(Q_S, W, t)$.

(2). If $\overline{R}_m = R_{cr}(W_{Y|X})$, then $\underline{R}_m \le R_{cr}(W_{Y|X})$.

(3). $\overline{R}_m \ge \underline{R}_m$.

Proof: We first show (1). If $\overline{R}_m > R_{cr}(W_{Y|X})$, then we have

$$\min_{R_{cr}(W_{Y|X}) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right] = \tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) + E_{sp}(\overline{R}_m, W_{Y|X}),$$

which means that

$$\min_{R_{cr}(W_{Y|X}) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X}) \right] = \tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) + E_r(\overline{R}_m, W_{Y|X}).$$

Since the minimum of the convex function $\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X})$ of R is achieved by $\overline{R}_m > R_{cr}(W_{Y|X})$, we must have

$$\min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X})\right] = \tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) + E_r(\overline{R}_m, W_{Y|X}),$$

i.e., $\underline{R}_m = \overline{R}_m$. On the other hand, if $\underline{R}_m \ge R_{cr}(W_{Y|X})$, then we have

$$\begin{aligned} e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) &= e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_{sp}(\underline{R}_m, W_{Y|X}) \\ &\ge \min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X})\right], \end{aligned}$$

but by definition we have

$$\begin{aligned} e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) &= \min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X})\right] \\ &\le \min_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X})\right]. \end{aligned}$$

It then follows from the above two inequalities that $\overline{R}_m = \underline{R}_m$.

$$\begin{aligned} e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) &\le e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X}) \\ &\le e\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \end{aligned}$$

for all $R > 0$, which means that $\overline{R}_m = \underline{R}_m$.

We next show (b). If $\overline{R}_m = R_{cr}(W_{Y|X})$, then by Lemma 5.3 and (d), $\rho(R_{cr}(W_{Y|X})) \ge -s_r(R_{cr}(W_{Y|X})) = -r_r(R_{cr}(W_{Y|X}))$. Using Lemma 5.3 again we obtain (2). To show (3), we only need to show the case when $\overline{R}_m < R_{cr}(W_{Y|X})$. According to Lemma 5.3 together with (c) and (d), we see $\rho(\overline{R}_m) > 1$ and $\rho(\underline{R}_m) = 1$. It follows from (e) that $\overline{R}_m > \underline{R}_m$.

■

This lemma emphasizes that when the JSCC error exponent upper bound is achieved at a rate equal to the channel critical rate $R_{cr}(W_{Y|X})$, the lower bound could be achieved at a rate smaller than $R_{cr}(W_{Y|X})$.

In the sequel we shall use properties (c)-(f), and Lemmas 5.2, 5.3 and 5.4 to prove Theorem 5.2. To show $A \Longleftrightarrow B \Longleftrightarrow C$, we only need to show: $A \Longrightarrow B$ (Forward) and

$B \Longrightarrow C \Longrightarrow A$ (Converse).

1. *Converse Part.* We start from

$$
\begin{array}{llll}
 & \overline{\rho}^* < 1 & & \\
\Longrightarrow & \rho(\overline{R}_m) < 1 & \text{(by (f))} & \\
\Longrightarrow & \overline{R}_m < \tau R_{cr}^{(s)}(Q_S) & \text{(by (e))} & \\
\text{and} & s_r(\overline{R}_m) > -1 & \text{(by Lemma 5.3)} & \\
\Longrightarrow & \overline{R}_m \geq R_{cr}(W_{Y|X}) & \text{(by (c))} & \\
\Longrightarrow & \tau R_{cr}^{(s)}(Q_S) > \underline{R}_m = \overline{R}_m > R_{cr}(W_{Y|X}) & \text{(by Lemma 5.4 (1))} & (5.18) \\
\text{or} & \tau R_{cr}^{(s)}(Q_S) > \overline{R}_m = R_{cr}(W_{Y|X}) \geq \underline{R}_m & \text{(by Lemma 5.4 (2))} & (5.19) \\
\Longrightarrow & 0 < \underline{\rho}^* = \overline{\rho}^* < 1 & & (5.20) \\
\text{and} & \tau R_{cr}^{(s)}(Q_S) > \overline{R}_m = \underline{R}_m \geq R_{cr}(W_{Y|X}), & & (5.21)
\end{array}
$$

where (9.56) and (5.21) are explained as follows. We first claim $\underline{\rho}^* < 1$, because $\underline{\rho}^* = 1$ would yield $\underline{R}_m \geq \tau R_{cr}^{(s)}(Q_S)$ by Lemma 5.2 (3), which is contradicted with (9.54) and (9.55). Since now $\underline{\rho}^* < 1$, from Lemma 5.3 and (d) we know $\underline{R}_m \geq R_{cr}(W_{Y|X})$. Thus in (9.55) we must have $\underline{R}_m = R_{cr}(W_{Y|X})$ and consequently (9.54) and (9.55) can both be summarized by (5.21). Meanwhile, $\underline{\rho}^* = \overline{\rho}^*$ follows by Lemma 5.2. If now

$$
\begin{array}{llll}
 & \overline{\rho}^* = 1 & & \\
\Longrightarrow & \rho(\overline{R}_m) = 1 & \text{(by (f))} & \\
\Longrightarrow & \overline{R}_m = \tau R_{cr}^{(s)}(Q_S) & \text{(by (e))} & \\
\text{and} & s_l(\overline{R}_m) \leq -1 \leq s_r(\overline{R}_m) & \text{(by Lemma 5.3)} & \\
\Longrightarrow & \overline{R}_m \geq R_{cr}(W_{Y|X}) & \text{(by (c))} & \\
\Longrightarrow & \tau R_{cr}^{(s)}(Q_S) = \underline{R}_m = \overline{R}_m > R_{cr}(W_{Y|X}) & \text{(by Lemma 5.4 (1))} & (5.22) \\
\text{or} & \tau R_{cr}^{(s)}(Q_S) = \overline{R}_m = R_{cr}(W_{Y|X}) \geq \underline{R}_m & \text{(by Lemma 5.4 (2))} & (5.23) \\
\Longrightarrow & \underline{\rho}^* = \overline{\rho}^* = 1 & & (5.24) \\
\text{and} & \tau R_{cr}^{(s)}(Q_S) = \underline{R}_m = \overline{R}_m \geq R_{cr}(W_{Y|X}), & & (5.25)
\end{array}
$$

where (5.24) and (5.25) are explained as follows. We first claim that $\underline{\rho}^* = 1$. If $\underline{\rho}^* < 1$, then by Lemma 5.2 (3) we have $\underline{R}_m < \tau R_{cr}^{(s)}(Q_S)$. In (5.22), we see $\underline{R}_m = \tau R_{cr}^{(s)}(Q_S)$, contradicted. In (5.23), it is still impossible that $\underline{R}_m < \tau R_{cr}^{(s)}(Q_S) = R_{cr}(W_{Y|X})$, because in that case we have $\rho(\underline{R}_m) < \rho(\tau R_{cr}^{(s)}(Q_S)) = 1$ by (e), which violates Lemma 5.3 since $\underline{R}_m < R_{cr}(W_{Y|X})$ implies $\rho(\underline{R}_m) = 1$. Thus we must have $\underline{\rho}^* = 1$ and (5.24) follows. According to Lemma 5.2 (3) again, $\underline{\rho}^* = 1$ implies $\underline{R}_m \geq \tau R_{cr}^{(s)}(Q_S)$. Hence in (5.23) we must have $\underline{R}_m = \tau R_{cr}^{(s)}(Q_S)$. (5.22) and (5.23) can both be summarized by (5.25). Next if

$$
\begin{aligned}
& \overline{\rho}^* > 1 & & \\
\Longrightarrow \quad & \rho(\overline{R}_m) > 1 & & \text{(by (f))} \\
\Longrightarrow \quad & \overline{R}_m > \tau R_{cr}^{(s)}(Q_S) & & \text{(by (e))} \qquad (5.26) \\
\text{and} \quad & s_l(\overline{R}_m) < -1 & & \text{(by Lemma 5.3)} \\
\Longrightarrow \quad & \overline{R}_m \leq R_{cr}(W_{Y|X}) & & \text{(by (c))} \\
\Longrightarrow \quad & \underline{R}_m \leq \overline{R}_m \leq R_{cr}(W_{Y|X}) & & \text{(by Lemma 5.4 (1) and (3))} \\
\Longrightarrow \quad & \underline{R}_m < R_{cr}(W_{Y|X}) & & \qquad (5.27) \\
\Longrightarrow \quad & r_l(\underline{R}_m) = -1 = r_r(\underline{R}_m) & & \text{(by (d))} \\
\Longrightarrow \quad & \rho(\underline{R}_m) = 1 & & \text{(by Lemma 5.3)} \\
\Longrightarrow \quad & \underline{R}_m = \tau R_{cr}^{(s)}(Q_S) & & \text{(by (e))} \qquad (5.28) \\
\Longrightarrow \quad & \underline{\rho}^* = 1 & & \text{(by Lemma 5.2 (3))} \\
\text{and} \quad & \overline{R}_m > \underline{R}_m. & & \text{(by (5.26) and (5.28)).}
\end{aligned}
$$

To see (5.27), we let $\underline{R}_m = \overline{R}_m = R_{cr}(W_{Y|X})$. Then using (d) and Lemma 5.3 yields $\rho(\underline{R}_m) \leq 1$, which is contradicted with the assumption $\rho(\underline{R}_m) = \rho(\overline{R}_m) > 1$. To show the last step, we assume $\underline{\rho}^* < 1$, then Lemma 5.2 (3) ensures $\underline{R}_m = \tau H_{Q_S^{(\underline{\rho}^*)}}(S) < \tau R_{cr}^{(s)}(Q_S)$, which is contradicted with the last second step.

2. *Forward Part.* First recall that $\rho(\tau R_{cr}^{(s)}(Q_S)) = 1$ by (e). Now if $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$, then $\overline{R}_m$ cannot be strictly larger than $\tau R_{cr}^{(s)}(Q_S)$ because in that case $\rho(\overline{R}_m) > \rho(\tau R_{cr}^{(s)}(Q_S)) = 1$, $-s_l(\overline{R}_m) \leq 1$ by (c), which violates Lemma 5.3. It then follows

$\overline{R}_m \le \tau R_{cr}^{(s)}(Q_S)$ and hence $\overline{\rho}^* \le 1$ by (e). Conversely, if $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$, then $\overline{R}_m$ cannot be less than (or equal to) $\tau R_{cr}^{(s)}(Q_S)$ because in that case $\rho(\overline{R}_m) \le \rho(\tau R_{cr}^{(s)}(Q_S)) = 1$, $-s_r(\overline{R}_m) > 1$ by (c), which violates Lemma 5.3. It then follows $\overline{R}_m > \tau R_{cr}^{(s)}(Q_S)$ and hence $\overline{\rho}^* > 1$ by (e).

Finally, we should note that when $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$, or $\overline{\rho}^* > 1$, the lower bound is achieved by $\underline{R}_m = \tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$ and $\underline{\rho}^* = 1$. Thus

$$\begin{aligned}
\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) &= \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) \\
&= \left[\underline{\rho}^* \underline{R}_m - \tau E_s(\underline{\rho}^*, Q_S)\right] + \left[E_o(1, W_{Y|X}) - \underline{\rho}^* \underline{R}_m\right] \\
&= E_o(1, W_{Y|X}) - \tau E_s(1, Q_S).
\end{aligned}$$

Meanwhile, Corollary 5.1 immediately follows by the above argument. ■

5.3.3 DMS and Symmetric DMC

Consider a DMS Q_S and a *symmetric*[3] DMC $W_{Y|X}$ with rate τ, where the channel transition matrix $W_{Y|X}$ can be partitioned along its columns into sub-matrices $W_{Y|X,1}$, $W_{Y|X,2}$, $\cdots$, $W_{Y|X,s}$, such that in each $W_{Y|X,i}$ with size $|\mathcal{X}| \times |\mathcal{Y}_i|$, each row is a permutation of each other row and each column is a permutation of each other column. Denote the transition probabilities in any column of sub-matrix $W_{Y|X,i}$, $i = 1, 2, \cdots, s$, by $\{p_{i1}, p_{i2}, ..., p_{i|\mathcal{X}|}\}$. Then both $E_o(\rho, W_{Y|X})$ and the channel capacity are achieved by the uniform distribution $P_X = 1/|\mathcal{X}|$ and have the form

$$E_0(\rho, W_{Y|X}) = (1+\rho)\log|\mathcal{X}| - \log\left\{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\} \tag{5.29}$$

and

$$C(W_{Y|X}) = \log|\mathcal{X}| - \frac{1}{|\mathcal{X}|}\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}\right) H_{P_i^{(0)}}(I_{\mathcal{X}}),$$

[3]Here symmetry is defined in the Gallager sense [42, p. 94]; it is a generalization of the standard notion of symmetry [29] (which corresponds to $s = 1$ above).

where the tilted distribution $P_i^{(\alpha)}$, $\alpha \geq 0$, for each $i = 1, 2, \cdots, s$, is defined on $\mathcal{I}_{\mathcal{X}} \triangleq \{1, 2, \cdots, |\mathcal{X}|\}$ by

$$P_i^{(\alpha)}(j) \triangleq \frac{p_{ij}^{\frac{1}{1+\alpha}}}{(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\alpha}})}, \quad j \in \mathcal{I}_{\mathcal{X}}.$$

Since now $E_0(\rho, W_{Y|X})$ is a concave and differentiable function of ρ, the bounds $\underline{E}_{Jr}$ and $\overline{E}_{Jsp}$ can be analytically obtained. If

$$\frac{1}{|\mathcal{X}|} \sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij} \right) H_{P_i^{(0)}}(I_{\mathcal{X}}) + \tau H_{Q_S}(S) < \log_2 |\mathcal{X}| \tag{5.30}$$

and

$$\frac{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} \sqrt{p_{ij}} \right)^2 H_{P_i^{(1)}}(I_{\mathcal{X}})}{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} \sqrt{p_{ij}} \right)^2} + \tau H_{Q_S^{(1)}}(S) \geq \log_2 |\mathcal{X}|, \tag{5.31}$$

then the source-channel exponent is positive and is exactly determined by

$$\begin{aligned} & E_J(Q_S, W_{Y|X}, \tau) \\ = \; & (1+\overline{\rho}^*) \log |\mathcal{X}| - \log \left\{ \left[\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\overline{\rho}^*}} \right)^{1+\overline{\rho}^*} \right] \left(\sum_{s \in \mathcal{S}} Q_S^{\frac{1}{1+\overline{\rho}^*}}(s) \right)^{\tau(1+\overline{\rho}^*)} \right\}, \end{aligned} \tag{5.32}$$

where $\overline{\rho}^*$ is the unique root of the equation

$$\frac{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}} \right)^{1+\rho} H_{P_i^{(\rho)}}(I_{\mathcal{X}})}{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}} \right)^{1+\rho}} + \tau H_{Q_S^{(\rho)}}(S) = \log_2 |\mathcal{X}|. \tag{5.33}$$

In the case when (5.30) does not hold, which means $\tau H_{Q_S}(S) \geq C(W_{Y|X})$, E_J is zero. When (5.30) holds but (5.31) does not hold, the right-hand side of (5.32) becomes the upper bound $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ and meanwhile, E_J is lower bounded by $E_o(1, W_{Y|X}) - \tau E_s(1, Q_S)$, where $E_o(\rho, W_{Y|X})$ is given by (5.29).

Example 5.2 Now we apply the conditions (5.30) and (5.31) to a communication system with a binary source with distribution $\{q, 1-q\}$, a binary symmetric channel (BSC) with crossover probability ε and transmission rates τ =0.5, 0.75, 1, and 1.25. Note that

$$R_{cr}(W_{Y|X}) = 1 - h_b \left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon} + \sqrt{1-\varepsilon}} \right)$$

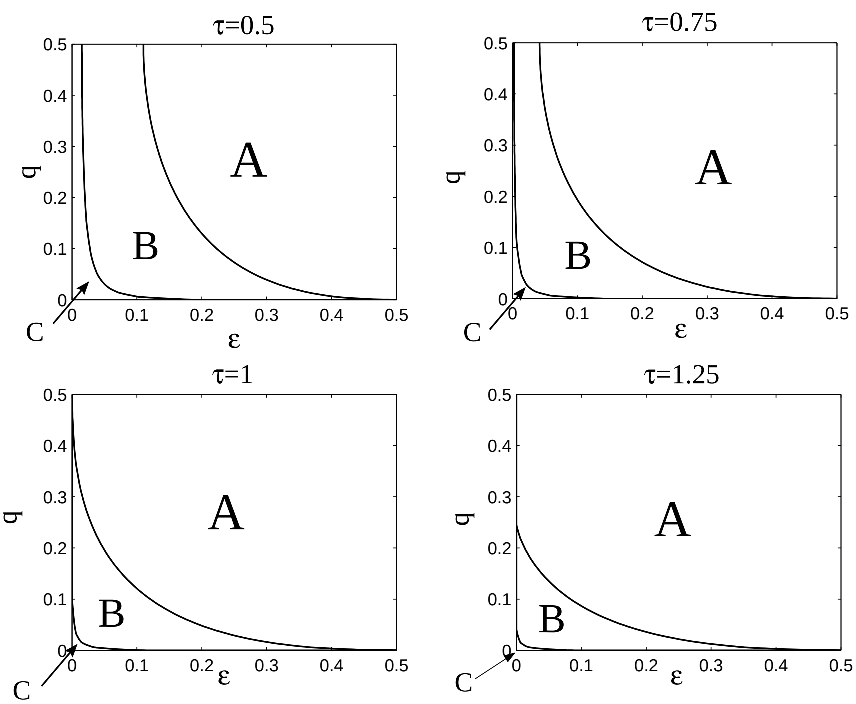

Figure 5.4: The regions for the (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 5.2 for different transmission rates τ. Note that $E_J = 0$ on the boundary between Regions **A** and **B**; E_J is exactly determined on the boundary between Regions **B** and **C**. In Region **A**, $E_J = 0$. In Region **B**, E_J is positive and known exactly. In Region **C**, E_J is positive and can be bounded above and below.

and

$$R_{cr}^{(s)}(Q_S) = h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right),$$

where $h_b(\cdot)$ is binary entropy function. In Fig. 5.4, we partition the set of possible points for the (ε, q) pairs into three regions: **A**, **B** and **C**. If $(\varepsilon, q) \in \mathbf{B}$, where conditions (5.30) and (5.31) hold, i.e., $\tau H_{Q_S}(S) < C(W_{Y|X})$ and $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$, then the corresponding E_J is positive and exactly known. Furthermore, if $(\varepsilon, q) \in \mathbf{C}$, then E_J is bounded above (below, respectively) by the right-hand side of (5.32) ($E_o(1, W_{Y|X}) - \tau E_s(1, Q_S)$, respectively). When $(\varepsilon, q) \in \mathbf{A}$, where $\tau H_{Q_S}(S) > C(W_{Y|X})$, E_J is zero, and the error probability of this communication system converges to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in \mathbf{B} \cup \mathbf{C}$.

5.4 Csiszár's Expurgated Bound

5.4.1 Equivalent Expression

In [31], Csiszár extended his work and obtained another lower bound to E_J for a class of source-channel pairs: for a DMS and a DMC with zero-error capacity equal to 0, if $E_{ex}(R, W_{Y|X}) = \max_{P_X \in \mathcal{P}(\mathcal{X})} E_{ex}(R, P_X, W_{Y|X})$ is attained for a P_X not depending on R, then

$$E_J(Q_S, W_{Y|X}, \tau) \geq \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau) \tag{5.34}$$

where

$$\underline{E}_{Jex}(Q_S, W_{Y|X}, \tau) \triangleq \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e \left(\frac{R}{\tau}, Q_S \right) + E_{ex}(R, W_{Y|X}) \right] \tag{5.35}$$

is called Csiszár's source-channel expurgated lower bound since it contains $E_{ex}(R, W_{Y|X})$ in its expression. We then use Fenchel's Duality Theorem to derive an equivalent expression of $\underline{E}_{Jex}$.

Theorem 5.3 *For a DMS and a DMC with zero-error capacity equal to 0, if*

$$E_{ex}(R, W_{Y|X}) = \max_{P_X \in \mathcal{P}(\mathcal{X})} E_{ex}(R, P_X, W_{Y|X})$$

is attained for a P_X not depending on R, then

$$\underline{E}_{Jex}(Q_S, W_{Y|X}, t) = \sup_{\rho \geq 1} [E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \tag{5.36}$$

Proof: Recall that $E_x(\rho, P_X, W_{Y|X})$ is concave in ρ on the interval $G = [1, +\infty)$ [42, pp. 153–154]. Note that

$$-E_{ex}(R, P_X, W_{Y|X}) \triangleq -\sup_{\rho \in G} [E_x(\rho, P_X, W_{Y|X}) - \rho R] = \inf_{\rho \in G} [\rho R - E_x(\rho; P_X, W_{Y|X})]$$

is the concave transform of $E_x(\rho, P_X, W_{Y|X})$ on $R \in G^* = \{R : -E_{ex}(R, P_X, W_{Y|X}) > -\infty\} = [0, +\infty)$ for DMCs with zero-error capacity equal to 0. Also recall that $\tau E_s(\rho, Q_S)$

is strictly convex in ρ on the interval $F = [0, +\infty)$. Its convex transform

$$\sup_{\rho \in F}[\rho R - \tau E_s(\rho, Q_S)] = \tau e\left(\frac{R}{\tau}, Q_S\right)$$

is a function of R on $F^* = \{R : \tau e(R/\tau, Q_S) < +\infty\} = (-\infty, \tau \log_2 |\mathcal{S}|]$. Fenchel duality theorem states that

$$\inf_{\rho \in F \cap G}[\tau E_s(\rho, Q_S) - E_x(\rho, P_X, W_{Y|X})] = \max_{R \in F^* \cap G^*}\left[-E_{ex}(R, P_X, W_{Y|X}) - \tau e\left(\frac{R}{\tau}, Q_S\right)\right]$$

or

$$\sup_{\rho \geq 1}[E_x(\rho, P_X, W_{Y|X}) - \tau E_s(\rho, Q_S)] = \min_{0 < R \leq \tau \log_2 |\mathcal{S}|}\left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{ex}(R, P_X, W_{Y|X})\right].$$

We can now maximize over P_X and get the two equivalent lower bounds:

$$\begin{aligned}
&\sup_{\rho \geq 1}[E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)] \\
&= \max_{P_X} \min_{0 < R \leq \tau \log_2 |\mathcal{S}|}\left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{ex}(R, P_X, W_{Y|X})\right] \\
&\overset{(a)}{=} \min_{0 < R \leq \tau \log_2 |\mathcal{S}|}\left[\tau e\left(\frac{R}{\tau}, Q_S\right) + \max_{P_X} E_{ex}(R, P_X, W_{Y|X})\right] \\
&\overset{(b)}{=} \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|}\left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{ex}(R, W_{Y|X})\right] \\
&= \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau),
\end{aligned}$$

where (a) follows by assumption that the maximizing P_X does not depend on R and (b) holds since the convex function $\tau e(R/\tau, Q_S) + E_{ex}(R, W_{Y|X})$ is either infinity or strictly decreasing for $R < \tau H_{Q_S}(S)$. ■

In the following lemma we note that the supremum in (5.36) can be replaced by a maximum, and the relation between the maximizer $\underline{\rho}_x$ and its dual minimizer $\underline{R}_{xm}$ is given.

Lemma 5.5 *For DMC with zero-error capacity equal to 0, the function $E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$ has a global maximum at a finite $\rho \geq 1$. Let*

$$\underline{\rho}_x \triangleq \arg\max_{\rho \geq 1}[E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)] \tag{5.37}$$

and

$$\underline{R}_{xm} \triangleq \arg \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|}\left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{ex}(R, W_{Y|X})\right]. \tag{5.38}$$

Then $\underline{R}_{xm} = \tau H_{Q_S^{(\underline{\rho}_x)}}(S)$ *if* $\underline{\rho}_x > 1$; $\underline{R}_{xm} \le \tau R_{cr}^{(s)}(Q_S)$ *if* $\underline{\rho}_x = 1$.

Remark 5.5 Since the function between brackets to be optimized in (5.37) (or (5.38)) is strictly concave (or convex), $\underline{\rho}_x$ and $\underline{R}_{xm}$ are well-defined and unique.

Proof: We first show that $\underline{\rho}_x$ is finite. Recall that for any P_X, Gallager's source function $E_s(\rho, Q_S)$ given in (2.7) and $E_x(\rho; P_X, W_{Y|X})$ given in (2.25) at $\rho = 1$ reduce to

$$E_s(1, Q_S) = \log_2 \left(\sum_{s \in \mathcal{S}} \sqrt{Q_S(s)} \right)^2$$

and

$$E_x(1; P_X, W_{Y|X}) = -\log_2 \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_X(x) \sqrt{P_{Y|X}(y|x)} \right)^2.$$

Using Jensen's inequality [29] on the convex function x^2, we obtain

$$E_s(1, Q_S) \le \log_2 \sum_{s \in \mathcal{S}} (Q_S(s) Q_S(s)^{-1}) = \log_2 |\mathcal{S}|$$

with equality if and only if Q_S is uniform, and

$$E_x(1; P_X, W_{Y|X}) \ge -\log_2 \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_X(x) W_{Y|X}(y|x) = 0.$$

Therefore,

$$E_x(1, W_{Y|X}) - \tau E_s(1, Q_S) > -\log_2 |\mathcal{S}|$$

because of the nonuniform source assumption. On the other hand, because the zero-error capacity is 0 we know that $\lim_{\rho \to \infty} \frac{E_x(\rho, W_{Y|X})}{\rho} = 0$ (from [42, p. 155]) and hence

$$\lim_{\rho \to \infty} \frac{E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)}{\rho} \le -\tau \log_2 |\mathcal{S}|.$$

Clearly, since the concave function $E_x(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$ is finite (bounded below) at $\rho = 1$, and approaches to $-\infty$ as $\rho \to \infty$, there exists a global maximum at a finite $\underline{\rho}_x$. We next show the relation between $\underline{\rho}_x$ and $\underline{R}_{xm}$. Following the proof of Theorem 5.3, let $f^*(y)$ be $\tau e(R/\tau, Q_S)$ and let $f(x)$ be $E_s(\rho, Q_S)$. Fenchel's Duality Theorem (4.2) says that $\underline{\rho}_x$ and $\underline{R}_{xm}$ should satisfy

$$\max_{\rho \ge 1} [\rho \underline{R}_{xm} - \tau E_s(\rho, Q_S)] = \underline{\rho}_x \underline{R}_{xm} - \tau E_s(\rho, Q_S).$$

If $\underline{\rho}_x > 1$, then $\underline{\rho}_x$ is the stationary point of the concave function $\rho \underline{R}_{xm} - \tau E_s(\rho, Q_S)$, and hence

$$\underline{R}_{xm} = \tau H_{Q_S^{(\underline{\rho}_x)}}(S).$$

Otherwise (if $\underline{\rho}_x = 1$), which means that the stationary point is less than or equal to 1, $\underline{R}_{xm} \leq \tau R_{cr}^{(s)}(Q_S)$. ■

Analogously to Theorem 5.2, we have the following explicit conditions regarding the expurgated lower bound to the JSCC exponent.

Theorem 5.4 *For the expurgated lower bound in Theorem 5.3, the following conditions are equivalent.*

- $\tau R_{cr}^{(s)}(Q_S) < R_{ex}(W_{Y|X}) \Longleftrightarrow \underline{\rho}_x > 1 \Longleftrightarrow \tau R_{cr}^{(s)}(Q_S) < \underline{R}_{xm} \leq R_{ex}(W_{Y|X})$. *Thus,*

$$E_J(Q_S, W_{Y|X}, t) \geq E_x(\underline{\rho}_x, W_{Y|X}) - \tau E_s(\underline{\rho}_x, Q_S).$$

- $\tau R_{cr}^{(s)}(Q_S) \geq R_{ex}(W_{Y|X}) \Longleftrightarrow \underline{\rho}_x = 1 \Longleftrightarrow \underline{R}_{xm} = \tau R_{cr}^{(s)}(Q_S) \geq R_{ex}(W_{Y|X})$. *Thus,*

$$E_J(Q_S, W_{Y|X}, t) \geq E_x(1, W_{Y|X}) - \tau E_s(1, Q_S).$$

The proof of Theorem 5.4 is similar to that of Theorem 5.2 and is hence omitted. We next use Theorems 5.2 and 5.4 to compare Csiszár's random-coding and expurgated lower bounds.

5.4.2 Random-coding Lower Bound vs Expurgated Lower Bound

Of clear interest is the case when the expurgated bound improves upon the random-coding bound.

Corollary 5.4 *The source-channel random-coding bound is improved by the expurgated bound (i.e., $\underline{E}_{Jr} < \underline{E}_{Jex}$) if and only if $\tau R_{cr}^{(s)}(Q_S) < R_{ex}(W_{Y|X})$, where $R_{ex}(W_{Y|X})$ is defined in (2.35).*

Proof: When $\tau R_{cr}^{(s)}(Q_S) < R_{ex}(W_{Y|X})$, we must have that $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$, since $R_{ex}(W_{Y|X})$ is never larger than $R_{cr}(W_{Y|X})$. It follows from Theorem 5.2 that the random-coding lower bound is attained at $\underline{R}_m = \tau R_{cr}^{(s)}(Q_S)$. By Theorem 5.4 the expurgated lower bound is attained at $R_{ex}(W_{Y|X}) \geq \underline{R}_{xm} > \tau R_{cr}^{(s)}(Q_S)$. On account of Lemma 5.5, this must happen if $\underline{R}_{xm} = \tau H_{Q_S^{(\underline{\rho}_x)}}(S)$ with $\underline{\rho}_x > 1$. Thus, $\underline{R}_{xm} > \underline{R}_m$ and

$$\begin{aligned}
\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) &= E_r(\underline{R}_m, W_{Y|X}) + \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) \\
&< E_r(\underline{R}_{xm}, W_{Y|X}) + \tau e\left(\frac{\underline{R}_{xm}}{\tau}, Q_S\right) \\
&\leq E_{ex}(\underline{R}_{xm}, W_{Y|X}) + \tau e\left(\frac{\underline{R}_{xm}}{\tau}, Q_S\right) \\
&= \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau).
\end{aligned}$$

In this case, the source-channel expurgated lower bound is tighter than the random-coding lower bound. We then show that $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) \geq \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau)$ if $\tau R_{cr}^{(s)}(Q_S) \geq R_{ex}(W_{Y|X})$.

When $R_{ex}(W_{Y|X}) \leq \tau R_{cr}^{(s)}(Q_S) \leq R_{cr}(W_{Y|X})$, it follows from Theorems 5.2 and 5.4 that

$$\begin{aligned}
\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) &= E_o(1, W_{Y|X}) - \tau E_s(1, Q_S) \\
&= E_x(1, W_{Y|X}) - \tau E_s(1, Q_S) \\
&= \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau),
\end{aligned}$$

where the second equality follows from the fact that, for any P_X, Gallager's channel functions $E_o(1, P_X, W_{Y|X})$ and $E_x(1, P_X, W_{Y|X})$ are equal [42], and hence their maxima are equal. In this case, the source-channel random-coding lower bound is identical to the expurgated lower bound.

When $\tau R_{cr}^{(s)}(Q_S) > R_{cr}(W_{Y|X})$, we must have $\tau R_{cr}^{(s)}(Q_S) > R_{ex}(W_{Y|X})$. Then the expurgated lower bound is attained at $\underline{R}_{xm} = \tau R_{cr}^{(s)}(Q_S)$ by Theorem 5.4. On account of Theorems 5.2 and Corollary 5.1, the random-coding lower bound is attained at $\underline{R}_m =$

$\tau H_{Q_S^{(\underline{\rho}^*)}}(S) \geq R_{cr}(W_{Y|X})$ with $\underline{\rho}^* \leq 1$. Consequently,

$$\begin{aligned}\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) &= E_r(\underline{R}_m, W_{Y|X}) + \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) \\ &\geq E_{ex}(\underline{R}_m, W_{Y|X}) + \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) \\ &\geq E_{ex}(\underline{R}_{xm}, W_{Y|X}) + \tau e\left(\frac{\underline{R}_{xm}}{\tau}, Q_S\right) \\ &= \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau).\end{aligned}$$

In this case, the source-channel random-coding lower bound is tighter than or equal to the expurgated lower bound. ■

5.4.3 DMS and Equidistant DMC

A DMC $W_{Y|X}$ is called equidistant if there exists a number $\beta > 0$ such that for all pairs of inputs $x \neq \widetilde{x}$,

$$\sum_y \sqrt{W_{Y|X}(y|x) W_{Y|X}(y|\widetilde{x})} = \beta.$$

Note that equidistant DMCs have 0 zero-error capacity, and every DMC with binary input alphabet is equidistant. It is shown in [57] that for an equidistant channel, $E_x(\rho, W_{Y|X})$ is achieved in the range $\rho \geq 1$ by a uniform input distribution $P_X(x) = 1/|\mathcal{X}|$. Therefore, we can write $E_x(\rho, W_{Y|X})$ as

$$E_x(\rho, W_{Y|X}) = -\rho \log_2\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|}\right) \qquad \text{for} \qquad \rho \geq 1.$$

Now we apply Theorems 5.3 and 5.4 to DMS Q_S and equidistant DMC $W_{Y|X}$ with transmission rate τ. We then see that if

$$\tau H_{Q_S^{(1)}}(S) + \log_2\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta + \frac{1}{|\mathcal{X}|}\right) \leq \frac{\beta \log \beta}{\beta + \frac{1}{|\mathcal{X}|-1}}, \tag{5.39}$$

the expurgated JSCC lower bound is tighter than the random-coding lower bound and is given by

$$E_J(Q_S, W_{Y|X}, t) \geq -\underline{\rho}_x \log_2\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\underline{\rho}_x}} + \frac{1}{|\mathcal{X}|}\right) - \tau(1+\underline{\rho}_x)\log_2 \sum_{s\in\mathcal{S}} Q_S^{\frac{1}{1+\underline{\rho}_x}}(s), \tag{5.40}$$

where $\underline{\rho}_x$ is the unique root of the equation

$$\tau H_{Q_S^{(\rho)}}(S) + \log_2\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|}\right) = \frac{\rho^{-1}\beta^{\frac{1}{\rho}}\log_2\beta}{\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|-1}}.$$

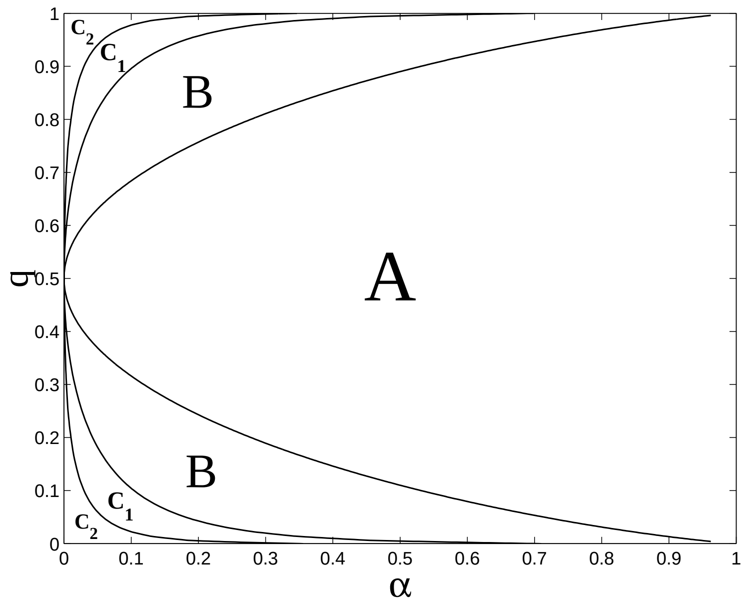

Figure 5.5: The regions for the (α, q) pairs in the binary DMS $\{q, 1-q\}$ and BEC (α) system of Example 5.3 with $\tau = 1$. Note that $E_J = 0$ on the boundary between Regions **A** and **B**; E_J is determined on the boundary between Regions **B** and $\mathbf{C}_1$; The random-coding bound and expurgated bound to E_J are equal on the boundary between Regions $\mathbf{C}_1$ and $\mathbf{C}_2$.

Example 5.3 Consider a communication system with a binary source with distribution $\{q, 1-q\}$, a binary erasure channel (BEC) with erasure probability α and transmission rate $\tau = 1$ (similar results hold for other cases, as in the last example). Using the conditions (5.30), (5.31) in Section 5.3.3, and together with (5.39), we present in Fig. 5.5 the set of (α, q) points, partitioned into four regions. If the pair (α, q) is located in Region **B**, then the system E_J is positive and exactly known. If $(\alpha, q) \in \mathbf{C} = \mathbf{C}_1 \cup \mathbf{C}_2$, then upper and lower bounds for E_J are known. Here, Region $\mathbf{C}_2$ consists of the values of (α, q) for which the source-channel expurgated lower bound given in (5.40) is tighter than the source-channel random-coding lower bound. Finally, when $(\alpha, q) \in \mathbf{A}$, $E_J(Q_S, W_{Y|X}, \tau) = 0$. In Fig. 5.6, we plot the random-coding and expurgated lower bounds for different source and BEC pairs. We observe that when the source distribution is Q_S={0.1,0.9} (respectively Q_S={0.2,0.8}), the expurgated lower bound for E_J is tighter than the random-coding lower bound if $\alpha < 0.0297$ (respectively if $\alpha < 0.0102$).

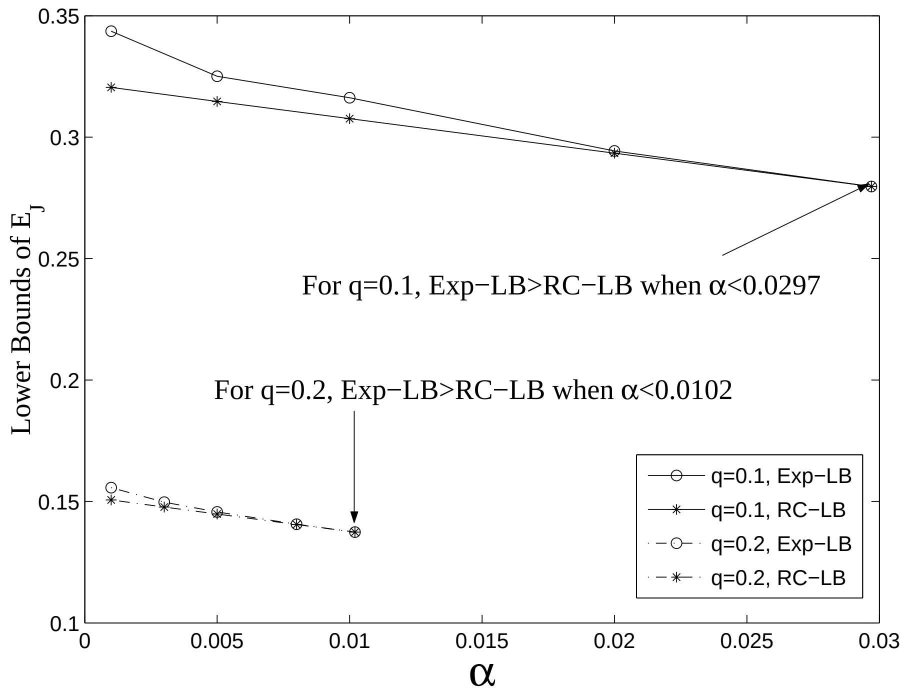

Figure 5.6: Improvement due to the expurgated lower bound for the binary DMS (α, q) and BEC (α) system of Example 5.3 with $\tau = 1$. Exp-LB and RC-LB stand for the expurgated and random-coding lower bounds, respectively.

5.5 JSCC Excess Distortion Exponent with Hamming Distortion Measure

Since in this section we study the (lossy) JSCC excess distortion exponent with a criterion fidelity, we allow the source distribution Q_S to be uniform. Given a distortion measure $d(\cdot,\cdot)$ on $\mathcal{S}\times\mathcal{S}$, a JSC code $(f_n, \varphi_n, \Delta, \tau)$ with blocklength n and transmission rate $\tau > 0$ for a τn-length DMS $Q_S \in \mathcal{P}(\mathcal{S})$ and a DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with a threshold Δ of tolerated distortion is a pair of mappings (see Fig. 5.1)

$$f_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n}.$$

The probability of failing to decode the JSC code $(f_n, \varphi_n, \Delta, \tau)$ within a prescribed distortion level $\Delta > 0$ is called the probability of excess distortion and defined by

$$P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \tau) \triangleq \sum_{\{(\mathbf{s},\mathbf{y}):d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta\}} Q_S^{(n)}(\mathbf{s}) W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})).$$

Definition 5.2 The JSCC excess distortion exponent $E_J^{\Delta}(Q_S, W_{Y|X}, \Delta, \tau)$ is defined as the supremum of the set of all numbers E^{Δ} for which there exists a sequence of JSC codes $(f_n, \varphi_n, \Delta, \tau)$ with blocklength n and transmission rate τ such that

$$E^{\Delta} \le \liminf_{n\to\infty} -\frac{1}{n}\log P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \tau).$$

When there is no possibility of confusion, $E_J^{\Delta}(Q_S, W_{Y|X}, t)$ will often be written E_J^{Δ}. In [31], Csiszár proved that for a DMS Q_S and a DMC $W_{Y|X}$, the JSCC error exponent under distortion threshold Δ satisfies

$$\underline{E}_{Jr}^{\Delta}(Q_S, W_{Y|X}, \tau) \le E_J^{\Delta}(Q_S, W_{Y|X}, \Delta, \tau) \le \overline{E}_{Jsp}^{\Delta}(Q_S, W_{Y|X}, \tau), \tag{5.41}$$

where

$$\underline{E}_{Jr}^{\Delta}(Q_S, W_{Y|X}, \tau) \triangleq \inf_{R>0}\left[\tau F\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_r(R, W_{Y|X})\right] \tag{5.42}$$

and

$$\overline{E}^{\Delta}_{Jsp}(Q_S, W_{Y|X}, \tau) \triangleq \inf_{R>0} \left[\tau F\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R, W_{Y|X}) \right]. \tag{5.43}$$

In the above, $F(R, Q_S, \Delta)$ is defined by (2.11) and is the true value of the DMS excess distortion exponent $e_\Delta(R, P_S)$, and $E_r(R, W_{Y|X})$ and $E_{sp}(R, W_{Y|X})$ are the random-coding and sphere-packing bounds to the channel error exponent. Likewise, if the infimum in (5.42) or (5.43) is attained for a rate larger than the channel critical rate, then the lower and upper bounds coincide (cf. Lemma 5.4), and we can determine E^{Δ}_J exactly. Of course, the two bounds are nontrivial if and only if $\tau R(Q_S, \Delta) < C(W_{Y|X})$ by the lossy JSCC theorem.

By definition $F(R, Q_S, \Delta)$ is a nondecreasing, but not necessarily convex or even continuous in R (cf. Section 2.2.2). Therefore, it is hard to analytically compute the JSCC exponent E^{Δ}_J in general. In this section we only address the computation of E^{Δ}_J for a binary DMS and an arbitrary DMC under the Hamming distortion measure $d_H(\cdot,\cdot)$, given by

$$d_H(s, \widetilde{s}) = \begin{cases} 1, & \text{if } s \neq \widetilde{s}, \\ 0, & \text{if } s = \widetilde{s}. \end{cases} \tag{5.44}$$

We first need to derive a parametric form of $F(R, Q_S, \Delta)$. Define

$$E^{\Delta}_s(\rho, Q_S) \triangleq (1+\rho)\log\left(q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}\right) - \rho h_b(\Delta). \tag{5.45}$$

Lemma 5.6 For binary DMS $Q_S \triangleq \{q, 1-q\}$ ($q \leq 1/2$) under the Hamming distortion measure (5.44) and distortion threshold Δ such that $\Delta \leq 1/2$, the function $F(R, Q_S, \Delta)$ given by (2.11) is equivalent to

$$F(R, Q_S, \Delta) = \begin{cases} +\infty, & R > 1 - h_b(\Delta), \\ \sup_{\rho \geq \rho_0}[\rho R - \tau E^{\Delta}_s(\rho, Q_S)], & R(Q_S, \Delta) < R \leq 1 - h_b(\Delta), \\ 0, & R \leq R(Q_S, \Delta), \end{cases} \tag{5.46}$$

where the rate-distortion function $R(Q_S, \Delta) = h_b(q) - h_b(\Delta)$ and $\rho_0 = 0$ if $q \geq \Delta$; otherwise $R(Q_S, \Delta) = 0$ and ρ_0 is the unique root of equation $H(Q_S^{(\rho)}) = h_b(\Delta)$ such that $\rho_0 > 0$.

Proof: Recall that the rate-distortion function $R(Q_S, \Delta)$ for a binary DMS $Q_S = \{q, 1-q\}$ under the Hamming distortion measure is given by (e.g., [29])

$$R(Q_S, \Delta) = \begin{cases} h_b(q) - h_b(\Delta), & 0 \le \Delta \le q, \\ 0, & \Delta > q. \end{cases} \tag{5.47}$$

Clearly, $F(R, Q_S, \Delta) = 0$ for $R \le 0$ since the infimum in (2.11) is attained at $P_S = Q_S$. Similarly, since $R(P_S, \Delta) \le 1 - h_b(\Delta)$ for all P_S, $F(R, Q_S, \Delta) = \infty$ for $R > 1 - h_b(\Delta)$. For the remainder of the proof, we assume $0 < R \le 1 - h_b(\Delta)$.

(1) *Case of* $0 \le \Delta \le q$. For $R \le R(Q_S, \Delta) = h_b(q) - h_b(\Delta)$, we have

$$F(R, Q_S, \Delta) = \inf_{P_S : R(P_S,\Delta) > R} D(P_S \parallel Q_S) = D(P_S \parallel Q_S)\Big|_{P_S = Q_S} = 0.$$

For $h_b(q) - h_b(\Delta) < R \le 1 - h_b(\Delta)$, we have

$$\begin{aligned} F(R, Q_S, \Delta) &= \inf_{P_S : R(P_S,\Delta) > R} D(P_S \parallel Q_S) \\ &= \min_{P_S \triangleq \{p, 1-p\} : R(P_S,\Delta) = R} D(P_S \parallel Q_S) \qquad (5.48) \\ &= \min_{p : h_b(p) - h_b(\Delta) = R} D(P_S \parallel Q_S) \\ &= e(R + h_b(\Delta), Q_S), \qquad \text{for } H(Q_S) \le R + h_b(\Delta) \le \log|\mathcal{S}| \qquad (5.49) \\ &= \sup_{\rho \ge 0} [\rho(R + h_b(\Delta)) - E_s(\rho)] \qquad (5.50) \\ &= \sup_{\rho \ge 0} [\rho R - (E_s(\rho) - \rho h_b(\Delta))]. \end{aligned}$$

Here (5.48) follows from the facts that the continuous function $\theta(p) \triangleq p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$ is increasing for $p \ge q$ and $R(P_S, \Delta)$ given in (5.56) is continuous and increasing in p for $\Delta \le p \le \frac{1}{2}$. In (5.49), we note that $H(Q_S) = h_b(q)$ and that $\log|\mathcal{S}| = 1$ as the source is binary. (5.50) follows by the well known parametric form of source exponent function introduced by Blahut [19] and noting that $R' \triangleq R + h_b(\Delta) \in [H(Q_S), \log|\mathcal{S}|]$.

(2) *Case of* $\Delta > q$. For $0 < R \le 1 - h_b(\Delta)$, similarly as (5.49), we have

$$F(R, Q_S, \Delta) = e(R', Q_S) = \sup_{\rho \in A} [\rho R' - E_s(\rho)],$$

where $R' = R + h_b(\Delta)$ such that $H(Q_S) < h_b(\Delta) < R' \le 1 = \log|\mathcal{S}|$ and

$$\begin{aligned} A &= \left\{\rho^* : \left.\frac{\partial[\rho R' - E_s(\rho)]}{\partial \rho}\right|_{\rho=\rho^*} = 0, \qquad h_b(\Delta) \le R' \le 1\right\} \\ &= \left\{\rho^* : h_b(\Delta) \le R' = H(Q_S^{(\rho^*)}) \le 1\right\} \\ &= \{\rho^* : \rho_0 \le \rho^* < \infty\}, \end{aligned} \tag{5.51}$$

where ρ_0 is the unique root of equation $H(Q_S^{(\rho)}) = h_b(\Delta)$ and $\rho_0 > 0$. Here (5.51) follows from the monotone property of $H(Q_S^{(\rho)})$. Therefore, we write

$$F(R, Q_S, \Delta) = \sup_{\rho \ge \rho_0} [\rho R - (E_s(\rho) - \rho h_b(\Delta))].$$

In fact, it can be shown that ρ_0 is the right slope of $F(R, Q_S, \Delta)$ at $R = R(Q_S, \Delta)$. ■

Define the binary divergence by

$$\widetilde{D}(\Delta \parallel q) \triangleq \Delta \log \frac{\Delta}{q} + (1-\Delta) \log \frac{1-\Delta}{1-q}. \tag{5.52}$$

Theorem 5.5 *Given a binary DMS ($q \le 1/2$) and a DMC $W_{Y|X}$ under the Hamming distortion measure and distortion threshold Δ ($\Delta \le 1/2$), the JSCC exponent satisfies the following.*

1) Lower Bound: If $0 \le \Delta < \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, then $\rho_0 < 1$ and

$$\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau) = \max_{\rho_0 \le \rho \le 1} [T_r(\rho, W_{Y|X}) - \tau E_s^{\Delta}(\rho, Q_S)], \tag{5.53}$$

Otherwise, if $\Delta \ge \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, then

$$\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau) = \tau \widetilde{D}(\Delta \parallel q) + E_0(1, W_{Y|X}). \tag{5.54}$$

2) Upper Bound:

$$\overline{E}_{sp}^{\Delta}(Q_S, W_{Y|X}, \tau) = \sup_{\rho \ge \rho_0} [T_{sp}(\rho, W_{Y|X}) - \tau E_s^{\Delta}(\rho, Q_S)]. \tag{5.55}$$

Proof: It can be easily verified that $F(R, Q_S, \Delta)$ is continuous and convex in $R \in (-\infty, 1-h_b(\Delta)]$ if $q \ge \Delta$ and $F(R, Q_S, \Delta)$ is continuous and convex in $R \in (0, 1-h_b(\Delta)]$ and has a jump at $R = R(Q_S, \Delta) = 0$ if $q < \Delta$. In either case, $F(R, Q_S, \Delta)$ has a right-slope ρ_0 at $R = 0$, where ρ_0 is defined in Lemma 5.6.

It follows from a geometric argument (as in Lemma 5.3) regarding the right- and left-slopes that if $\rho_0 \geq 1$, or equivalently, if $\Delta \geq \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, $\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau)$ in (5.42) is achieved at $R \downarrow 0^+$, and

$$\begin{aligned}
\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau) &= \lim_{R \downarrow 0^+} \left[\tau F\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_r(R, W_{Y|X}) \right] \\
&= \lim_{R \downarrow 0^+} \left[\tau \inf_{P_S : R(P_S, \Delta) > \frac{R}{\tau}} D(P_S \parallel Q_S) + E_0(1, W_{Y|X}) - R \right] \\
&= \tau \widetilde{D}(\Delta \parallel q) + E_0(1, W_{Y|X}).
\end{aligned}$$

Otherwise, if $\rho_0 < 1$, $\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau)$ in (5.42) is achieved at $R > 0$. In this case, we define

$$\widetilde{F}(R, Q_S, \Delta) = \begin{cases} F(R, Q_S, \Delta), & R > 0, \\ \lim_{R \downarrow 0} F(R, Q_S, \Delta), & R = 0. \end{cases} \tag{5.56}$$

Clearly, replacing $F(R, Q_S, \Delta)$ by $\widetilde{F}(R, Q_S, \Delta)$ in $\underline{E}_{Jr}^{\Delta}(Q_S, W_{Y|X}, \tau)$ does not affect the lower bound. In fact, we can rewrite

$$\underline{E}_{Jr}^{\Delta}(Q_S, W_{Y|X}, \tau) = \inf_{0 \leq R \leq 1 - h_b(\Delta)} \left[\tau \widetilde{F}\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_r(R, W_{Y|X}) \right].$$

For the upper bound, noting that ρ_0 is finite, we always can write

$$\overline{E}_{Jsp}^{\Delta}(Q_S, W_{Y|X}, \tau) \triangleq \inf_{0 \leq R \leq 1 - h_b(\Delta)} \left[\tau \widetilde{F}\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R, W_{Y|X}) \right],$$

since the above infimum can never be achieved at $R = 0$ by a simple right- (left-) slope argument (cf. Lemma 5.3). Now $\widetilde{F}(R, Q_S, \Delta)$ is convex and continuous in $[0, 1 - h_b(\Delta)]$. It follows by Lemma 5.6 that $\widetilde{F}(R, Q_S, \Delta)$ and $E_s^{\Delta}(\rho, Q_S)$ are a pair of convex Fenchel transforms, i.e.,

$$\widetilde{F}(R, Q_S, \Delta) = (E_s^{\Delta}(\rho, Q_S))^*, \quad R \in [0, 1 - h_b(\Delta)]$$

and

$$E_s^{\Delta}(\rho, Q_S) = (\widetilde{F}(R, Q_S, \Delta))^*, \quad R \in [\rho_0, +\infty).$$

Adopting the approach of Section 5.2, we can apply Fenchel duality theorem to $\underline{E}_{Jr}^{\Delta}(Q_S, W_{Y|X}, \tau)$ for the case $\rho_0 < 1$, i.e., $0 \le \Delta < \sqrt{q}/(\sqrt{q} + \sqrt{1-q})$, and $\overline{E}_{Jsp}^{\Delta}(Q_S, W_{Y|X}, \tau)$ and obtain equivalent computable bounds (5.53) and (5.55). ■

Remark 5.6

1) As in the lossless case, if $\tau(h_b(q) - h_b(\Delta)) \ge C(W_{Y|X})$, then $\underline{E}_r^{\Delta}(Q_S, W_{Y|X}, \tau) = \overline{E}_{sp}^{\Delta}(Q_S, W_{Y|X}, \tau) = 0$. If $R_\infty(W_{Y|X}) > \tau(1 - h_b(\Delta))$, then $\overline{E}_{sp}^{\Delta}(Q_S, W_{Y|X}, \tau) = +\infty$.

2) In the special case where the binary source is uniform, i.e., $q = 1/2$, Theorem 5.5 reduces to

$$\begin{aligned} \max_{0\le\rho\le1} \left[-\rho\tau(1 - h_b(\Delta)) + T_r(\rho, W_{Y|X})\right] &\le E_J^{\Delta}(Q_S, W_{Y|X}, \tau) \\ &\le \sup_{\rho\ge0} \left[-\rho\tau(1 - h_b(\Delta)) + T_{sp}(\rho, W_{Y|X})\right]. \end{aligned}$$

This is clearly equivalent to

$$E_r\left(\tau(1 - h_b(\Delta)), W_{Y|X}\right) \le E_J^{\Delta}(Q_S, W_{Y|X}, \tau) \le E_{sp}\left(\tau(1 - h_b(\Delta)), W_{Y|X}\right) \quad (5.57)$$

by the definition of T_r and T_{sp}. In other words, E_J^{Δ} is bounded by the channel random-coding and sphere-packing bounds at rate $\tau(1-h_b(\Delta))$. If $\tau(1-h_b(\Delta)) \ge R_{cr}(W_{Y|X})$, then E_J^{Δ} is exactly determined.

3) When the source is nonuniform, $E_s^{\Delta}(\rho, Q_S) = E_s(\rho, Q_S) - \rho\tau h_b(\Delta)$ is strictly concave in ρ. In this case, the maximizer

$$\overline{\rho}^{\Delta} \triangleq \arg\sup_{\rho\ge\rho_0} [T_{sp}(\rho, W_{Y|X}) - \tau E_s^{\Delta}(\rho, Q_S)]$$

is strictly larger than ρ_0 if $\tau(h_b(q) - h_b(\Delta)) < C(W_{Y|X})$ and $R_\infty(W_{Y|X}) \le \tau(1 - h_b(\Delta))$. Particularly, $\overline{\rho}^{\Delta} < \infty$ if $R_\infty(W_{Y|X}) < t(1 - h_b(\Delta))$. As counterparts of Lemma 5.2 and Corollary 5.1, it can be shown that the upper bound $\overline{E}_{sp}^{\Delta}(Q_S, W_{Y|X}, \tau)$ in (5.43) is attained at $\overline{R}_m^{\Delta} = H_{Q_S^{(\overline{\rho}^{\Delta})}}(S) - h_b(\Delta)$ and the lower bound in (5.42) is attained at $\underline{R}_m^{\Delta} = H_{Q_S^{(\underline{\rho}^{\Delta})}}(S) - h_b(\Delta)$, where $\underline{\rho}^{\Delta} = \min\{\overline{\rho}^{\Delta}, 1\}$. Consequently, other similar results to the lossless case regarding these optimizers can be obtained.

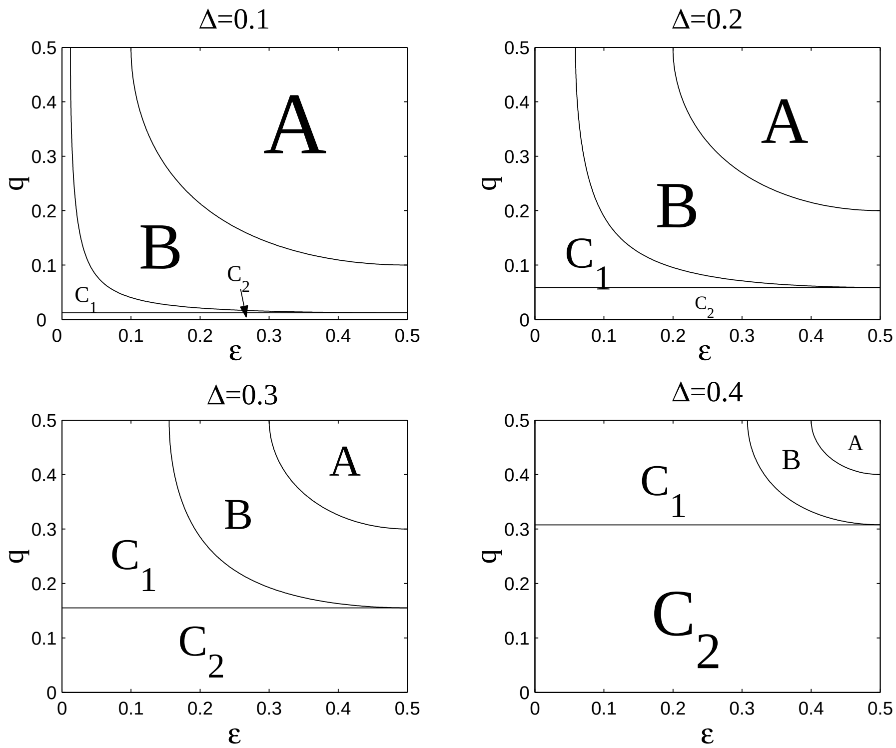

Figure 5.7: The regions for the (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 5.4 with Hamming distortion for different values of the distortion threshold Δ with $t = 1$. Note that $E_J^\Delta = 0$ on the boundary between Regions **A** and **B**, and $E_J^\Delta > 0$ is determined on the boundary between Regions **B** and $\mathbf{C}_1$.

Example 5.4 For a binary DMS $\{q, 1-q\}$ $(q \leq 0.5)$ and a BSC (ε) under transmission rate $t = 1$, we compute the JSCC error exponent under the Hamming distortion measure with distortion threshold Δ $(\Delta < \frac{1}{2})$. In Fig. 5.7, if the pair (ε, q) is located in Region **B**, then the corresponding JSCC exponent can be determined exactly (the lower and upper bounds are equal). If (ε, q) is located in Region $\mathbf{C}_1$, then E_J^Δ is bounded by (5.53) and (5.55). If (ε, q) is located in Region $\mathbf{C}_2$, then E_J^Δ is bounded by (5.54) and (5.55). When $(\varepsilon, q) \in \mathbf{A}$, E_J^Δ is zero, and the error probability of this communication system converges to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in \mathbf{B} \cup \mathbf{C}_1 \cup \mathbf{C}_2$.

Fig. 5.8 shows the JSCC error exponent lower bound of the binary DMS $\{q, 1-q\}$ ($q < 0.5$) and BSC (ε) pairs under different distortion thresholds. We fix the BSC parameter $\varepsilon = 0.2$, and vary q from 0 to 0.5. In Fig. 5.8, Segment 1 is determined by (5.54), and Segments 2 and 3 are determined by (5.53). Furthermore, the lower bound coincides with the upper bound (5.55) in Segment 3; i.e., the JSCC exponent is exactly determined in Segment 3.

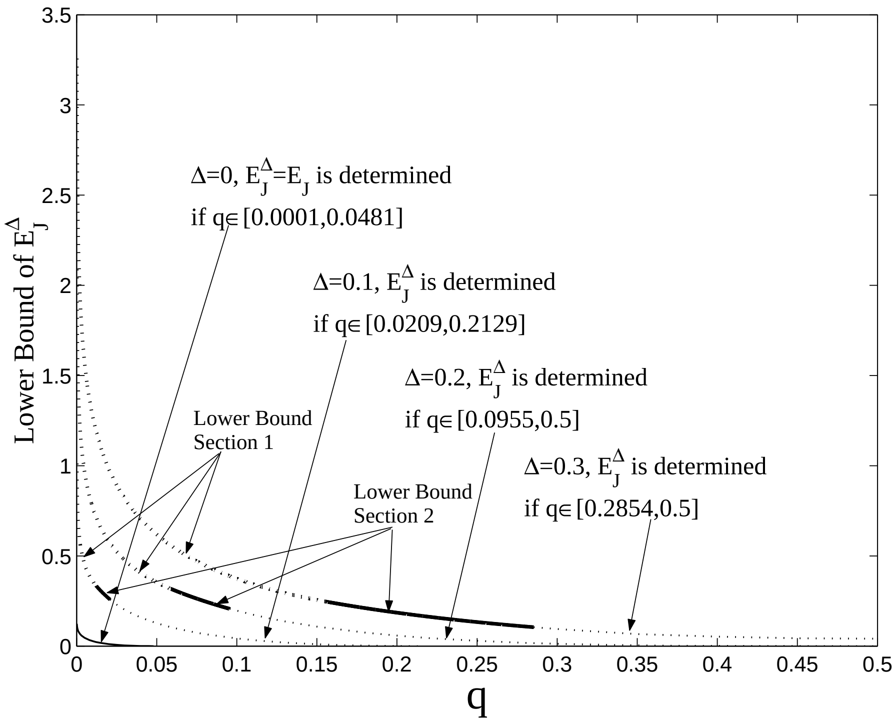

Figure 5.8: Fix $\varepsilon = 0.2$. The JSCC excess ditortion exponent lower bound of the binary DMS $\{q, 1-q\}$ ($q \leq 0.5$) and BSC (ε) pairs under Hamming distortion with $\tau = 1$. For $\Delta = 0$, E_J^Δ is determined if $q \in [0.0001, 0.0481]$, which is the same as the random-coding lower bound for the lossless JSCC error exponent. For $\Delta = 0.1$, E_J^Δ is determined if $q \in [0.0209, 0.2129]$. For $\Delta = 0.2$, E_J^Δ is determined if $q \in [0.0955, 0.5]$. For $\Delta = 0.3$, E_J^Δ is determined if $q \in [0.2854, 0.5]$.

5.6 Conclusion

In this chapter, we established equivalent parametric representations of Csiszár's lower and upper bounds, $\underline{E}_{Jr}$, $\underline{E}_{Jex}$ and $\overline{E}_{Jsp}$, for the JSCC exponent E_J of a communication system with a DMS and a DMC. As a result, the computation of the bounds for E_J is facilitated for arbitrary DMS-DMC pairs. Furthermore, the bounds enjoy closed-form expressions when the channel is symmetric (in the Gallager sense). Notice that Csiszár's random-coding lower bound for E_J is in general larger than Gallager's lower bound; they are identical if the channel is symmetric. We obtained explicit sufficient and necessary conditions for $\underline{E}_{Jr} = \overline{E}_{Jsp}$ and $\underline{E}_{Jex} > \underline{E}_{Jr}$. Finally, we partially investigated the computation of Csiszár's lower and upper bounds for the lossy JSCC exponent for binary sources and DMCs under the Hamming distortion measure, and obtained equivalent representations for these bounds using the same approach as for the lossless JSCC exponent.

Chapter 6

JSCC Error Exponent with Feedback/Source Side Information

In this chapter we discuss the JSCC error exponent with feedback or source side information (SI). The question we aim to answer, is whether output feedback or source SI can strictly increase the JSCC error exponent.

In Section 6.1, we consider the discrete memoryless JSCC system with perfect (noiseless and instantaneous) causal feedback. We obtain an upper and a lower bound for the JSCC error exponent with feedback $E_{J,fb}$. The upper bound follows directly from the definition of the exponent and the corresponding channel error exponent with feedback by using Csiszár's approach based on the method of types. More specifically, we actually show that Csiszár's JSCC sphere packing upper bound $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ given by (5.6) is still valid for the feedback case, i.e., $E_{J,fb} \leq \overline{E}_{Jsp}$. We next establish a Gallager-type lower bound for $E_{J,fb}$, expressed in terms of source and channel functions (see (5.9)), using an iterative coding scheme proposed by Zigangirov [115]. We then examine this lower bound for channels with binary input alphabet and a symmetric distribution (in the Gallager sense). A sufficient condition for which $E_{J,fb}$ is determined exactly by its lower and upper bound is provided. In Section 6.2, we investigate the situation for which the JSCC error exponent with feedback can be strictly larger than the exponent when there is no feedback.

By numerically comparing the lower bound for $E_{J,fb}$ and the upper bound of E_J, the JSCC error exponent without feedback, we present a few examples to show that $E_{J,fb}$ could be considerably larger than E_J.

We next extend the JSCC problem by considering the availability of SI on the transmitted source at the decoder. We establish an (achievable) lower bound for the JSCC error exponent. The lower bound follows from a two-stage encoding two-stage decoding scheme which combines the approaches of Csiszár [30] and Oohama and Han [73] and is based on the method of types. In particular, at the decoding side, we employ a generalized maximum mutual information decoder followed by a minimum conditional entropy decoder.

Furthermore, in Section 6.5, we analytically compare the lower bound for the exponent with source SI at the decoder, $\underline{E}_J^{SID}$, with $\overline{E}_{Jsp}$, the sphere-packing upper bound for the exponent without SI. We derive a sufficient condition for which the source SI at the decoder can strictly enlarge the JSCC error exponent for a system consisting of a binary source and a symmetric channel. The sufficient and necessary condition for which the source can be reliably transmitted over the channel, i.e., the JSCC theorem, is also formulated in Section 6.4. It is seen that the source-channel separation principle holds. Finally, a conclusion is drawn in Section 6.6.

6.1 Systems with Feedback

Before we deal with the JSCC feedback system, we first introduce the channel coding problem with feedback and review some results on the channel error exponent with feedback.

6.1.1 Literature Review: Channel Coding with Perfect Feedback

Given a message set $\mathcal{M}_n \triangleq \{1, 2, ..., M_n\}$ and a DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$, a causal channel code with block length n and perfect (noiseless and instantaneous) output feedback (see Fig. 6.1) consists a set of encoder-mappings $\{f_{c,r}\}_{r=1}^n$, where

$$f_{c,r} : \mathcal{M}_n \times \mathcal{Y}^{r-1} \longrightarrow \mathcal{X}, \quad 1 \le r \le n,$$

and one decoder-mapping

$$\varphi_{cn} : \mathcal{Y}^n \longrightarrow \mathcal{M}_n.$$

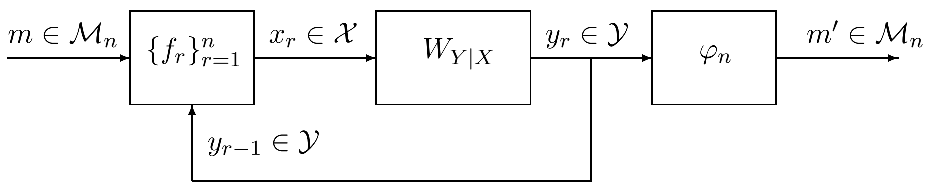

Figure 6.1: Causal channel coding system with perfect feedback.

Let the transmitted message m be uniformly and independently drawn from the message set $\mathcal{M}_n$. The rate of the channel code $(\{f_{c,r}\}_{r=1}^n, \varphi_{cn})$ is defined by

$$R_n \triangleq \frac{\log_2 M_n}{n} \quad \text{bits/channel use.}$$

Let the corresponding n-length codeword be $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ and the received codeword be $\mathbf{y} = (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$. Then the probability of receiving y_r $(1 \leq r \leq n)$ at the r-th instant under the conditions that the message m is transmitted, that the input codeword is $x_1, x_2, ...x_r$, and that $y_1, y_2, ..., y_{r-1}$ has been previously accepted is given by

$$\Pr(Y_r = y_r | Y_1 = y_1, Y_2 = y_2, ..., Y_{r-1} = y_{r-1}, m) = W_{Y|X}(y_r | x_r).$$

For the sake of convenience, we denote the r-th component of the codeword

$$x_r = f_{c,r}(y_1, y_2, ..., y_{r-1}, m)$$

by $f_{c,r}(m)$. Therefore, the probability that a sequence $\mathbf{y}$ is received conditional on that the message m has been transmitted is given by

$$P_{W^n, f_c}(\mathbf{y}|m) \triangleq \prod_{r=1}^{n} W_{Y|X}(y_r | f_{cr}(m)),$$

and the probability of error for the channel code $(\{f_{c,r}\}_{r=1}^n, \varphi_{cn})$ with rate R_n is given by

$$P_{efb}^{(n)}(R_n, W_{Y|X}) = \frac{1}{M_n} \sum_{m=1}^{M_n} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y}) \neq m} P_{W^n, f_c}(\mathbf{y}|m). \tag{6.1}$$

Definition 6.1 For any $R > 0$, the channel error exponent $E_{fb}(R, W_{Y|X})$ with perfect feedback is defined as the supremum of all numbers E_c for which there exists a sequence of channel codes $(\{f_{c,r}\}_{r=1}^{n}, \varphi_{cn})$ with

$$E_c \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{efb}^{(n)}(R, W_{Y|X}).$$

and

$$R \leq \liminf_{n\to\infty} R_n.$$

Lower and upper bounds for the (fixed-length) channel coding error exponent with feedback have been studied in [11, 22, 37, 88, 115]. Sheverdyaev [88] proved that the sphere-packing upper bound for the DMC without feedback is also valid in the feedback case, i.e., for $R > 0$, $E_{fb}(R, W_{Y|X}) \leq E_{sp}(R, W_{Y|X})$. On the other hand, Zigangirov [115] proposed a coding scheme based on a set of likelihood functions and obtained a lower bound for the error exponent for the BSC with feedback. His lower bound was extended by D'yachkov [37] for K-ary symmetric channels with feedback and the bound for the BSC was later improved for small rates by Burnashev [22]. These works show that at least for the DMC with binary input, the error exponent for channel coding with feedback is determined exactly by the sphere-packing bound in the interval $R > R_{cr,fb}(W_{Y|X})$ for some $R_{cr,fb}(W_{Y|X})$ strictly less than the critical rate $R_{cr}(W_{Y|X})$ for the channel. Furthermore, it is shown in [37] that at zero rate, the channel error exponent with feedback strictly outperforms the one without feedback for both the K-ary symmetric DMC and the DMC with binary input, whenever the zero-error capacity of the channel is equal to zero.

6.1.2 JSCC System with Perfect Feedback

We now extend the channel coding with feedback to a JSCC setup. Consider the following causal JSCC system with perfect feedback (see Fig. 6.2). Given a DMS $\{Q_S : \mathcal{S}\}$ with alphabet $\mathcal{S}$, a DMC $\{W_{Y|X} : \mathcal{X} \to \mathcal{Y}\}$ with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$, and the transmission rate τ (source symbol/channel use), a causal JSC code with block length n with perfect feedback consists a set of encoder-mappings $\{f_r\}_{r=1}^{n}$ where

$$f_r : \mathcal{S}^{\tau n} \times \mathcal{Y}^{r-1} \longrightarrow \mathcal{X}, \quad 1 \leq r \leq n,$$

and one decoder-mapping

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n}.$$

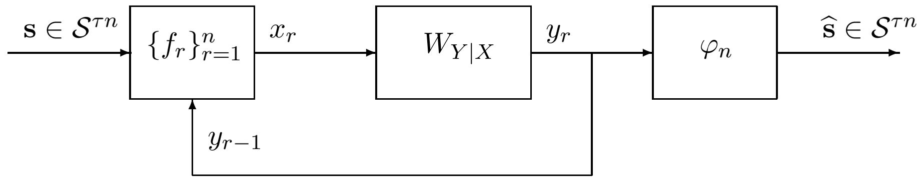

Figure 6.2: Causal JSCC system with perfect feedback.

Let the transmitted source message be $\mathbf{s} \in \mathcal{S}^{\tau n}$, the corresponding n-length codeword be $\mathbf{x} \in \mathcal{X}^n$ and the received codeword be $\mathbf{y} \in \mathcal{Y}^n$. Then similarly to the channel coding system, the probability of receiving y_r $(1 \leq r \leq n)$ at the r-th instant under the conditions that the message $\mathbf{s}$ is transmitted, that the input codeword is $x_1, x_2, ...x_r$, and that $y_1, y_2, ..., y_{r-1}$ has been previously accepted is given by

$$\Pr(Y_y = y_r | Y_1 = y_1, Y_2 = y_2, ..., Y_{r-1} = y_{r-1}, S^n = \mathbf{s}) = W_{Y|X}(y_r|x_r).$$

For the sake of convenience, we denote the r-th component of the codeword

$$x_r = f_r(y_1, y_2, ..., y_{r-1}, \mathbf{s})$$

by $f_r(\mathbf{s})$. The probability that a sequence $\mathbf{y}$ is received conditional on that the source $\mathbf{s}$ has been transmitted is given by

$$P_{W^n,f}(\mathbf{y}|\mathbf{s}) \triangleq \prod_{r=1}^{n} W_{Y|X}(y_r|f_r(\mathbf{s})),$$

and the probability of error for the JSC code $(\{f_r\}_{r=1}^n, \varphi_n)$ is given by

$$P_{efb}^{(n)}(Q_S, W_{Y|X}, \tau) = \sum_{\{(\mathbf{s},\mathbf{y}):\varphi_n(\mathbf{y}) \neq \mathbf{s}\}} Q_{S^{\tau n}}(\mathbf{s}) P_{W^n,f}(\mathbf{y}|\mathbf{s}). \tag{6.2}$$

Definition 6.2 The JSCC error exponent $E_{Jfb}(Q_S, W_{Y|X}, \tau)$ with perfect feedback is defined as the supremum of all numbers E for which there exists a sequence of JSC codes $(\{f_r\}_{r=1}^n, \varphi_n)$ with

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{efb}^{(n)}(Q_S, W_{Y|X}, \tau).$$

When there is no possibility of confusion, $E_{Jfb}(Q_S, W_{Y|X}, \tau)$ will be written as E_{Jfb}. In the following we shall derive an upper bound and a lower bound for E_{Jfb}.

6.1.3 Upper Bound for JSCC Error Exponent with Feedback

First of all, we can upper bound $E_{Jfb}(Q_S, W_{Y|X}, \tau)$ in the same way as Csiszár's did for the JSCC upper bound (5.4) without feedback by using a simple type-partitioning (or type counting) argument.

Theorem 6.1 *Given Q_S, $W_{Y|X}$, and $\tau > 0$ with perfect feedback,*

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) \leq \inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{fb}(R, W_{Y|X}) \right], \tag{6.3}$$

where $e(R, Q_S)$ is the source error exponent, and $E_{fb}(R, W_{Y|X})$ is the channel error exponent with feedback defined by Definition 6.1.

Proof: Due to (2.5), we first write

$$\begin{aligned} &\inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{fb}(R, W_{Y|X}) \right] \\ &= \inf_{P_S \in \mathcal{P}(\mathcal{S})} \left[\tau D(P_S \| Q_S) + E_{fb}(\tau H_{P_S}(S), W_{Y|X}) \right]. \end{aligned}$$

We assume that the above is finite (the upper bound is trivial if it is infinity) and the infimum actually becomes a minimum. Let the minimum be achieved by distribution $P_S^* \in \mathcal{P}(\mathcal{S})$, then there must exist a sequence of types $\left\{ \widehat{P}_S \in \mathcal{P}_{\tau n}(\mathcal{S}) \right\}_{n=n_o}^{\infty}$ such that $\widehat{P}_S \to P_S^*$ uniformly.[1]

Next rewrite the probability of error given in (6.2) as a sum of probabilities of types and lower bound it by

$$P_{efb}^{(n)}(Q_S, W_{Y|X}, \tau) = \sum_{P_S \in \mathcal{P}_{\tau n}(\mathcal{S})} Q_S^{(\tau n)}(\mathbb{T}_{P_S}) P_{efb}(\mathbb{T}_{P_S}) \geq Q_S^{(\tau n)}(\mathbb{T}_{\widehat{P}_S}) P_{efb}(\mathbb{T}_{\widehat{P}_S}) \tag{6.4}$$

[1] The sequence $\left\{ \widehat{P}_S \in \mathcal{P}_{\tau n}(\mathcal{S}) \right\}_{n=n_o}^{\infty}$ here denotes a sequence for $n = n_o, 2n_o, 3n_o, \ldots$, where n_o is the smallest integer such that τn is also an integer.

where

$$P_{efb}(\mathbb{T}_{\widehat{P}_S}) \triangleq \frac{1}{|\mathbb{T}_{\widehat{P}_S}|} \sum_{\mathbf{s}\in\mathbb{T}_{\widehat{P}_S}} \sum_{\mathbf{y}:\varphi_n(\mathbf{y})\neq\mathbf{s}} P_{W^n,f}(\mathbf{y}|\mathbf{s}).$$

Note that $P_{efb}(\mathbb{T}_{\widehat{P}_S})$ can be interpreted as the probability of error of an n-block *channel* code $(\{f_{c,r}\}_{r=1}^n, \varphi_{cn})$ with message set $\mathcal{M}_n = \mathbb{T}_{\widehat{P}_S}$ for the channel $W_{Y|X}$, since all the sequences $\mathbf{s} \in \mathbb{T}_{\widehat{P}_S}$ are equiprobable. Now setting $R_n = \frac{1}{n}\log_2 |\mathbb{T}_{\widehat{P}_S}|$, by the definition of the channel error exponent with feedback and Lemma 3.1,

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{efb}(\mathbb{T}_{\widehat{P}_S}) \le E\left(\liminf_{n\to\infty} R_n, W_{Y|X}\right) = E(\tau H_{\widehat{P}_S}(S), W_{Y|X}) \tag{6.5}$$

for any sequence of JSC codes $(\{f_r\}_{r=1}^n, \varphi_n)$. According to Lemma 3.1 again, we know that for any $\widehat{P}_S \in \mathcal{P}_{\tau n}(\mathcal{S})$

$$-\frac{1}{\tau n}\log_2 Q_S^{(\tau n)}(\mathbb{T}_{\widehat{P}_S}) \le D(\widehat{P}_S\|Q_S) + |\mathcal{S}|\frac{1}{\tau n}\log_2(1+\tau n)$$

which implies

$$\limsup_{n\to\infty} -\frac{1}{n}\log_2 Q_S^{(\tau n)}(\mathbb{T}_{\widehat{P}_S}) \le \tau D(\widehat{P}_S\|Q_S). \tag{6.6}$$

It then follows from (6.4), (6.5), and (6.6) that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{efb}^{(n)}(Q_S, W_{Y|X}, \tau) \\
\le\ & \liminf_{n\to\infty}\left[-\frac{1}{n}\log_2 Q_S^{(\tau n)}(\mathbb{T}_{\widehat{P}_S}) - \frac{1}{n}\log_2 P_{efb}(\mathbb{T}_{P_S})\right] \\
\le\ & \limsup_{n\to\infty} -\frac{1}{n}\log_2 Q_S^{(\tau n)}(\mathbb{T}_{\widehat{P}_S}) + \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{efb}(\mathbb{T}_{\widehat{P}_S}) \\
\le\ & \tau D(P_S^*\|Q_S) + E_{fb}(\tau H_{P_S^*}(S), W_{Y|X}).
\end{aligned}$$

Since the above bound holds for any sequence of JSC codes, we complete the proof of Theorem 6.1. ■

Observation 6.1 As we mentioned before, we can see from the proof that this approach to prove the JSCC upper bound, based on a type counting argument, can be adapted to other discrete JSCC systems. Indeed, as long as we can partition the source space via a polynomial number of type classes, and we can rewrite the averaged probability of error for each type class as a channel coding probability error, then we can obtain a similar conceptual bound expressed in terms of the sum of source and channel error exponents. Note that this conceptual bound cannot currently be computed as the channel error exponent is not yet fully known for all coding rates, but it directly implies that any upper bound for the channel error exponent yields a corresponding upper bound for the JSCC error exponent.

Since $E_{fb}(R, W_{Y|X}) \leq E_{sp}(R, W_{Y|X})$ due to Sheverdyaev [88], the following bound is obvious.

Corollary 6.1 *Given Q_S, $W_{Y|X}$, and $\tau > 0$ with perfect feedback,*

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) \leq \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) \tag{6.7}$$

where $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ is Csiszár's source-channel sphere-packing upper bound given by (5.6).

By definition, the error exponent for systems with feedback must be larger than the one without feedback; otherwise we just ignore the feedback information and then we can achieve the same performance as the one without feedback. Thus, a trivial lower bound for $E_{Jfb}(Q_S, W_{Y|X}, \tau)$ follows

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) \geq E_J(Q_S, W_{Y|X}, \tau) \geq \max\{\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau), \underline{E}_{Jex}(Q_S, W_{Y|X}, \tau)\}, \tag{6.8}$$

and consequently the following condition follows from the results of Chapter 5.

Corollary 6.2 *If $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$, then feedback cannot improve the JSCC reliability function, i.e.,*

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) = E_J(Q_S, W_{Y|X}, \tau).$$

6.1.4 Lower Bound for JSCC Error Exponent with Feedback

In the sequel, we are hence interested in the case when $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$. A Gallager-type lower bound will be derived by modifying the approach in [115] (also see [22, 37]) for the channel coding exponent bound with feedback.

We first introduce the coding scheme. Let $P_X(x)$ be an arbitrary distribution on $\mathcal{X}$ with $P_X(x) > 0$ for any $x \in \mathcal{X}$, and let $B = [b_{xy}]_{|\mathcal{X}|\times|\mathcal{Y}|}$ be an arbitrary $|\mathcal{X}| \times |\mathcal{Y}|$ matrix with nonnegative components and nonzero columns (B will be specified in the proof of Theorem 6.2). For each transmission instant $r = 1, 2, ..., n$, we consider a set of likelihood functions $T_r(\mathbf{s})$ for each source message $\mathbf{s} \in \mathcal{S}^{\tau n}$ such that

$$\sum_{\mathbf{s}} T_r(\mathbf{s}) = 1 \quad \text{and} \quad T_r(\mathbf{s}) \geq 0.$$

Thus each $T_r = \{T_r(\mathbf{s})\}$ can be regarded as a probability distribution. The initial distribution T_1, and the iterative algorithm between T_r and T_{r+1} will be specified below.

When $r = 1$, let the initial distribution T_1 be the tilted distribution (see Section 5.1.2) of the source distribution $T_1 = Q_{S^{\tau n}}^{(\lambda)}$ defined on $\mathcal{S}^{\tau n}$, where $\lambda \geq 0$ is arbitrary and will be optimized later. Note that this is different with the coding scheme in [115] for channel coding with feedback, where T_1 is set to be a uniform distribution, i.e., $T_1 = 1/M_n$ where M_n is the size of the message set. Now both the encoder and the decoder know T_1, and they would employ the same algorithm to calculate the next likelihood function for each $\mathbf{s}$. We assume at the r-th transmission instant, the encoder and decoder have T_r.

Encoding rule. Denote $M = |S|^{\tau n}$ and $K = |\mathcal{X}|$ with $\mathcal{X} = \{x^{(1)}, x^{(2)}, ..., x^{(K)}\}$. Before each transmission and based on T_r, the encoder distributes all the M possible source messages into K groups in the following way. The encoder orders the likelihood functions and the distribution P_X decreasingly, say, $T_r(\mathbf{s}_1) \geq T_r(\mathbf{s}_2) \geq \cdots \geq T_r(\mathbf{s}_M)$ and $P_X(x^{(1)}) \geq P_X(x^{(2)}) \geq \cdots \geq P_X(x^{(K)})$. For the first K source messages, we assign $\mathbf{s}_1$ to group G_1, assign $\mathbf{s}_2$ to group G_2,..., and finally, we assign $\mathbf{s}_K$ to group G_K. This procedure is to make each group nonempty. For the source messages $\mathbf{s}_{K+1}$ up to $\mathbf{s}_M$, we successively assign each source message to a group G_j according to the following rule:

$$j = \arg \min_{1 \leq i \leq K} \frac{\sum_{\mathbf{s} \in G_i} T_r(\mathbf{s})}{P_X(x^{(i)})}.$$

With the grouping completed, if the transmitted source message $\mathbf{s} \in G_i$, then transmit the channel symbol $x^{(i)}$.

Iterative algorithm. For both the encoder and the decoder, after transmitting x_r at the r-th instant, assume the channel output is y_r. The encoder and the decoder calculate the next likelihood function $T_{r+1}(\cdot)$ for each $\mathbf{s}$ using T_r and y_r in the following way:

$$T_{r+1}(\mathbf{s}) = \frac{b_{x(\mathbf{s})y_r} T_r(\mathbf{s})}{\sum_{\mathbf{s}} b_{x(\mathbf{s})y_r} T_r(\mathbf{s})}$$

where $x(\mathbf{s}) = x^{(i)}$ if $\mathbf{s} \in G_i$, $i = 1, 2, ..., K$. Note that the decoder also makes the grouping to obtain the same G_i's. Clearly, $T_{r+1}(\mathbf{s})$ is also a probability distribution.

Decoding rule. After the n-th transmission, the decoder would make a decision to say which source message was transmitted. Based on the likelihood functions and the last received symbol y_n, the decoder computes T_{n+1} and output the source message $\mathbf{s}$ with the largest likelihood function

$$\widehat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \mathcal{S}^{\tau n}} T_{n+1}(\mathbf{s}).$$

We next show that under the above coding and decoding procedure, the following bound is achievable.

Theorem 6.2 *Given Q_S, $W_{Y|X}$, and $\tau > 0$ with perfect feedback,*

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) \geq \sup_{\rho \geq 0} \left[E_o^{fb}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S) \right], \tag{6.9}$$

where

$$E_o^{fb}(\rho, W_{Y|X}) = \max_{\mu \geq 0} \max_{P_X} \min_{a \in \mathcal{X}} \min_{\boldsymbol{q}: q(a) \leq P_X(a)} E_o^{fb}(\rho, W_{Y|X}, \mu, P_X, a, \boldsymbol{q})$$

where the minimum is taken over the probability distribution $\boldsymbol{q}$ on $\mathcal{X}$ (i.e., $\sum_{x \in \mathcal{X}} q(x) = 1$, $q(x) \geq 0$) such that $q(a) \leq P_X(a)$ and

$$E_o^{fb}(\rho, W, \mu, P_X, a, \boldsymbol{q}) = -\log_2 \sum_{y \in \mathcal{Y}} W_{Y|X}^{\frac{1+\mu-\rho}{1+\mu}}(y|a) \left[\sum_{x \in \mathcal{X}} q(x) W_{Y|X}^{\frac{1}{1+\mu}}(y|x) \right]^{\rho}.$$

Proof: According to the decoding rule, an error occurs if $T_{n+1}(\mathbf{s})$ is not the largest when $\mathbf{s}$ is the transmitted source message. Clearly, an upper bound follows

$$\begin{aligned} P_e^{(n)}(Q_S, W_{Y|X}, \tau) &= \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \Pr\left(T_{n+1}(\mathbf{s}) \text{ is not the largest} \middle| \mathbf{s} \text{ is transmitted}\right) \\ &\leq \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \Pr\left(T_{n+1}(\mathbf{s}) \leq \frac{1}{2} \middle| \mathbf{s} \text{ is transmitted} \right). \end{aligned}$$

Let

$$V_0 \triangleq \frac{T_1(\mathbf{s})}{1-T_1(\mathbf{s})} = \frac{\frac{Q_{S^{\tau n}}(\mathbf{s})^{\frac{1}{1+\lambda}}}{\sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})^{\frac{1}{1+\lambda}}}}{1-\frac{Q_{S^{\tau n}}(\mathbf{s})^{\frac{1}{1+\lambda}}}{\sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})^{\frac{1}{1+\lambda}}}}$$

and

$$V_r \triangleq \frac{T_{r+1}(\mathbf{s})}{1-T_{r+1}(\mathbf{s})}\frac{1-T_r(\mathbf{s})}{T_r(\mathbf{s})}, \quad r = 1, 2, ..., n.$$

We may write, for any $\rho \geq 0$,

$$\begin{aligned}
&P_e^{(n)}(Q_S, W_{Y|X}, \tau) \\
&\leq \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})\Pr\left(\frac{T_{n+1}(\mathbf{s})}{1-T_{n+1}(\mathbf{s})} \leq 1 \middle| \mathbf{s} \text{ is transmitted}\right) \\
&= \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})\Pr\left(\prod_{r=1}^{n} V_r^{-\rho} \geq V_0^{\rho} \middle| \mathbf{s} \text{ is transmitted}\right) \\
&\overset{(a)}{\leq} \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) V_0^{-\rho}\mathbb{E}\left[\prod_{r=1}^{n} V_r^{-\rho} \middle| \mathbf{s} \text{ is transmitted}\right] \\
&\leq \left(\sum_{\mathbf{s}'} Q_{S^{\tau n}}(\mathbf{s}')^{\frac{1}{1+\lambda}}\right)^{\rho} \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})^{\frac{1+\lambda-\rho}{1+\lambda}}\mathbb{E}\left[\prod_{r=1}^{n} V_r^{-\rho} \middle| \mathbf{s} \text{ is transmitted}\right], \quad (6.10)
\end{aligned}$$

where (a) follows from Markov's inequality. At this point we need to borrow an important result from [37]. It has been shown in [37] that for arbitrary channel input distribution P_X,

$$\mathbb{E}\left[\prod_{r=1}^{n} V_r^{-\rho} \middle| \mathbf{s} \text{ is transmitted}\right] \leq H(\rho, B, P_X, W_{Y|X})^n,$$

where

$$H(\rho, B, P_X, W_{Y|X}) = \max_{a \in \mathcal{X}} H(\rho, B, P_X, W_{Y|X}, a)$$

is independent of $\mathbf{s}$ with

$$H(\rho, B, P_X, W_{Y|X}, a) = \max_{\mathbf{q}: q(a) \leq P_X(a)} \sum_{y \in \mathcal{Y}} W_{Y|X}(y|a)\left[\frac{\sum_{x \in \mathcal{X}} q(x) b_{xy}}{b_{ay}}\right]^{\rho},$$

where the maximum is taken over the probability distribution $\mathbf{q}$ on $\mathcal{X}$ (i.e., $\sum_{x \in \mathcal{X}} q(x) = 1$, $q(x) \geq 0$) such that $q(a) \leq P_X(a)$. Now setting $b_{xy} = W_{Y|X}(y|x)^{\frac{1}{1+\mu}}$ for an arbitrary $\mu \geq 0$, and noting that the distribution P_X is arbitrary, we obtain

$$\mathbb{E}\left[\prod_{r=1}^{n} V_r^{-\rho} \middle| \mathbf{s} \text{ is transmitted}\right] \leq 2^{-nE_o^{fb}(\rho, W_{Y|X})}. \quad (6.11)$$

On the other hand, by noting that the source is a DMS with $Q_{S^{\tau n}}(\mathbf{s}) = \prod_{i=1}^{\tau n} Q_S(s_i)$, we have

$$\left(\sum_{\mathbf{s}'} Q_{S^{\tau n}}(\mathbf{s}')^{\frac{1}{1+\lambda}}\right)^{\rho} \left(\sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s})^{\frac{1+\lambda-\rho}{1+\lambda}}\right) = f(\lambda)^{\tau n},$$

where

$$f(\lambda) = \left(\sum_{s} Q_S(s)^{\frac{1+\lambda-\rho}{1+\lambda}}\right) \left(\sum_{s} Q_S(s)^{\frac{1}{1+\lambda}}\right)^{\rho}, \quad \lambda \geq 0.$$

To obtain a good upper bound on the probability of error, we need to optimize the parameter λ for fixed ρ. It is not hard to check that

$$\left.\frac{\partial f(\lambda)}{\partial \lambda}\right|_{\lambda=\rho} = 0 \quad \text{and} \quad \left.\frac{\partial^2 f(\lambda)}{\partial \lambda^2}\right|_{\lambda=\rho} \geq 0.$$

Thus

$$\min_{\lambda \geq 0} f(\lambda) \leq \left(\sum_{s} Q_S(s)^{\frac{1}{1+\rho}}\right)^{1+\rho} = 2^{E_s(\rho,Q)}. \tag{6.12}$$

In the above, we write "$\leq$" instead of "$=$" because we do not know whether $\lambda = \rho$ is a global minimum point. Finally, substituting (6.11) and (6.12) into (6.10) and maximizing the exponent over all $\rho \geq 0$ yields the desired lower bound (6.9). ■

Remark 6.1 Unlike λ in the proof, we are not able to find the best μ for a fixed ρ (especially for $\rho \geq 1$) for general DMC's.

The lower bound (6.9) has a similar parametric form as Gallager's lower bound for JSCC error exponent without feedback (5.9), i.e.,

$$\max_{0 \leq \rho \leq 1} [E_o(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)],$$

and we have shown that, Gallager's lower bound is equal to Csiszár's lower bound $\underline{E}_{Jr}$ if the channel function $E_o(\rho, W_{Y|X})$ is achieved by a distribution P_X independent of ρ. In fact, we have the following relation.

Corollary 6.3 *For DMC with binary input alphabet $\mathcal{X} = \{0, 1\}$, the lower bound given in (6.9) is at least as large as Gallager's random-coding lower bound without feedback.*

Proof: We first restrict the range of ρ by

$$\sup_{\rho\geq 0}\left[E_o^{fb}(\rho,W_{Y|X})-\tau E_s(\rho,Q_S)\right]\geq \max_{0\leq\rho\leq 1}\left[E_o^{fb}(\rho,W_{Y|X})-\tau E_s(\rho,Q_S)\right].$$

It then suffices to show $E_o^{fb}(\rho,W_{Y|X})\geq E_o(\rho,W_{Y|X})$ for any given $\rho\in[0,1]$. When $0\leq\rho\leq 1$, the function

$$\left[\sum_{x'\in\mathcal{X}} q(x')W_{Y|X}^{\frac{1}{1+\mu}}(y|x')\right]^{\rho}$$

is a concave function of $\mathbf{q}$ and the constraints $\sum_x q(x)=1$ and $q(a)\leq P_X(a)$ will be achieved with equality. Thus,

$$\begin{aligned}&\max_{a\in\mathcal{X}}\max_{\mathbf{q}:q(a)\leq P_X(a)}\sum_{y\in\mathcal{Y}} W_{Y|X}^{\frac{1+\mu-\rho}{1+\mu}}(y|a)\left[\sum_{x\in\mathcal{X}} q(x)W_{Y|X}^{\frac{1}{1+\mu}}(y|x)\right]^{\rho}\\ =\;&\max_{a\in\{0,1\}}\sum_{y\in\mathcal{Y}} W_{Y|X}^{\frac{1+\mu-\rho}{1+\mu}}(y|a)\left[\sum_{x\in\mathcal{X}} P_X(x)W_{Y|X}^{\frac{1}{1+\mu}}(y|x)\right]^{\rho}.\end{aligned}$$

Next we set $\mu=\rho$ and choose P_X^* $(P_X^*(a)>0)$ such that

$$\sum_{y\in\mathcal{Y}} W_{Y|X}^{\frac{1}{1+\rho}}(y|a)\left[\sum_{x\in\mathcal{X}} P_X^*(x)W_{Y|X}^{\frac{1}{1+\rho}}(y|x)\right]^{\rho}=\sum_{y\in\mathcal{Y}}\left[\sum_{x\in\mathcal{X}} P_X^*(x)W_{Y|X}^{\frac{1}{1+\rho}}(y|x)\right]^{1+\rho}$$

for every $a\in\mathcal{X}$. We know from [42, Theorem 5.6.5] that such P_X^* must achieve the maximum of $E_o(\rho,W_{Y|X})$. Thus

$$E_o^{fb}(\rho,W_{Y|X})\geq \min_{a\in\mathcal{X}}\min_{\mathbf{q}:q(a)\leq P_X^*(a)} E_o^{fb}(\rho,W,\mu=\rho,P_X=P_X^*,a,\mathbf{q})=E_o(\rho,W_{Y|X}).$$

■

However, it is difficult to evaluate the bound (6.9) for general DMC's, and we do not know if the lower bound can improve Gallager's bound in general, since when $\rho\geq 1$, it turns out that

$$\sum_{y\in\mathcal{Y}} W_{Y|X}(y|x)\left[\frac{\sum_{x'\in\mathcal{X}} q(x')b_{x'y}}{b_{x'y}}\right]^{\rho}$$

is a convex function of $\mathbf{q}$, and the maximum would be achieved at some boundary points. When $W_{Y|X}$ is a K-ary symmetric channel[2] and a DMC with binary input, an analytic formula for $E_o^{fb}(\rho,W_{Y|X})$ is given in [37].

[2] In this case $W_{Y|X}$ is determined by only two parameters, i.e., the elements on the diagonal of the $K\times K$ transition matrix are the same, and the other elements are the same.

To be simple, we next use the results of [37] to specialize the lower bound given in (6.9) for binary input channels with symmetric distribution (in the Gallager sense, cf. Section 5.3.3), and derive a sufficient condition in terms of information rates for which E_{Jfb} is determined exactly. Let the DMC have binary input alphabet $\mathcal{X} = \{0, 1\}$. For $x \geq 0$ define functions

$$g_0(x) \triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|0)^{\frac{1}{1+x}} W_{Y|X}(y|1)^{\frac{x}{1+x}}$$

and

$$g_1(x) \triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|1)^{\frac{x}{1+x}} W_{Y|X}(y|0)^{\frac{1}{1+x}}.$$

It has been shown in [37] that for such channel

$$E_o^{fb}(\rho, W_{Y|X}) = E_o(\rho, W_{Y|X}) \tag{6.13}$$

for $\rho \leq \rho^*$ where $\rho^* = \min\{\rho_1, \rho_2\}$ and $\rho_i > 1$ $(i = 1, 2)$ is the unique root of

$$-\log_2 g_i(\rho) = E_o(\rho, W_{Y|X}).$$

Otherwise (if $\rho \geq \rho^*$ and hence $\rho > 1$),

$$E_o^{fb}(\rho, W_{Y|X}) = -\log_2 \min_{\mu \geq \rho - 1} \max \left\{ f_1(\rho, \mu, 0), f_2(\rho, \mu, 0), f_1\left(\rho, \mu, \frac{1}{2}\right), f_1\left(\rho, \mu, \frac{1}{2}\right) \right\} \tag{6.14}$$

where

$$\begin{aligned} f_1(\rho, \mu, q) &\triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|0)^{1-\frac{\rho}{1+\mu}} \left[W_{Y|X}(y|0)^{\frac{1}{1+\mu}} q + W_{Y|X}(y|1)^{\frac{1}{1+\mu}} (1-q) \right]^\rho, \\ f_2(\rho, \mu, q) &\triangleq \sum_{y \in \mathcal{Y}} W_{Y|X}(y|1)^{1-\frac{\rho}{1+\mu}} \left[W_{Y|X}(y|1)^{\frac{1}{1+\mu}} q + W_{Y|X}(y|0)^{\frac{1}{1+\mu}} (1-q) \right]^\rho. \end{aligned}$$

Theorem 6.3 *Let* $|\mathcal{X}| = 2$, $\tau H_{Q_S}(S) < C(W_{Y|X})$, *and* $\log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. *If* $E_o(\rho, W_{Y|X})$ *is achieved by a* P_X *independent of* ρ, *then*

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) = \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau), \tag{6.15}$$

for the source-channel pairs satisfying

$$\tau H_{Q_S^{(\rho^*)}}(S) \geq R_{cr,fb}(W_{Y|X}), \tag{6.16}$$

where $Q_S^{(\rho^*)}$ *is the tilted distribution with respect to* ρ^* *and* $R_{cr,fb}(W_{Y|X}) = E'_o(\rho^*, W_{Y|X})$ *is a number strictly less than the critical rate* $R_{cr}(W_{Y|X})$.

Remark 6.2 For symmetric channels (in the Gallager sense) $E_o(\rho, W_{Y|X})$ is achieved by uniform input distributions and hence it is differentiable with respect to ρ (cf. Section 5.3.3).

Proof of Theorem 6.3: The proof is straightforward since if (6.16) holds then $\overline{E}_{Jsp}$ given in the parametric form (5.8) would be achieved by a $\rho \leq \rho^*$ by noting that $T_{sp}(\rho, W_{Y|X}) = E_o(\rho, W_{Y|X})$ is differentiable, and hence $E_o^{fb}(\rho, W_{Y|X}) = E_o(\rho, W_{Y|X})$ by (6.13) and (6.15) follows. ■

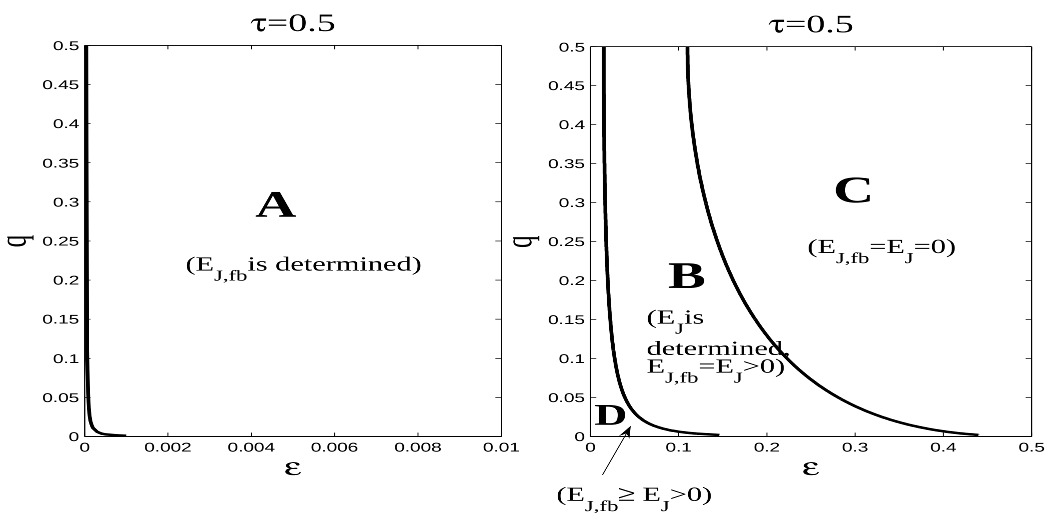

Figure 6.3: The regions for the (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 6.1 with $\tau = 0.5$. $E_{J,fb}$ is exactly determined in Region **A** (including the boundary). Furthermore, in Region **B** (including the boundary with **A**), $E_{J,fb} = E_J > 0$; in Region **C** (including the boundary with **B**), $E_{J,fb} = E_J = 0$. Note that in Region **D**, E_J is not exactly determined.

Example 6.1 Now we apply the conditions (6.16) to a communication system with a binary source with distribution $\{q, 1-q\}$ $(q < 0.5)$, a BSC with crossover probability ε $(\varepsilon < 0.5)$

and transmission rates $\tau = 0.5$. For such channel

$$g_0(x) = g_1(x) = \varepsilon^{\frac{1}{1+x}}(1-\varepsilon)^{\frac{x}{1+x}} + \varepsilon^{\frac{x}{1+x}}(1-\varepsilon)^{\frac{1}{1+x}}.$$

Fig. 6.3 (left) shows that the JSCC error exponent with feedback $E_{J,fb}$ is determined exactly by the sphere-packing upper bound $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ in Region **A** (including the boundary). Comparing with Region B in Fig. 6.3 (right), where E_J is determined by $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$, **A** is much bigger than $\mathbf{B} \cup \mathbf{C}$. In other words, the sphere-packing upper bound determines the JSCC error exponent with feedback for many source-channel pairs in Region **D**, where E_J is not determined there.

6.2 Feedback Can Increase the JSCC Error Exponent

For discrete memoryless source-channel systems, feedback does not improve the region for reliable transmissibility (i.e., we have the same JSCC theorem for systems with feedback), but it might improve the reliability function. In the last section we studied the lower and upper bounds for the JSCC error exponent with perfect feedback $E_{J,fb}$, and the most important result is that we obtain a nontrivial Gallager-type lower bound which can be easily computed when the channel has binary input alphabet and a symmetric distribution. Consequently, we can compare it with the upper bound of E_J and study the situation when $E_{J,fb}$ can be strictly larger than E_J.

From Fig. 6.3, we note that when E_J is determined $E_{J,fb}$ is equal to E_J in Regions **B** and **C**, and $E_{J,fb} > E_J$ is possible only if the source-channel pairs are in Region **D**. In that case, the lower and upper bounds $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ are not equal, and more specifically, the upper bound

$$\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) = \inf_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[te\left(\frac{R}{\tau}, Q_S\right) + E_{sp}(R, W_{Y|X}) \right]$$

would be achieved by some $R \le R_{cr}(W_{Y|X})$. Considering that the sphere-packing channel exponent $E_{sp}(R, W_{Y|X})$ is a loose bound for low rates and that

$$E_J(Q_S, W_{Y|X}, \tau) \le \inf_{\tau H_{Q_S}(S) \le R \le \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X}) \right],$$

we replace $E_{sp}(R, W_{Y|X})$ by the straight-line bound $E_{st}(R, W_{Y|X})$, which leads to a tighter upper bound for E_J

$$E_J(Q_S, W_{Y|X}, \tau) \leq \overline{E}_{Jst}(Q_S, W_{Y|X}, \tau) \triangleq \inf_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e \left(\frac{R}{\tau}, Q_S \right) + E_{st}(R, W_{Y|X}) \right]. \tag{6.17}$$

In the following we will analytically compute the new upper bound $\overline{E}_{Jst}(Q_S, W_{Y|X}, \tau)$ and numerically compare it with the lower bound given of Theorem 6.2 for binary input channels with symmetric distribution.

Since for channels with symmetric distribution $E_{sp}(R, W_{Y|X})$ is differentiable with respect to R, the supporting line of $E_{sp}(R, W_{Y|X})$ at every $R > 0$ is tangent to $E_{sp}(R, W_{Y|X})$ and we have an analytical form for its slope. We know that the supporting line of $-E_{sp}$ at R_l is given by (cf. [17, Section 7.1])

$$-E_{sp}(R_l, W_{Y|X}) = \rho_l R_l - (-E_{sp}(R_l, W_{Y|X}))_* = \rho_l R_l - E_o(\rho_l, W_{Y|X})$$

with slope ρ_l, where the second equality holds since $E_o(\rho_l, W_{Y|X})$ is concave. This means that

$$\sup_{\rho \geq 0} [\rho R_l - E_o(\rho, W_{Y|X})] = \rho_l R_l - E_o(\rho_l, W_{Y|X})$$

and hence $E_o'(\rho_l, W_{Y|X}) = R_l$ as $E_o(\rho_l, W_{Y|X})$ is differentiable.

On the other hand, for binary input channels (which are equidistant channels, see Section 5.4.3), $E_{ex}(0, W_{Y|X})$ is achieved by uniform input distribution and by [42, p. 156]

$$E_{ex}(0, W_{Y|X}) = -\sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{4} \log_2 \left[\sum_{y \in \mathcal{Y}} \sqrt{W_{Y|X}(y|x_1) W_{Y|X}(y|x_2)} \right] < \infty.$$

We can express the straight-line exponent analytically by

$$E_{st}(R, W_{Y|X}) = \begin{cases} E_{ex}(0, W_{Y|X}) - \rho_l R & \text{if } 0 \leq R \leq R_l, \\ E_{sp}(R, W_{Y|X}) & \text{if } R \geq R_l, \end{cases} \tag{6.18}$$

where ρ_l is the unique solution of

$$E_o(\rho_l, W_{Y|X}) = -\sum_{x_1, x_2 \in \mathcal{X}} \frac{1}{4} \log_2 \left[\sum_{y \in \mathcal{Y}} \sqrt{W_{Y|X}(y|x_1) W_{Y|X}(y|x_2)} \right]. \tag{6.19}$$

and $R_l = E_o'(\rho_l, W_{Y|X})$. By the Fenchel duality theorem, it can be shown in a similar manner as Theorem 5.1 that

$$\overline{E}_{Jst}(Q_S, W_{Y|X}, \tau) = \max_{0 \le \rho \le \rho_l} [E_o(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \tag{6.20}$$

We then compare the lower bound of E_{Jfb} with the new upper bound $\overline{E}_{Jst}(Q_S, W_{Y|X}, \tau)$ in the following example.

Example 6.2 Let the source be a binary source with distribution $\{q, 1-q\}$ ($q < 0.5$), and let the channel be a BSC with crossover probability ε ($\varepsilon < 0.5$). We choose small transmission rates $\tau = 0.25$ and 0.4 here since we want to make the infimum of (6.17) achieved by small R so that the lower bound of $E_{J,fb}$ given by Theorem 6.2 is able to outperform the upper bound of E_J computed from (6.20). Indeed, as seen from Fig. 6.4, when $\tau = 0.25$, the lower bound of $E_{J,fb}$ is obviously larger than the upper bound of E_J for $q = 0.1$ and 0.2 and small ε's. When $\tau = 0.4$, the lower bound of $E_{J,fb}$ still has slight advantage over the upper bound of E_J for $\varepsilon \le 0.001$. Thus we have demonstrated that feedback can improve the JSCC reliability function for some discrete memoryless systems.

Remark 6.3 It is known [37], [115] that for certain DMC's (e.g. BSC's) the channel error exponent with feedback is strictly larger than the one without feedback for rates in an interval below the channel critical rate. Our results coincide with the previous results since the upper and lower bounds for the JSCC reliability function reduces to the channel error exponent bounds if the source distribution is uniform.

6.3 Systems with Source Side Information at the Decoder

6.3.1 System Description

We consider in this section a communication system consisting of two correlated DMSs $Q_{SL} \in \mathcal{P}(\mathcal{S} \times \mathcal{L})$ with finite alphabet $\mathcal{S} \times \mathcal{L}$ and joint distribution Q_{SL}, and a DMC $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and transition

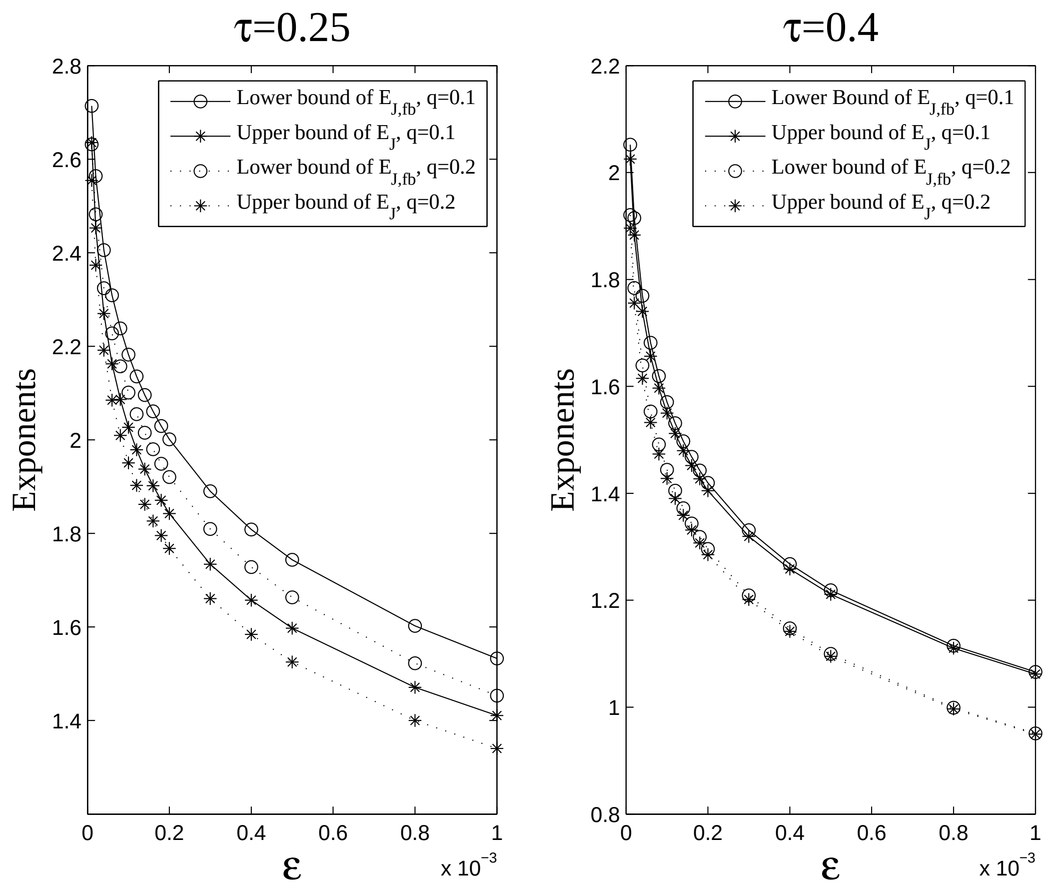

Figure 6.4: Feedback can enlarge the JSCC error exponent for different (ε, q) pairs in the binary DMS $\{q, 1-q\}$ and BSC (ε) system of Example 6.2.

probability distribution $W_{Y|X}$. We need to transmit the source Q_S over the channel $W_{Y|X}$ with side information Q_L available at the decoder only.

The system is depicted in Fig. 6.5. The source message pair $(\mathbf{s}, \mathbf{l})$ of length τn is drawn in an independent and identically distributed (i.i.d.) manner from a joint distribution $Q_{SL} \in \mathcal{P}(\mathcal{S} \times \mathcal{L})$. We need to transmit the source message $\mathbf{s}$ over the DMC $W_{Y|X}$ via JSCC block codes of length n and transmission rate τ. The source message $\mathbf{l}$, viewed as a noisy observation of $\mathbf{s}$, contains the source SI and helps the decoder reconstruct $\mathbf{s}$.

A JSC code of block length n and transmission rate $\tau > 0$ for the system with SI at the

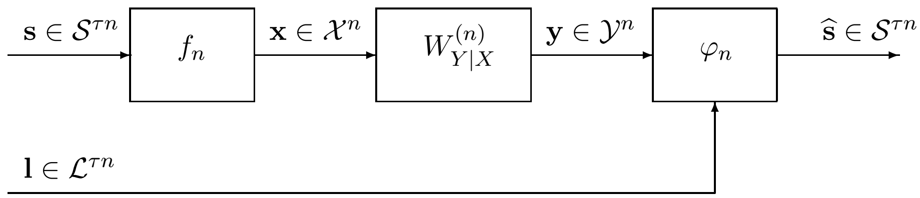

Figure 6.5: JSCC system with source SI.

decoder is a pair of mappings, (f_n, φ_n), where

$$f_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{X}^n$$

is the encoder, and

$$\psi_n : \mathcal{Y}^n \times \mathcal{L}^{\tau n} \longrightarrow \mathcal{S}^{\tau n}$$

is the decoder (see Fig. 6.5). The probability of error is given by

$$P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, \tau) \triangleq \sum_{(\mathbf{s},\mathbf{l}) \in \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n}} Q_{SL}^{(\tau n)}(\mathbf{s}, \mathbf{l}) \sum_{\mathbf{y}: \varphi_n(\mathbf{y},\mathbf{l}) \neq \mathbf{s}} W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})). \tag{6.21}$$

Definition 6.3 Given Q_{SL}, $W_{Y|X}$ and $\tau > 0$, the JSCC error exponent $E_J^{SID}(Q_{SL}, W_{Y|X}, \tau)$ is defined as supremum of the set of all numbers E for which there exists a sequence of JSC codes (f_n, φ_n) with blocklength n and transmission rate τ such that

$$E \leq \liminf_{n \to \infty} -\frac{1}{n} \log_2 P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, \tau). \tag{6.22}$$

6.3.2 A Lower Bound

For the joint distribution of the sources Q_{SL}, we can look at the conditional distribution $Q_{L|S} \in \mathcal{P}(\mathcal{L}|\mathcal{S})$ as a dummy channel between Q_S and Q_L. For any $P_S \in \mathcal{P}(\mathcal{S})$ and any $R \leq H_{P_S}(S)$, we define an exponent for the dummy channel $Q_{L|S}$ by

$$\begin{aligned} e_r(R, Q_{L|S}, P_S) &\triangleq E_r(H_{P_S}(S) - R, Q_{L|S}, P_S) \\ &= \min_{P_{L|S} \in \mathcal{P}(\mathcal{L}|\mathcal{S})} \left[D(P_{L|S} \| Q_{L|S} | P_S) + |R - H_{P_S P_{L|S}}(S|L)|^+ \right]. \end{aligned} \tag{6.23}$$

Recall that $E_r(R, W_{Y|U})$ is a strictly decreasing function of R and vanishes at the channel capacity of $W_{Y|X}$. Accordingly, $e_r(R, Q_{L|S}, P_S)$ is a strictly increasing function of R and is zero if and only if $R \le H_{P_S Q_{L|S}}(S|L)$. Let

$$E_{SI}(P_S, Q_{L|S}, W_{Y|X}) \triangleq \max_{0 \le R \le \tau H_{P_S}(S)} \min \left\{ E_r(R, W_{Y|X}), \tau e_r \left(\frac{R}{\tau}, Q_{L|S}, P_S \right) \right\}. \quad (6.24)$$

For $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^n$, since the joint type $P_{\mathbf{xy}}$ can also be represented as distributions of dummy RV's, we define the *empirical* mutual information and conditional entropy respectively by

$$I(\mathbf{x}; \mathbf{y}) \triangleq I_{P_{\mathbf{xy}}}(X; Y)$$

and

$$H(\mathbf{x}|\mathbf{y}) \triangleq H_{P_{\mathbf{xy}}}(X|Y).$$

To establish a lower bound for $E_J^{SID}(Q_{SL}, W_{Y|X}, \tau)$, we need the following auxiliary result.

Proposition 6.1 [30, Theorem 5] *Given finite sets $\mathcal{X}$ and $\mathcal{Y}$, a sequence of positive integers $\{m_n\}$ with*

$$\frac{1}{n} \log_2 m_n \to 0,$$

for $\delta > 0$, $n = n(\delta, |\mathcal{X}|, |\mathcal{Y}|)$ sufficiently large, arbitrary (not necessarily distinct) types $P_{X_i} \in \mathcal{P}_n(\mathcal{X})$, and positive integers N_i, $i = 1, 2, ..., m_n$ with

$$R_i \triangleq \frac{1}{n} \log_2 N_i < H_{P_{X_i}}(X) - \delta,$$

there exist m_n disjoint subsets $\Omega_i = \left\{ \boldsymbol{x}_p^{(i)} \right\}_{p=1}^{N_i} \subseteq \mathcal{T}_{P_{X_i}}$ for every i, and a mapping $\varphi_n^{(0)} : \mathcal{Y}^n \to \Omega$, where $\Omega \triangleq \bigcup_i \Omega_i$, such that the probability of erroneous transmission of an $\boldsymbol{x} \in \Omega$ using $\varphi_n^{(0)}$ is bounded for $W_{Y|X}$ as

$$\sum_{\boldsymbol{y}: \varphi_n^{(0)}(\boldsymbol{y}) \ne \boldsymbol{x}} W_{Y|X}^{(n)}(\boldsymbol{y}|\boldsymbol{x}) \le 2^{-n\left[E_r\left(R_i, P_{X_i}, W_{Y|X}\right) - \delta\right]} \quad (6.25)$$

if $\boldsymbol{x} \in \Omega_i$ for every i.

Proof: The proof is based on the type packing lemma (here we only need (3.4) of Lemma 3.2, which is Csiszár's type packing lemma [30, Theorem 5] for a single-letter type setting) and a generalized maximum mutual information decoding rule.

In the sequel of the proof, we look at X as the RV A in Lemma 3.2. For the $\{m_n\}$ and P_{X_i} given in Proposition 6.1, according to the first part of Lemma 3.2, there exist pairwise disjoint subsets Ω_i satisfying (3.4) for every $1 \leq i \leq m_n$, $1 \leq p \leq N_i$, $V_{X'|X} \in \mathcal{P}_n(\mathcal{X}|\mathcal{X})$, with the exception of the case that $i = k$ and $V_{X'|X}$ is the conditional distribution such that $V_{X'|X}(x'|x)$ is 1 if $x' = x$ and 0 otherwise. We shall show that for such Ω_i, there exists a mapping $\varphi_n^{(0)}$ such that (6.25) is satisfied.

For any $\mathbf{x} \in \Omega$ and $\mathbf{y} \in \mathcal{Y}^n$, let

$$\alpha(\mathbf{x};\mathbf{y}) \triangleq I(\mathbf{x};\mathbf{y}) - R_i,$$

where $R_i = \frac{1}{n}\log_2 N_i$ if $\mathbf{x} \in \Omega_i$. Define $\varphi_n^{(0)} : \mathcal{Y}^n \to \Omega$ by

$$\varphi_n^{(0)}(\mathbf{y}) \triangleq \arg\max_{\mathbf{x}\in\Omega} \alpha(\mathbf{x};\mathbf{y}).$$

Using the decoder $\varphi_n^{(0)}$, we can upper bound the probability of error (assuming that $\mathbf{x} \in \Omega_i$ is sent through the channel) as follows

$$\begin{aligned} P_e^{(n)}(\mathbf{x}) &= W_{Y|X}^{(n)}\left(\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq \mathbf{x}\right\}\Big|\,\mathbf{x}\right) \\ &\leq \sum_{\widehat{V}_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|P_{X_i})} W_{Y|X}^{(n)}\left(\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq \mathbf{x}\right\}\Big|\,\mathbf{x}\right). \end{aligned} \tag{6.26}$$

Using the identity (Lemma 3.1) that for any $\mathbf{x} \in \Omega_i \subseteq \mathbb{T}_{P_{X_i}}$ and $\mathbf{y} \in \mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x})$

$$W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x}) = 2^{-n\left[D\left(\widehat{V}_{Y|X}\|W_{Y|X}|P_{X_i}\right) + H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X)\right]},$$

we obtain

$$\begin{aligned} &\sum_{\widehat{V}_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|P_{X_i})} W_{Y|X}^{(n)}\left(\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq \mathbf{x}\right\}\Big|\,\mathbf{x}\right) \\ &= \left|\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \underbrace{\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq \mathbf{x}\right\}}_{\triangleq \mathcal{E}}\right| 2^{-n\left[D\left(\widehat{V}_{Y|X}\|W_{Y|X}|P_{X_i}\right) + H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X)\right]}. \end{aligned}$$

Thus we only need to upper bound $\left|\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \mathcal{E}\right|$. If we fix a $k = 1, 2, ..., m_n$, then $\mathcal{E}$ is the set of all $\mathbf{y}$ such that there exist some $\mathbf{x}' \in \Omega_k$, $(\mathbf{x}, \mathbf{x}', \mathbf{y})$ admits a joint type

$P_{\mathbf{x}\mathbf{x}'\mathbf{y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y})$ and

$$I(\mathbf{x}';\mathbf{y}) - R_k \geq I(\mathbf{x};\mathbf{y}) - R_i. \tag{6.27}$$

Note that (6.27) can be represented as for dummy R.V.'s $X \in \mathcal{X}$, $X' \in \mathcal{X}$, and $Y \in \mathcal{Y}$, the following holds under the joint distribution $P_{XX'Y} = P_{\mathbf{x}\mathbf{x}'\mathbf{y}}$,

$$I_{P_{X'Y}}(X';Y) - R_k \geq I_{P_{XY}}(X;Y) - R_i,$$

where $P_{X'Y}$ and P_{XY} are the corresponding marginal distributions induced by $P_{XX'Y}$. Thus, $\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \mathcal{E}$ can be written as a union of subsets

$$\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \mathcal{E} = \bigcup_{k=1}^{m_n} \bigcup_{P_{XX'Y} \in \mathcal{C}_k(\mathbf{x})} \mathcal{F}_k(\mathbf{x}, P_{XX'Y}) \tag{6.28}$$

where

$$\mathcal{C}_k(\mathbf{x}) \triangleq \left\{ \begin{array}{ll} & P_X = P_{X_i}, \quad P_{X'} = P_{X_k} \\ P_{XX'Y} & P_{Y|X} = \widehat{V}_{Y|X}, \\ \in \mathcal{P}_n(\mathcal{X}^2 \times \mathcal{Y}) : & I_{P_{X'Y}}(X';Y) - R_k \\ & \quad \geq I_{P_{XY}}(X;Y) - R_i \end{array} \right\},$$

where P_X, $P_{X'}$ and $P_{Y|X}$, etc, are the corresponding marginal and conditional distributions induced from $P_{XX'Y}$, and

$$\mathcal{F}_k(\mathbf{x}, P_{XX'Y}) \triangleq \left\{ \mathbf{y} : \begin{array}{ll} \exists \quad \mathbf{x}' & (\mathbf{x}, \mathbf{x}', \mathbf{y}) \in \mathbb{T}_{XX'Y} \\ \text{such that} & \mathbf{x}' \in \Omega_k \end{array} \right\},$$

where $\mathbb{T}_{XX'Y} \triangleq \mathbb{T}_{P_{XX'Y}}$. Clearly, given any k, and $P_{XX'Y}$,

$$\begin{aligned} |\mathcal{F}_k(\mathbf{x}, P_{XX'Y})| &\leq \left| \left\{ (\mathbf{x}', \mathbf{y}) : \begin{array}{l} (\mathbf{x}, \mathbf{x}', \mathbf{y}) \in \mathbb{T}_{XX'Y} \\ \mathbf{x}' \in \Omega_k, \quad \mathbf{x}' \neq \mathbf{x} \end{array} \right\} \right| \\ &= \left| \left\{ \mathbf{x}' : \begin{array}{l} (\mathbf{x}, \mathbf{x}') \in \mathbb{T}_{XX'} \\ \mathbf{x}' \in \Omega_k, \quad \mathbf{x}' \neq \mathbf{x} \end{array} \right\} \right| \times \left| \mathbb{T}_{Y|XX'}(\mathbf{x}, \mathbf{x}') \right| \\ &\leq N_k 2^{-n\left[I_{P_{XX'}}(X;X') - \eta\right]} \times 2^{nH_{P_{XX'Y}}(Y|X,X')}, \end{aligned} \tag{6.29}$$

where the last inequality follows from Lemma 3.2 (3.4). Meanwhile, when $\mathbf{x} \in \Omega_i$, the following simple bound also holds

$$|\mathcal{F}_k(\mathbf{x}, P_{XX'Y})| \leq \left|\mathbb{T}_{Y|X}(\mathbf{x})\right| \leq 2^{nH_{P_{XY}}(Y|X)} = 2^{nH_{P_{X_i}\widehat{V}_{Y|X}}(Y|X)} \tag{6.30}$$

since for each $\mathbb{T}_{XX'Y} \in \mathcal{C}_k(\mathbf{x})$, we have $P_X = P_{X_i}$, $P_{Y|X} = \widehat{V}_{Y|X}$ and hence $P_{XY} = P_{X_i}\widehat{V}_{Y|X}$. Now substituting the following inequality

$$\begin{aligned}
& H_{P_{XX'Y}}(Y|X, X') - I_{P_{XX'}}(X; X') \\
= \; & H_{P_{XX'Y}}(X, X', Y) - H_{P_X}(X) - H_{P_{X'}}(X') \\
= \; & H_{P_{XY}}(X, Y) + H_{P_{XX'Y}}(U', X'|X, Y) - H_{P_X}(X) - H_{P_{X'}}(X') \\
= \; & H_{P_{XY}}(Y|X) - I_{P_{XX'Y}}(X'; X, Y) \\
\leq \; & H_{P_{XY}}(Y|X) - I_{P_{X'Y}}(X'; Y) \\
= \; & H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X) - I_{P_{X'Y}}(X'; Y) \qquad (6.31)
\end{aligned}$$

into (6.29), combining with (6.30) together, we obtain

$$|\mathcal{F}_k(\mathbf{x}, P_{XX'Y})| \leq 2^{n\left[H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X) - \left|I_{P_{X'Y}}(X';Y) - R_k\right|^+\right]}. \tag{6.32}$$

Again recall that for $P_{XX'Y} \in \mathcal{C}_k(\mathbf{x})$, $P_{XY} = P_{X_i}\widehat{V}_{Y|X}$, and note that

$$I_{P_{X'Y}}(X'; Y) - R_k \geq I_{P_{XY}}(X; Y) - R_i.$$

This implies when $P_{XX'Y} \in \mathcal{C}_k(\mathbf{x})$

$$|\mathcal{F}_k(\mathbf{x}, P_{XX'Y})| \leq 2^{n\left[H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X) - \left|I_{P_{X_i}\widehat{V}_{Y|X}}(X;Y) - R_i\right|^+\right]},$$

and hence

$$\begin{aligned}
\left|\mathbb{T}_{\widehat{V}_{Y|X}}(\mathbf{x}) \bigcap \mathcal{E}\right| \leq \; & m_n (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|} \\
& \times 2^{n\left[H_{P_{X_i}\widehat{V}_{Y|X}}(Y|X) - \left|(I_{P_{X_i}\widehat{V}_{Y|X}}(X;Y) - R_i\right|^+\right]},
\end{aligned} \tag{6.33}$$

since by Lemma 3.1

$$|\mathcal{C}_k(\mathbf{x})| \leq \left|\mathcal{P}_n(\mathcal{X}^2 \times \mathcal{Y})\right| \leq (n+1)^{|\mathcal{X}|^2|\mathcal{Y}|}.$$

Finally, we complete the proof by plugging (6.33) into (6.26) and by noting that $|\mathcal{P}_n(\mathcal{Y}|P_{X_i})|$ is also a polynomial function of n by Lemma 3.1. ■

Theorem 6.4 *Given Q_{SL}, $W_{Y|X}$ and $\tau > 0$, when the SI Q_L is available only at the decoder, the JSCC error exponent satisfies*

$$E_J^{SID}(Q_{SL}, W_{Y|X}, \tau) \geq \underline{E}_J^{SID}(Q_{SL}, W_{Y|X}, \tau) \triangleq \min_{P_S \in \mathcal{P}(\mathcal{S})} \left[\tau D(P_S \| Q_S) + E_r^*(P_S, W_{Y|X})\right], \tag{6.34}$$

where

$$E_r^*(P_S, W_{Y|X}) = \max\left\{E_r(\tau H_{P_S}(S), W_{Y|X}), E_{SI}(P_S, Q_{L|S}, W_{Y|X})\right\}.$$

Remark 6.4 Note that if sources Q_S and Q_L are independent, i.e., $Q_{SL} = Q_S Q_L$, then the lower bound (6.34) reduces to (5.5) as expected.

Proof:

Step 1: Outline of Proof

We employ a two-stage encoding two-stage decoding rule by combining the approaches of [30] and [73] to show the existence of JSC codes (f_n, φ_n) such that for any $\delta > 0$,

$$P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, \tau) \leq 2^{-n\left[\underline{E}_J^{SID}(Q_{SL}, W_{Y|X}, \tau) - \delta\right]}$$

for $n = n(\delta, |\mathcal{X}|, |\mathcal{Y}|)$ sufficiently large. In the first stage coding, there are two coding schemes for the encoder to choose based on the type $P_{\mathbf{s}}$ of the source word $\mathbf{s}$.

(a) If $\tau H_{P_{\mathbf{s}}}(S) > \widetilde{R}_i$, then encode $\mathbf{s}$ using a unique index in $\{1, 2, 3, ..., \lceil 2^{n\widetilde{R}_i} \rceil\}$,where $\widetilde{R}_i$ achieves the error exponent E_{SI} for the fixed type $P_{\mathbf{s}}$, i.e., the cross point of the channel coding error exponent and the source coding error exponent with source SI.

(b) If $\tau H_{P_{\mathbf{s}}}(S) = \widetilde{R}_i$, then map each $\mathbf{s}$ to an index in a one-to-one manner.

In the second stage coding, we employ Csiszár's JSCC scheme [30].

Step 2: First Stage Encoding

Set $m_n \triangleq |\mathcal{P}_{tn}(\mathcal{S})|$. Note that m_n here is polynomial of n and hence $\frac{1}{n}\log_2 m_n \to 0$. For each type $P_{S_i} \in \mathcal{P}_{tn}(\mathcal{S})$, $i = 1, 2, ..., m_n$, let $\mathcal{N}_i \triangleq \{1, 2, ..., N_i\}$ where $N_i = \lceil 2^{n\widetilde{R}_i} \rceil$ and $\widetilde{R}_i$ achieves $E_{SI}(P_{S_i}, Q_{L|S}, W_{Y|X})$ defined in (6.24), i.e., $\widetilde{R}_i$ is the intersection of $E_r(R, W_{Y|X})$ and $\tau e_r\left(\frac{R}{\tau}, Q_{L|S}, P_{S_i}\right)$ in the domain $[0, \tau H_{P_{S_i}}(S)]$ if any; in that case

$$E_{SI}(P_{S_i}, Q_{L|S}, W_{Y|X}) = E_r(\widetilde{R}_i, W_{Y|X}) = \tau e_r\left(\frac{\widetilde{R}_i}{\tau}, Q_{L|S}, P_{S_i}\right) \geq E_r(\tau H_{P_{S_i}}(S), W_{Y|X}), \tag{6.35}$$

where the last inequality holds since $E_r(R, W_{Y|X})$ is a decreasing function of R. By definition, we thus obtain that

$$E_r^*(P_{S_i}, W_{Y|X}) = E_{SI}(P_{S_i}, Q_{L|S}, W_{Y|X}). \tag{6.36}$$

If there is no intersection between $E_r(R, W_{Y|X})$ and $\tau e_r\left(\frac{R}{\tau}, Q_{L|S}, P_{S_i}\right)$ in the domain $R \leq \tau H_{P_{S_i}}(S)$, the maximum in (6.24) would be achieved by $\widetilde{R}_i = \tau H_{P_{S_i}}(S)$. In that case

$$E_r^*(P_{S_i}, W_{Y|X}) = E_r(\tau H_{P_{S_i}}(S), W_{Y|U}) \geq E_{SI}(P_{S_i}, Q_{L|S}, W_{Y|X}). \tag{6.37}$$

Lemma 6.1 *[32, p. 264, Problem 5], [43]* If $\widetilde{R}_i < \tau H_{P_{S_i}}(S)$, then there exists a pair of mappings $\widehat{f}_{n,i}^{(1)} : \mathcal{T}_{P_{S_i}} \to \mathcal{N}_i$ and $\widehat{\varphi}_{n,i}^{(1)} : \mathcal{N}_i \times \mathcal{L}^{\tau n} \to \mathcal{T}_{P_{S_i}}$ such that for $\delta > 0$,

$$\sum_{\mathbf{l}: \widehat{\varphi}_{n,i}^{(1)}\left(\widehat{f}_{n,i}^{(1)}(\mathbf{s}), \mathbf{l}\right) \neq \mathbf{s}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \leq 2^{-\tau\left[e_r\left(\frac{\widetilde{R}_i}{\tau}, Q_{L|S}, P_{S_i}\right) - \delta\right]} \tag{6.38}$$

if $\mathbf{s} \in \mathcal{T}_{S_i}$ for n sufficiently large.

Proof: Let $\mathbf{s} \in \mathcal{T}_{S_i}$ be the transmitted source message. Write

$$\sum_{\mathbf{l}: \widehat{\varphi}_{n,i}^{(1)}\left(\widehat{f}_{n,i}^{(1)}(\mathbf{s}), \mathbf{l}\right) \neq \mathbf{s}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) = \sum_{P_{L|S} \in \mathcal{P}_n(\mathcal{L}|P_{S_i})} \sum_{\mathbf{l} \in \mathcal{T}_{P_{L|S}}(\mathbf{s})} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \, \mathbb{1}\left\{\widehat{\varphi}_{n,i}^{(1)}\left(\widehat{f}_{n,i}^{(1)}(\mathbf{s}), \mathbf{l}\right) \neq \mathbf{s}\right\}. \tag{6.39}$$

To show the existence of mappings $(\widehat{f}_{n,i}^{(1)}, \widehat{\varphi}_{n,i}^{(1)})$ satisfying (6.38), we consider a family of mapping pairs $(f_{n,i}^{(1)}, \varphi_{n,i}^{(1)})$ as follows. $f_{n,i}^{(1)} : \mathcal{T}_{P_{S_i}} \to \mathcal{N}_i$ is a random binning function that maps each $\mathbf{s} \in \mathcal{T}_{P_{S_i}}$ to an index $w \in \mathcal{N}_i$ with probability $1/N_i$. $\varphi_{n,i}^{(1)} : \mathcal{N}_i \times \mathcal{L}^{\tau n} \to \mathcal{T}_{P_{S_i}}$ is a minimum conditional entropy decoder which searches all the source sequences $\mathbf{s} \in \mathcal{T}_{P_{S_i}}$ such that the empirical conditional entropy $H(\mathbf{s}|\mathbf{l})$ is minimized,

$$\varphi_{n,i}^{(1)}(w, \mathbf{l}) = \arg \min_{\mathbf{s} \in \mathcal{T}_{P_{S_i}} : f_{n,i}^{(1)}(\mathbf{s}) = w} H(\mathbf{s}|\mathbf{l}).$$

We then bound the above probability averaged over all possible pairs $(f_{n,i}^{(1)}, \varphi_{n,i}^{(1)})$,

$$\begin{aligned} &\mathbb{E}\left[\sum_{\mathbf{l}:\varphi_{n,i}^{(1)}\left(f_{n,i}^{(1)}(\mathbf{s}),\mathbf{l}\right)\neq \mathbf{s}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s})\right] \\ &= \sum_{P_{L|S} \in \mathcal{P}_n(\mathcal{L}|P_{S_i})} \sum_{\mathbf{l} \in \mathcal{T}_{P_{L|S}}(\mathbf{s})} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s})\, \mathbb{E}\left[\mathbb{1}\left\{\varphi_{n,i}^{(1)}\left(f_{n,i}^{(1)}(\mathbf{s}), \mathbf{l}\right) \neq \mathbf{s}\right\}\right], \end{aligned}$$

where the expectation is taken with respect to the distribution $P_W(w) = 1/N_i$, $w \in \mathcal{N}_i$. Let A_0 be the event that source messages $\mathbf{s}$ and $\mathbf{l}$ were transmitted. According to the minimum conditional entropy decoding rule

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}\left\{\varphi_{n,i}^{(1)}\left(f_{n,i}^{(1)}(\mathbf{s}), \mathbf{l}\right) \neq \mathbf{s}\right\}\right] \\ &= \Pr\left(\left\{\exists \quad \widehat{\mathbf{s}} \in \mathcal{T}_{S_i}, \widehat{\mathbf{s}} \neq \mathbf{s} \text{ such that } H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l}), f_{n,i}^{(1)}(\mathbf{s}) = f_{n,i}^{(1)}(\widehat{\mathbf{s}})\right\} \Big| A_0\right) \\ &\leq \sum_{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i}: \widehat{\mathbf{s}} \neq \mathbf{s}, H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l})} \Pr\left(\left\{f_{n,i}^{(1)}(\mathbf{s}) = f_{n,i}^{(1)}(\widehat{\mathbf{s}})\right\} \Big| A_0, \widehat{\mathbf{s}}\right) \\ &= \sum_{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i}: \widehat{\mathbf{s}} \neq \mathbf{s}, H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l})} \sum_{w=1}^{N_i} P_W(f_{n,i}^{(1)}(\mathbf{s}) = w) P_W(f_{n,i}^{(1)}(\widehat{\mathbf{s}}) = w) \\ &= \sum_{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i}: \widehat{\mathbf{s}} \neq \mathbf{s}, H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l})} \frac{1}{N_i} \\ &\leq \frac{|\{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i} : H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l})\}|}{N_i}. \end{aligned} \tag{6.40}$$

It follows from the method of types (cf. Lemma 3.1) that

$$\begin{aligned} &|\{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i} : H(\widehat{\mathbf{s}}|\mathbf{l}) \leq H(\mathbf{s}|\mathbf{l})\}| \\ &= \left| \bigcup_{P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L}): H_{P_{SL}}(S|L) \leq H(\mathbf{s}|\mathbf{l})} \{\widehat{\mathbf{s}} \in \mathcal{T}_{S_i} : (\widehat{\mathbf{s}}, \mathbf{l}) \in \mathcal{T}_{P_{SL}}\} \right| \\ &\leq \sum_{P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L}): H_{P_{SL}}(S|L) \leq H(\mathbf{s}|\mathbf{l})} 2^{\tau n H_{P_{SL}}(S|L)} \\ &\leq (\tau n + 1)^{|\mathcal{S}||\mathcal{L}|} 2^{\tau n H(\mathbf{s}|\mathbf{l})}. \end{aligned} \tag{6.41}$$

Plugging (6.41) into (6.40), and noting that the expectation should be no greater than 1, we obtain

$$\mathbb{E}\left[\mathbb{1}\left\{\varphi_{n,i}^{(1)}\left(f_{n,i}^{(1)}(\mathbf{s}),\mathbf{l}\right)\neq\mathbf{s}\right\}\right]\leq(\tau n+1)^{|\mathcal{S}||\mathcal{L}|}2^{-\tau n|\widetilde{R}_i/t-H(\mathbf{s}|\mathbf{l})|^+}.$$

Also, by Lemma 3.1,

$$Q_{L|S}^{(\tau n)}\left(\mathcal{T}_{P_{L|S}}(\mathbf{s})\Big|\,\mathbf{s}\right)\leq 2^{-\tau nD(P_{L|S}\|Q_{L|S}|P_{S_i})}.$$

Thus

$$\mathbb{E}\left[\sum_{\mathbf{l}:\varphi_{n,i}^{(1)}\left(f_{n,i}^{(1)}(\mathbf{s}),\mathbf{l}\right)\neq\mathbf{s}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s})\right]$$
$$\leq\quad(\tau n+1)^{2|\mathcal{S}||\mathcal{L}|}2^{-\tau n\min_{P_{L|S}}\left[D(P_{L|S}\|Q_{L|S}|P_{S_i})+|R_i/t-H_{P_{S_i}P_{L|S}}(S|L)|^+\right]}.$$

This implies that there exists a pair of mappings $(\widehat{f}_{n,i}^{(1)},\widehat{\varphi}_{n,i}^{(1)})$ such that for any $\mathbf{s}\in\mathcal{T}_{S_i}$ the upper bound (6.38) holds. ■

We next define the first encoding function $\widetilde{f}_n^{(1)}=\left\{\widetilde{f}_{n,i}^{(1)}\right\}_{i=1}^{m_n}:\mathcal{S}^{\tau n}\to\bigcup_{i=1}^{m_n}\mathcal{N}_i$ as follows. For every $i=1,2,...,m_n$, and the choice of $\mathcal{N}_i$,

- if $\widetilde{R}_i=\tau H_{P_{S_i}}(S)$, then $\widetilde{f}_{n,i}^{(1)}:\mathcal{T}_{P_{S_i}}\to\mathcal{N}_i$ maps each $\mathbf{s}\in\mathcal{T}_{P_{S_i}}$ to a unique index $w\in\mathcal{N}_i$, since $|\mathcal{T}_{P_{S_i}}|\leq 2^{nH_{P_{S_i}}(S)}$.
- if $\widetilde{R}_i<\tau H_{P_{S_i}}(S)$, then let $\widetilde{f}_{n,i}^{(1)}=\widehat{f}_{n,i}^{(1)}$.

Note that the decoding function $\widehat{\varphi}_{n,i}^{(1)}$ corresponding to each of the above cases will be used in the decoding stage.

Step 3: Second Stage Encoding

For the index set $\mathcal{N}_i=\{1,2,...,N_i\}$ with $R_i\triangleq\frac{1}{n}\log_2 N_i$ (note that $R_i\to\widetilde{R}_i$ as $n\to\infty$), let $P_{X_i}^*\in\mathcal{P}_n(\mathcal{X})$ be a type[3] that maximizes the exponent $E_r(R_i,W_{Y|X})=E_r(R_i,W_{Y|X},P_{X_i}^*)$ for each $i=1,2,...,m_n$. Assume (without loss of generality) that $R_i<H_{P_{X_i}^*}(X)-\delta$ is

[3]Note that the maximizer can be approximated (with arbitrarily high accuracy) by a type.

satisfied for $i = 1, 2, ..., m'_n$ (where $m'_n \leq m_n$). Then for the m'_n types $P^*_{X_i}$'s, according to Proposition 6.1, there exist disjoint sets $\Omega_i \in \mathcal{T}_{P^*_{X_i}}$ with $|\Omega_i| = N_i$, $i = 1, 2, ..., m'_n$ and a mapping $\varphi_n^{(0)}$ such that (9.15) holds. Now for each $i = 1, 2, ..., m_n$, define the second encoding function $f_n^{(2)} = \left\{ f_{n,i}^{(2)} \right\}_{i=1}^{m_n} : \bigcup_{i=1}^{m_n} \mathcal{N}_i \to \mathcal{X}^n$ as follows.

- If $i \leq m'_n$, i.e., if $R_i < H_{P^*_{X_i}}(X) - \delta$ is satisfied for such i, then $f_{n,i}^{(2)}$ maps each $w \in \mathcal{N}_i$ to a unique codeword $\mathbf{x} \in \Omega_i$.

- If $i > m'_n$, i.e., if $R_i \geq H_{P^*_{X_i}}(X) - \delta$ holds for such i, then let $f_{n,i}^{(2)}(w) = \mathbf{0} \in \mathcal{X}^n$ for every $w \in \mathcal{N}_i$ (assume without loss of generality that $\mathbf{0} \notin \Omega$), and an error is declared.

Step 4: First Stage Decoding

Based upon received n-length sequence $\mathbf{y}$ at the channel output, we first employ $\varphi_n^{(0)} : \mathcal{Y}^n \to \Omega$ to estimate the codeword $\mathbf{x}$. According to Proposition 6.1, there exist such function $\varphi_n^{(0)}$ such that if the transmitted index $\mathbf{x} \in \Omega_i$, $i = 1, 2, ..., m'_n$, then the probability of error is bounded by

$$\sum_{\mathbf{y}:\varphi_n^{(0)}(\mathbf{y}) \neq \mathbf{x}} W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x}) \leq 2^{-n\left[E_r\left(R_i, W_{Y|X}, P^*_{X_i}\right) - \delta\right]} = 2^{-n\left[E_r\left(R_i, W_{Y|X}\right) - \delta\right]}.$$

If the transmitted codeword $\mathbf{x} \notin \Omega$, then the probability of error is bounded by 1. We then use $\varphi_n^{(0)}$ to define the decoding function $\varphi_n^{(2)} : \mathcal{Y}^n \to \bigcup_{i=1}^{m_n} \mathcal{N}_i$ as follows. Let $\varphi_n^{(0)}(\mathbf{y}) = \widehat{\mathbf{x}} \in \Omega_i$, then the decoder $\varphi_n^{(2)}$ outputs the index $\widehat{w} \in \mathcal{N}_i$ such that $f_{n,i}^{(2)}(\widehat{w}) = \widehat{\mathbf{x}}$.

Step 5: Second Stage Decoding

Let $\widehat{w} \in \mathcal{N}_i$ be the output of $\varphi_n^{(2)}$. Define the decoding function $\widetilde{\varphi}_n^{(1)} = \{\widetilde{\varphi}_{n,i}^{(1)}\}_{i=1}^{m_n} : \bigcup_{i=1}^{m_n} \mathcal{N}_i \times \mathcal{L}^{\tau n} \to \mathcal{S}^{\tau n}$ as follows.

- If $\widetilde{R}_i = \tau H_{P_{S_i}}(S)$, then $\widetilde{\varphi}_{n,i}^{(1)} : \mathcal{N}_i \times \mathcal{L}^{\tau n} \to \mathcal{T}_{P_{S_i}}$ outputs the unique $\mathbf{s}$ such that $\widetilde{f}_{n,i}^{(1)}(\mathbf{s}) = \widehat{w}$.

- If $\widetilde{R}_i < \tau H_{P_{S_i}}(S)$, then set $\widetilde{\varphi}_{n,i}^{(1)} = \widehat{\varphi}_{n,i}^{(1)}$, where $\widehat{\varphi}_{n,i}^{(1)}$ is the minimum conditional entropy decoder corresponding to $\widetilde{f}_{n,i}^{(1)}$ used in the first stage decoding.

Step 6: Analysis of the Probability of Error

Let $f_n(\mathbf{s}) \triangleq f_n^{(2)}(\widetilde{f}_n^{(1)}(\mathbf{s}))$ and $\varphi_n(\mathbf{y},\mathbf{l}) \triangleq \widetilde{\varphi}_n^{(1)}(\varphi_n^{(2)}(\mathbf{y}),\mathbf{l})$. Rewrite the probability of error (6.21) as

$$P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, t) = \sum_{P_{S_i}\in\mathcal{P}_n(\mathcal{S})} \sum_{\mathbf{s}\in\mathcal{T}_{P_{S_i}}} Q_S^{(\tau n)}(\mathbf{s}) \sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \sum_{\mathbf{y}:\varphi_n(\mathbf{y},\mathbf{l})\neq\mathbf{s}} W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})). \tag{6.42}$$

In the following we bound

$$P(\mathbf{s}) \triangleq \sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \sum_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))\mathbb{1}\left\{\varphi_n(\mathbf{y},\mathbf{l}) \neq \mathbf{s}\right\}$$

assuming that the transmitted source messages is $\mathbf{s} \in \mathcal{T}_{P_{S_i}}$. There are two cases to consider.

Case 1: If $\widetilde{R}_i = \tau H_{P_{S_i}}(S)$, then

$$P(\mathbf{s}) = \sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \sum_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w))\mathbb{1}\left\{\varphi_n^{(2)}(\mathbf{y}) \neq w\right\} = \sum_{\mathbf{y}:\varphi_n^{(2)}(\mathbf{y})\neq w} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w)), \tag{6.43}$$

where $w = \widetilde{f}_n^{(1)}(\mathbf{s})$. If $f_n^{(2)}(w) \in \Omega$ (i.e., $R_i < H_{P_{X_i}^*}(X) - \delta$), it follows from Step 3 that

$$\sum_{\mathbf{y}:\varphi_n^{(2)}(\mathbf{y})\neq w} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w)) \leq 2^{-n\left[E_r\left(R_i, W_{Y|X}\right)-\delta\right]}$$

for n sufficiently large. If $f_n^{(2)}(w) \notin \Omega$ (i.e., $R_i \geq H_{P_{X_i}^*}(X) - \delta$), the above bound trivially holds since $R_i \geq H_{P_{X_i}^*}(X) - \delta$ yields

$$E_r\left(R_i, W_{Y|X}\right) = E_r\left(R_i, W_{Y|X}, P_{X_i}^*\right) \leq |I_{P_{X_i}^* W_{Y|X}}(X;Y) - R_i|^+ \leq \delta. \tag{6.44}$$

Above all, we can bound for n sufficiently large,

$$P(\mathbf{s}) \leq 2^{-n\left[E_r\left(R_i, W_{Y|X}\right)-\delta\right]} \leq 2^{-n\left[E_r\left(\widetilde{R}_i, W_{Y|X}\right)-2\delta\right]} = 2^{-n\left[E_r^*\left(P_{S_i}, W_{Y|X}\right)-2\delta\right]} \tag{6.45}$$

if $\mathbf{s} \in \mathcal{T}_{S_i}$, where the second inequality holds since $R_i = \frac{1}{n}\log_2\lceil 2^{n\widetilde{R}_i}\rceil$ can be arbitrarily close to $\widetilde{R}_i$ as n goes to infinity, and the last equality follows from (6.35)–(6.37).

Case 2: If $\widetilde{R}_i < \tau H_{P_{S_i}}(S)$, then

$$\begin{aligned} P(\mathbf{s}) &\le \sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s}) \sum_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w)) \\ &\quad \left[\mathbb{1}\left\{\varphi_n^{(2)}(\mathbf{y}) \neq w\right\} + \mathbb{1}\left\{\varphi_n^{(2)}(\mathbf{y}) = w\right\}\mathbb{1}\left\{\widetilde{\varphi}_n^{(1)}\left(\varphi_n^{(2)}(\mathbf{y}),\mathbf{l}\right) \neq \mathbf{s}\right\}\right] \\ &\le \sum_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w))\mathbb{1}\left\{\varphi_n^{(2)}(\mathbf{y}) \neq w\right\} + \sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s})\mathbb{1}\left\{\widetilde{\varphi}_n^{(1)}(w,\mathbf{l}) \neq \mathbf{s}\right\} \end{aligned}$$

where $w = \widetilde{f}_n^{(1)}(\mathbf{s})$. As in Case 1, for any $\mathbf{s} \in \mathcal{T}_{S_i}$, we can bound, for $\delta' = \delta - \frac{1}{2n}$,

$$\begin{aligned} \sum_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|f_n^{(2)}(w))\mathbb{1}\left\{\varphi_n^{(2)}(\mathbf{y}) \neq w\right\} &\le 2^{-n\left[E_r\left(R_i, W_{Y|X}\right)-\delta'\right]} \\ &\le 2^{-n\left[E_r\left(\widetilde{R}_i, W_{Y|X}\right)-2\delta'\right]} \\ &= \frac{1}{2}2^{-n\left[E_r^*\left(P_{S_i}, W_{Y|X}\right)-2\delta\right]} \end{aligned}$$

for n sufficiently large, where the last equality follows from (6.35) and (6.36). On the other hand, it follows from Lemma 6.1, (6.35), and (6.36) that, for $\delta'' = \delta - \frac{1}{\tau n}$,

$$\sum_{\mathbf{l}\in\mathcal{L}^{\tau n}} Q_{L|S}^{(\tau n)}(\mathbf{l}|\mathbf{s})\mathbb{1}\left\{\widetilde{\varphi}_n^{(1)}(w,\mathbf{l}) \neq \mathbf{s}\right\} \le 2^{-\tau n\left[e_r\left(\frac{\widetilde{R}_i}{\tau}, Q_{L|S}, P_S\right)-\delta''\right]} = \frac{1}{2}2^{-n\left[E_r^*\left(P_{S_i}, W_{Y|X}\right)-2\delta\right]}$$

for n sufficiently large. Therefore, we can also bound

$$P(\mathbf{s}) \le 2^{-n\left[E_r^*\left(P_{S_i}, W_{Y|X}\right)-2\delta\right]} \tag{6.46}$$

for n sufficiently large. Substituting (6.45) and (6.46) into (6.42) we obtain

$$P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, t) \le \sum_{P_{S_i}\in\mathcal{P}_n(\mathcal{S})} \sum_{\mathbf{s}\in\mathcal{T}_{P_{S_i}}} Q_S^{(\tau n)}(\mathbf{s}) 2^{-n\left[E_r^*\left(P_{S_i}, W_{Y|X}\right)-2\delta\right]} \tag{6.47}$$

for n sufficiently large. Since $|\mathcal{P}_n(\mathcal{S})| \le (\tau n+1)^{|\mathcal{S}|}$, $Q_S^{(\tau n)}(\mathcal{T}_{P_{S_i}}) \le 2^{-\tau n D(P_{S_i}\|Q_S)}$, we obtain that for such JSC codes

$$P_{e,n}^{SID}(Q_{SL}, W_{Y|X}, t) \le 2^{-n\left\{\min_{P_S\in\mathcal{P}(\mathcal{S})}\left[\tau D(P_S\|Q_S)+E_r^*(P_S, W_{Y|X})\right]-2\delta\right\}}$$

for n sufficiently large. ■

Observation 6.2 Note that the above proof generalize the one of Csiszár's [30, Theorem 3] for the JSCC lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ given in (5.5). In Csiszár's proof, R_i in the second-stage encoding is simply chosen to be $\tau H_{P_{S_i}}(S)$, and there is no first-stage encoding and second-stage decoding. Thus Theorem 6.4 applies to the JSCC system without any SI. In fact, if the source SI is independent of the transmitted source, i.e., if we can write $Q_{SL}(s,l) = Q_S(s)Q_L(l)$ for any $s \in \mathcal{S}$ and $l \in \mathcal{L}$, then $e_r(R, Q_{L|S}, P_S)$ is zero identically. and hence $E_{SI}(P_S, Q_{L|S}, W_{Y|X}) = E_r(H_{P_S}(S), W_{Y|X})$. It then follows that

$$\begin{aligned}\underline{E}_J^{SID}(Q_{SL}, W_{Y|X}, \tau) &= \min_{P_S \in \mathcal{P}(\mathcal{S})} \left[\tau D(P_S \| Q_S) + E_r(H_{P_S}(S), W_{Y|X})\right] \\ &= \min_R \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_r(R, W_{Y|X})\right] = \underline{E}_{Jr}(Q_S, W_{Y|X}, \tau).\end{aligned}$$

One may ask what happens if the source SI Q_L is available at both the encoder and decoder; do we have a lower and/or upper bound for the JSCC system? The answer is yes. In fact, when Q_L is available at both the encoder and decoder, the encoding function f_n is a mapping

$$f_n : \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n} \longrightarrow \mathcal{X}^n,$$

and the probability of error is given by

$$P_{e,n}^{SIED}(Q_{SL}, W_{Y|X}, \tau) \triangleq \sum_{(\mathbf{s},\mathbf{l}) \in \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n}} Q_{SL}^{(\tau n)}(\mathbf{s}, \mathbf{l}) \sum_{\mathbf{y}: \varphi_n(\mathbf{y}, \mathbf{l}) \neq \mathbf{s}} W_{Y|X}^{(n)}(\mathbf{y} | f_n(\mathbf{s}, \mathbf{l})). \tag{6.48}$$

We can write

$$P_{e,n}^{SIED}(Q_{SL}, W_{Y|X}, \tau) = \sum_{\mathbf{l} \in \mathcal{L}^{\tau n}} Q_L^{(n)}(\mathbf{l}) P_{e,n}^{SIED}(Q_{SL}, W_{Y|X}, \tau, \mathbf{l})$$

where

$$P_{e,n}^{SIED}(Q_{SL}, W_{Y|X}, \tau, \mathbf{l}) = \sum_{\mathbf{s} \in \mathcal{S}^{\tau n}} Q_{S|L}^{(\tau n)}(\mathbf{s}|\mathbf{l}) \sum_{\mathbf{y}: \varphi_n(\mathbf{y}, \mathbf{l}) \neq \mathbf{s}} W_{Y|X}^{(n)}(\mathbf{y} | f_n(\mathbf{s}, \mathbf{l}))$$

can be interpreted as the JSCC conditional probability of error given the SI $\mathbf{l}$ (cf. Section 5.1). Now simply applying the results of Chapter 5 and using a type counting argument as in the proof of Theorem 6.1, it is easy to show that the error exponent is lower bounded by

$$\min_{P_{SL}} [\tau D(P_{SL} \| Q_{SL}) + E_r(\tau H_{P_{SL}}(S|L), W_{Y|X})]$$

and upper bounded by

$$\inf_{P_{SL}} [\tau D(P_{SL} \| Q_{SL}) + E_{sp}(\tau H_{P_{SL}}(S|L), W_{Y|X})],$$

where the bounds coincide if the minimum (or infimum) is achieved by some P_{SL} such that $\tau H_{P_{SL}}(S|L)$ is greater than the critical rate of the DMC, hence determining the exponent exactly. So the case when source SI is available at both the encoder and decoder can be viewed as an easy consequence of the results of Chapter 5 and is of less interest in this chapter. Finally, note that we do not yet know whether the availability of the source SI at the encoder would yield a strictly larger JSCC error exponent. To answer this question, we may need to establish an upper bound for E_J^{SID}; this may be considered in future research. Another interesting problem for future research is whether the lower bound $\underline{E}_J^{SID}$ still holds if the random-coding channel exponent $E_r(\tau H_{P_S}(S), W_{Y|X})$ is replaced by the expurgated channel exponent $E_{ex}(\tau H_{P_S}(S), W_{Y|X})$.

6.4 JSCC Theorem for Systems with Source Side Information

By examining the sufficient condition for the positivity of the lower bound $\underline{E}_J^{SID}$, we obtain a sufficient condition for which the source Q_S can be reliably transmitted over the channel. We also can prove a necessary condition using Fano's inequality [29], and thus complete the JSCC theorem.

Theorem 6.5 *[86]* (JSCC Theorem with source SI) Given Q_{SL}, $W_{Y|X}$ and $\tau > 0$, when SI Q_L is available either at only the decoder, we have the following conditions.

(a) The source Q_S can be transmitted over the channel $W_{Y|X}$ with an arbitrarily small probability of error if $tH_{Q_{SL}}(S|L) < C(W_{Y|X})$, where $C(W_{Y|X})$ is the channel capacity of $W_{Y|X}$.

(b) Conversely, if the source Q_S can be transmitted over the channel $W_{Y|X}$ with an arbitrarily small probability of error, then $tH_{Q_{SL}}(S|L) \leq C(W_{Y|X})$.

We point out that the JSCC theorem for systems with source SI available at decoder was initially obtained in [86], and a lossy version JSCC theorem was established in [87]. It was further extended to JSCC systems allowing a Gel'fand-Pinsker channel (rather than a DMC) in [69], whose state is known to the encoder. We still give a proof here for the sake of completeness.

Proof of Theorem 6.5:

Forward Part: Since $E_r(\tau H_{P_S}(S), W_{Y|U})$ is positive if and only if $\tau H_{P_S}(S) < C(W_{Y|X})$, and $E_{SI}(P_S, Q_{L|S}, W_{Y|U})$ is positive if and only if $H_{P_S Q_{L|S}}(S|L) < C(W_{Y|X})$ by definition, it immediately follows that $E_r^*(P_S, W_{Y|X})$ is positive if and only if $H_{P_S Q_{L|S}}(S|L) < C(W_{Y|X})$ since $H_{P_S Q_{L|S}}(S|L) < H_{P_S}(S)$. Now if $\underline{E}_J^{SID}$ given by (6.34) is achieved by a P_S not equal to Q_S, $\underline{E}_J^{SID}$ must be positive. If $\underline{E}_J^{SID}$ is achieved by a $P_S = Q_S$, and additionally if $\tau H_{Q_{SL}}(S|L) < C(W_{Y|X})$, then $\underline{E}_J^{SID} = E_r^*(Q_S, W_{Y|X}) > 0$. Above all, we see that the lower bound $\underline{E}_J^{SID}$ given by (6.34) is positive if $\tau H_{Q_{SL}}(S|L) < C(W_{Y|X})$. The forward part follows.

Converse Part: Set $k = \tau n$. Fano's inequality gives

$$H(S^k|L^k, Y^n) \leq P_{e,n}^{SID} \log_2 |\mathcal{S}^k| + H(P_{e,n}^{SID}) \triangleq n\varepsilon_n.$$

Clearly, $P_{e,n}^{SID} \to 0$ implies that $\varepsilon_n \to 0$ (as $n \to \infty$). It then follows that

$$\begin{aligned}
kH(S|L) &= H(S^k|L^k) \\
&= I(S^k; Y^n|L^k) + H(S^k|L^k, Y^n) \\
&\overset{(a)}{\leq} I(X^n; Y^n|L^k) + n\varepsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} I(X^n; Y_i|L^k, Y^{i-1}) + n\varepsilon_n \\
&= \sum_{i=1}^{n} [H(Y_i|L^k, Y^{i-1}) - H(Y_i|L^k, Y^{i-1}, X^n)] + n\varepsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} [H(Y_i) - H(Y_i|X_i)] + n\varepsilon_n \\
&= \sum_{i=1}^{n} I(X_i; Y_i) + n\varepsilon_n \\
&\leq nC(W_{Y|X}) + n\varepsilon_n,
\end{aligned}$$

where (a) follows from Fano's inequality and the data processing inequality [29] since $S^k \to X^n \to Y^n$ form a Markov chain, in (b) $Y^{i-1} \triangleq (Y_1, ..., Y_{i-1})$, and (c) holds since conditioning reduces entropy and Y_i is only dependent on X_i since the channel is memoryless. The proof of the converse part is completed by letting $P_{e,n}^{SID} \to 0$. ■

Observation 6.3 (Separation Principle Holds) It is readily verified that the condition $tH_{Q_{SL}}(S|L) < C(W_{Y|X})$ can be achieved by separate source coding with SI and channel coding. Therefore, separation of source and channel coding is optimal from the point of view of reliable transmissibility with source SI.

Remark 6.5 Note that the same JSCC theorem holds for the JSCC system when source SI is available at both the encoder and decoder, and the separation principle also holds for this case.

6.5 Source Side Information Can Increase the JSCC Error Exponent

We next observe that the SI does not only enlarge the achievable region for transmission (see Theorem 6.5 and recall that $H_{Q_{SL}}(S|L) \leq H_{Q_S}(S)$), but also improves the reliability of transmission. Obviously, if the sources Q_{SL} and the channel $W_{Y|X}$ satisfy $\tau H_{Q_{SL}}(S|L) < C(W_{Y|X}) < \tau H_{Q_S}(S)$, then we have

$$E_J^{SID}(Q_{SL}, W_{Y|X}, \tau) \geq \underline{E}_J^{SID}(Q_{SL}, W_{Y|X}, \tau) > 0 = E_J(Q_{SL}, W_{Y|X}, \tau).$$

Recalling that we also have an upper bound for E_J given by (5.6), thus, we can study the benefits of E_J^{SID} over E_J by comparing the lower bound $\underline{E}_J^{SID}(Q_{SL}, W_{Y|X}, \tau)$ with the upper bound $\overline{E}_{Jsp}(Q_{SL}, W_{Y|X}, \tau)$.

Given a nonuniform DMS Q_S, a DMC $W_{Y|X}$ and $\tau > 0$ such that $\tau H_{Q_S}(S) < C(W_{Y|X})$, we recall from Theorem 5.2 that $\overline{E}_{Jsp}(Q_{SL}, W_{Y|X}, \tau) = \underline{E}_{Jr}(Q_{SL}, W_{Y|X}, \tau)$ if and only if $\overline{\rho}^* \leq 1$, where

$$\overline{\rho}^* \triangleq \arg \max_{0 \leq \rho < \infty} [T_{sp}(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)]. \tag{6.49}$$

Furthermore, the upper and lower bound would be achieved by $R_m = \overline{R}_m = \underline{R}_m = \tau H_{Q_S^{(\overline{\rho}^*)}}(S) < C(W_{Y|X})$, i.e.,

$$\begin{aligned} \overline{E}_J(Q_S, W_{Y|X}, \tau) &= \underline{E}_J(Q_S, W_{Y|X}, \tau) = T_{sp}(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S) \\ &= \tau e\left(\frac{R_m}{\tau}, Q_S\right) + E_{sp}(R_m, W_{Y|X}) > 0. \end{aligned} \quad (6.50)$$

This means that the minimums in

$$\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) = \min_{P_S}[\tau D(P_S \| Q_S) + E_{sp}(\tau H(P_S), W_{Y|X})]$$

and

$$\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) = \min_{P_S}[\tau D(P_S \| Q_S) + E_r(\tau H(P_S), W_{Y|X})]$$

would be uniquely achieved by $P_S^* = Q_S^{(\overline{\rho}^*)}$.

Assume that for P_S satisfying $\tau H_{P_S}(S) = R_m$, $E_{SI}(P_S, Q_{L|S}, W_{Y|X})$ in (6.24) is achieved by an $R^* < \tau H_{P_S}(S) = R_m$, then we must have

$$E_r^*(P_S, W_{Y|X}) = E_{SI}(P_S, Q_{L|S}, W_{Y|X}) = E_r(R^*, W_{Y|X}) > E_r(R_m, W_{Y|X})$$

since $E_r(R, W_{Y|X})$ is strictly decreasing at R_m. This yields

$$\begin{aligned} &\min_{P_S:\tau H_{P_S}(S)=R_m} \left[\tau D(P_S \| Q_S) + E_r^*(H_{P_S}(S), W_{Y|X})\right] \\ &> \min_{P_S:\tau H_{P_S}(S)=R_m} [\tau D(P_S \| Q_S) + E_r(\tau H_{P_S}(S), W_{Y|X})] \\ &\geq \min_{P_S}[\tau D(P_S \| Q_S) + E_r(\tau H_{P_S}(S), W_{Y|X})] \\ &= \underline{E}_J(Q_S, W_{Y|X}, \tau) = \overline{E}_J(Q_S, W_{Y|X}, \tau) = E_J(Q_S, W_{Y|X}, \tau). \end{aligned} \quad (6.51)$$

Now we claim that if $\overline{\rho}^* \leq 1$ and (6.24) is achieved by an $R^* < \tau H_{P_S}(S) = R_m$, then

$$\underline{E}_J^{SID}(Q_S, W_{Y|X}, \tau) > E_J(Q_S, W_{Y|X}, \tau).$$

Indeed, if $\underline{E}_J^{SID}$ given in (6.34) is achieved by a $\widehat{P}_S$ such that $\tau H_{\widehat{P}_S}(S) \neq R_m$, then by definition of $E_r^*(P_S, W_{Y|X})$,

$$\begin{aligned} \underline{E}_J^{SID}(Q_S, W_{Y|X}, t) &\geq \tau D(\widehat{P}_S \| Q_S) + E_r(\tau H_{\widehat{P}_S}(S), W_{Y|X}) \\ &> \tau D(Q_S^{(\overline{\rho}^*)} \| Q_S) + E_r(\tau H_{Q_S^{(\overline{\rho}^*)}}(S), W_{Y|X}) \\ &= \underline{E}_{Jr}(Q_S, W_{Y|X}, t) = \overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau) = E_J(Q_S, W_{Y|X}, \tau). \end{aligned}$$

Otherwise if $\tau H_{\widehat{P}_S}(S) = R_m$, then (6.51) ensures that $\underline{E}_J^{SID}(Q_S, W_{Y|X}, t) > E_J(Q_S, W_{Y|X}, t)$. In order to guarantee that (6.24) is achieved by an $R^* < \tau H_{P_S}(S) = R_m$, noting that E_r is decreasing in R and that e_r is increasing in R, we only need that there exists an intersection between E_r and e_r, i.e., we need the condition

$$E_r(R_m, W_{Y|X}) < \min_{P_S : \tau H_{P_S}(S) = R_m} \tau e_r(H_{P_S}(S), Q_{L|S}, P_S). \tag{6.52}$$

Since

$$\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau) = \tau e\left(\frac{R_m}{\tau}, Q_S\right) + E_r(R_m, W_{Y|X}) = \tau D(Q_S^{(\overline{\rho}^*)} \| Q_S) + E_r(R_m, W_{Y|X}),$$

and

$$e_r(H_{P_S}(S), Q_{L|S}, P_S) = E_r(0, P_S, Q_{L|S}) = E_o(1, P_S, Q_{L|S}),$$

where $E_o(1, P_S, Q_{L|S})$ is given in (2.19), we may write (6.52) by

$$T_{sp}(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S) + \tau D(Q_S^{(\overline{\rho}^*)} \| Q_S) < \min_{H_{P_S}(S) = H_{Q_S^{(\overline{\rho}^*)}}(S)} t E_o(1, P_S, Q_{L|S}).$$

This together with $\overline{\rho}^* \leq 1$ yields a sufficient condition for which $\underline{E}_J^{SID} > E_J$. In general, we need to first compute $E_o(\overline{\rho}^*, W_{Y|X})$ and take a concave hull to obtain $T_{sp}(\overline{\rho}^*, W_{Y|X})$, and we need to minimize $E_o(1, P_S, Q_{L|S})$ over P_S, but if the source is binary and the channel is symmetric, the above condition can be further simplified and easily verified.

Corollary 6.4 *Let $Q_S = \{q, 1-q\}$ ($q < 0.5$) be a binary DMS, and $W_{Y|X}$ be symmetric such that $\tau H_{Q_S}(S) < C(W_{Y|X})$. If $\overline{\rho}^* \leq 1$ and*

$$E_o(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S) < \tau E_o(1, Q_S^{(\overline{\rho}^*)}, Q_{L|S}) - \tau D(Q_S^{(\overline{\rho}^*)} \| Q_S),$$

then $E_J^{SID} > E_J$, where $\overline{\rho}^*$ achieves the maximum of $E_o(\rho, W_{Y|X}) - \tau E_s(\rho, Q_S)$ and is given by (5.33).

Example 6.3 Let the transmitted source Q_S be a binary DMS with distribution $Q_S = \{q, 1-q\}$ ($q < 0.5$), and let the channel $W_{Y|X}$ be a binary symmetric channel (BSC) with crossover probability $\varepsilon \in (0, 0.5)$. The source Q_L is a noisy version of Q_S described by $L = S \oplus N \mod 2$ ($\mathcal{L} = \mathcal{N} = \{0,1\}$) with noise distribution $P_N(N=1) = 0.05$, i.e., the SI is transmitted through a dummy BSC $Q_{L|S}$ with crossover probability 0.05. Set the transmission rate $t = 0.75$. Fig. 6.6 shows the regions of the binary source and the BSC parameters, i.e., (ε, q) pairs, for which the source can be reliably transmitted over the channel and E_J^{SID} can be strictly larger than E_J by Lemma 6.4. Region **A** (including the boundary with **B**) is the region where $\tau H_{Q_{SL}}(S|L) \geq C(W_{Y|X})$, i.e., where both E_J^{SID} and E_J are zero. Region **B** (including the boundary with **C**) is the region where $\tau H_{Q_{SL}}(S|L) < C(W_{Y|X}) \leq \tau H_{Q_S}(S)$, i.e., where E_J is zero, but E_J^{SID} is positive. Region **C** (not including the boundary with **D**) is the region where both E_J^{SID} and E_J are positive, but the condition given in Lemma 6.4 holds, i.e., $E_J^{SID} > E_J > 0$. In Region **D**, both exponents E_J^{SID} and E_J are positive, and the condition in Lemma 6.4 is not satisfied. Note that Lemma 6.4 only gives a sufficient condition which can be easily verified. This condition is however not necessary for having $E_J^{SID} > E_J$; this is illustrated in Fig. 6.7, where we note that $E_J^{SID} > E_J$ for some $(\varepsilon, q) \in \mathbf{D}$.

We plot in Fig. 6.7 the lower bound E_J^{SID} given in (6.34), the sphere-packing upper bound $\overline{E}_{Jsp}$ given in (5.6), and the random-coding lower bound $\underline{E}_{Jr}$ given in (5.5) for the above DMS(q)–BSC(ε) system with $q = 0.1$. The plots show that $\underline{E}_J^{SID}$ is strictly larger than $\overline{E}_J$ for $\varepsilon > 0.0045$, and $\underline{E}_J^{SID}$ coincides with $\underline{E}_J$ for $\varepsilon \leq 0.002$. We note that when the channel has large noise ($\varepsilon > 0.01$), the SI can substantially improve the error exponent. Furthermore, E_J is zero for $\varepsilon \geq 0.175$, but $\underline{E}_J^{SID}$ is still positive until $\varepsilon = 0.29$. Thus with the SI Q_L at the decoder, $E_J^{SID} > E_J$ holds for a large class of source-channel conditions.

6.6 Conclusion

In Chapter 5 we mainly focused on the analytical computation of lower and upper bounds for the JSCC error exponent, while in this chapter we dealt with more on the bounding technique.

For the system with feedback, we established a conceptual upper bound for $E_{J,fb}$ by using a simple type counting argument. We will employ the same approach to other discrete

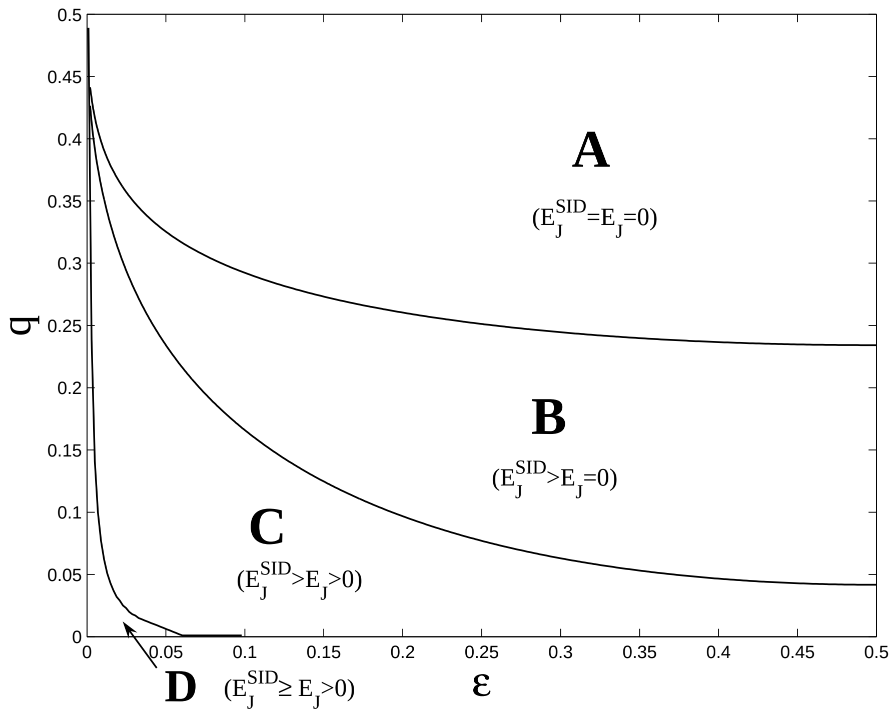

Figure 6.6: Side information at the decoder can enlarge the source-channel parameters for reliable transmissibility for the binary DMS(q)–BSC(ε) system of Example 6.3, $\tau = 0.75$.

systems (in Chapters 7 and 9). The lower bound for $E_{J,fb}$ is obtained by slightly modifying Zigangirov's iterative coding scheme. Although this bound is hard to compute in general, we showed that it is at least as good as Gallager's lower bound for E_J for binary input channels, and that it does coincide with the upper bound $\overline{E}_{Jsp}$ for much more source-channel pairs than Gallager's bound. Using this lower bound, we numerically illustrated that feedback can strictly increase the JSCC error exponent for channels with binary input alphabet and a symmetric distribution.

For the system with source SI at decoder, an achievable lower bound for the JSCC error exponent is obtained by using the method of types. The proofs generalize the one of Csiszár's for the random-coding lower bound $\underline{E}_{Jr}$. In particular, to prove the lower bound, we combined the maximum mutual information decoder of [30] and the minimum conditional entropy decoder of [73]. Consequently, JSCC theorem for system with source SI is formulated. Finally, we compared the lower bound $\underline{E}_J^{SID}$ with the upper bound $\overline{E}_{Jsp}$,

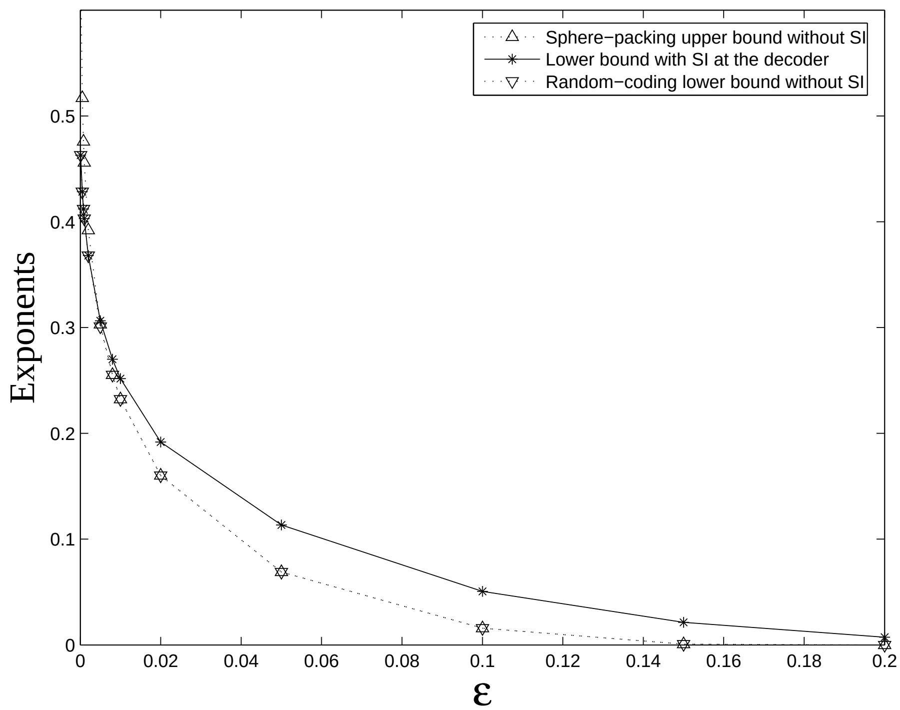

Figure 6.7: SI at the decoder can increase the JSCC error exponent for the binary DMS(q)–BSC(ε) system of Example 6.3, $q = 0.1$, $\tau = 0.75$.

and a sufficient condition for which $E_J^{SID} > E_J$ is given for binary DMSs and symmetric DMCs. Numerical results show that SI (at the decoder) not only enlarges the region of the source-channel parameters for which reliable transmissibility is possible, but it can also provide a noticeable increase in the JSCC error exponent for a large class of source-channel pairs.

Chapter 7

JSCC Error Exponent for Discrete Systems with Markovian Memory

In this chapter, we investigate the JSCC error exponent, E_J, for a discrete communication system with Markovian memory. Specifically, we establish a (computable) upper bound for E_J for transmitting a stationary ergodic (irreducible) Markov (SEM) source $\mathbf{Q_S}$ over a channel $\mathbf{W_{Y|X}}$ with additive SEM noise $\mathbf{P_Z}$ (for the sake of brevity, we hereafter refer to this channel as the SEM channel $\mathbf{W_{Y|X}}$). Note that Markov sources are widely used to model realistic data sources, and binary SEM channels can approximate well binary input hard-decision demodulated fading channels with memory (e.g., see [75], [100], [101]).

Section 7.1 contains JSCC system description and preliminaries on the information rates for systems with memory. In Section 7.2, we prove a strong converse JSCC theorem for systems consisting of an ergodic discrete source and a discrete channel with modulo-additive ergodic noise. We then introduce the notion of artificial (or auxiliary) Markov sources adopted from [94] in Section 7.3. Some interesting results involving Markov sources and their artificial counterparts are presented.

In Section 7.4, we deal with the main results regarding the upper and lower bounds for the JSCC error exponent E_J for SEM source-channel pairs. We first derive a computable sphere-packing type upper bound for E_J. The proof of the bound, following the standard lower bounding technique for the average probability of error, is based on the judicious construction from the original SEM source-channel pair ($\mathbf{Q_S}$,$\mathbf{W_{Y|X}}$) of an artificial Markov source $\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha^*)}$ and an artificial channel $\mathbf{V_{Y|X}}$ with additive Markov noise $\widetilde{\mathbf{P}}_{\mathbf{Z}}^{(\alpha^*)}$, where α^* is a parameter to be optimized, such that the stationarity and ergodicity properties are retained by $\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha^*)}$ and $\widetilde{\mathbf{P}}_{\mathbf{Z}}^{(\alpha^*)}$. The proof then employs the strong converse JSCC Theorem for ergodic sources and channels with ergodic additive noise and the fact that the normalized log-likelihood ratio between n-tuples of two SEM sources asymptotically converges (as $n \to \infty$) to their Kullback-Leibler divergence rate. As by-products, we obtain upper bounds for the error exponent for SEM sources and SEM channels.

We next examine Gallager's lower bound for E_J (which is valid for arbitrary source-channel pairs with memory), when specialized to the SEM source-channel system. By comparing our upper bound with Gallager's lower bound, we provide the condition under which they coincide, hence exactly determining E_J. We note that this condition holds for a large class of SEM source-channel pairs. Using the Fenchel duality theorem, we provide equivalent representations for these bounds. We show that our upper bound (respectively Gallager's lower bound) to E_J, can also be represented by the minimum of the sum of SEM source error exponent and the upper (respectively lower) bound of SEM channel error exponent. In this regard, our result is a natural extension of Csiszár's bounds from the case of memoryless systems to the case of SEM systems.

We next introduce Markov types and employ the method of types to prove another upper bound for the JSCC error exponent in terms of the SEM source exponent and the SEM channel exponent. This upper bound may not be computable, but we will use it as a tool to compare the JSCC error exponent with tandem coding error exponent in Chapter 10. In Section 7.5, we briefly remark the extension of our results to SEM systems with arbitrary Markovian orders. Finally, a conclusion is drawn in Section 7.6.

7.1 System Description and Definitions

7.1.1 System

We consider a communication system with transmission rate τ (source symbols/channel use) consisting of a discrete source with finite alphabet $\mathcal{S}$ described by the sequence of τn-dimensional distributions $\mathbf{Q_S} \triangleq \{Q_{S^{\tau n}} \in \mathcal{P}(\mathcal{S}^{\tau n})\}_{\tau n=1}^{\infty}$, and a discrete channel described by the sequence of n-dimensional transition distributions $\mathbf{W_{Y|Z}} \triangleq \{W_{Y^n|X^n} \in \mathcal{P}(\mathcal{Y}^n|\mathcal{X}^n)\}_{n=1}^{\infty}$ with common input and output alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1, ..., B-1\}$.

A JSC code with blocklength n and transmission rate τ is a pair of mappings:

$$f_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n}.$$

In this chapter, we confine our attention to discrete channels with (modulo B) additive noise of n-dimensional distribution $\mathbf{P_Z} \triangleq \{P_{Z^n} \in \mathcal{P}(\mathcal{Z}^n)\}_{n=1}^{\infty}$; see Fig. 7.1. The channels are described by

$$Y_i = X_i \oplus Z_i \pmod{B},$$

where Y_i, X_i and Z_i are the channel's output, input and noise symbols at time i, and $Z_i \in \mathcal{Z} = \{0, 1, ..., B-1\}$ is independent of X_i, $i = 1, 2, ..., n$.

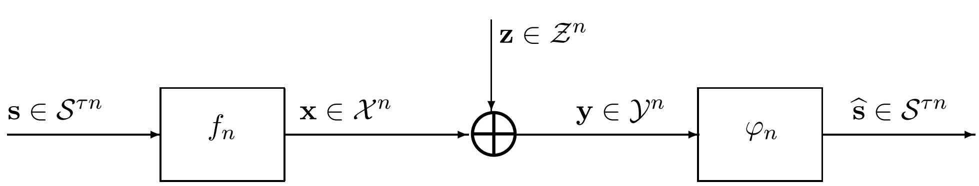

Figure 7.1: JSCC system for discrete sources and discrete channels with additive noise.

Denote the transmitted source message by $\mathbf{s} \in \mathcal{S}^{\tau n}$, the corresponding n-length codeword by $f_n(\mathbf{s}) = \mathbf{x} \triangleq (x_1, x_2, ..., x_n) \in \mathcal{X}^n$ and the received codeword at the channel output by

$\mathbf{y} \triangleq (y_1, y_2, ..., y_n) \in \mathcal{Y}^n$. Then the probability of receiving $\mathbf{y}$ under the conditions that the message $\mathbf{s}$ is transmitted (i.e., the input codeword is $f_n(\mathbf{s}) = \mathbf{x}$) is given by

$$\Pr(Y^n = \mathbf{y}|S^{\tau n} = \mathbf{s}) = W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) = W_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = W_{Y^n|X^n}(\mathbf{y} \ominus \mathbf{x}|\mathbf{x}) = P_{Z^n}(\mathbf{z}),$$

where the last equality follows by the independence of input codeword $\mathbf{x}$ and the additive noise $\mathbf{z} = \mathbf{y} \ominus \mathbf{x}$, noting that $\ominus$ is modulo-B subtraction here. The decoding operation φ_n is the rule decoding on a set of non-intersecting sets of output words $A_{\mathbf{s}}$ such that $\bigcup_{\mathbf{s}} A_{\mathbf{s}} = \mathcal{Y}^n$. If $\mathbf{y} \in A_{\mathbf{s}'}$, then we conclude that the source message $\mathbf{s}'$ has been transmitted. If the source message $\mathbf{s}$ has been transmitted, the conditional error probability in decoding is given by

$$\Pr\left(\{\mathbf{y} : \mathbf{y} \in A_{\mathbf{s}}^c\} | \, \mathbf{s}\right) \triangleq \sum_{\mathbf{y}:\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})),$$

and the probability of error of the code (f_n, φ_n) is

$$P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y} \in A_{\mathbf{s}}^c} Q_{S^{\tau n}}(\mathbf{s}) W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})). \tag{7.1}$$

Definition 7.1 The JSCC error exponent $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ for source $\mathbf{Q_S}$ and channel $\mathbf{W_{Y|X}}$ is defined as the supremum of the set of all numbers E for which there exists a sequence of JSC codes (f_n, φ_n) with transmission rate τ blocklength n such that

$$E \leq \liminf_{n \to \infty} -\frac{1}{n} \log_2 P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

When there is no possibility of confusion, $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ will be written as E_J. A lower bound for E_J for arbitrary discrete source-channel pairs with memory was already obtained by Gallager [42, Problem 5.16]. In Section 7.4, we will establish a computable upper bound for E_J and compare it with Gallager's bound.

7.1.2 Information Rates

For a discrete source $\mathbf{Q_S}$, its (limsup) entropy rate is defined by

$$\overline{H}_{\mathbf{Q_S}}(S) \triangleq \limsup_{k \to \infty} \frac{1}{k} H_{Q_{S^k}}(S^k),$$

where $H_{Q_{S^k}}(S^k)$ is the Shannon entropy of Q_{S^k}. $\overline{H}_{\mathbf{Q_S}}(S)$ admits an operational meaning (in the sense of the lossless fixed length source coding theorem) if $\mathbf{Q_S}$ is information stable [49]. The source Rényi entropy rate of order α ($\alpha \geq 0$) is defined by

$$\overline{H}^{(\alpha)}_{\mathbf{Q_S}}(S) \triangleq \limsup_{k\to\infty} \frac{1}{k} H^{(\alpha)}_{Q_{S^k}}(S^k),$$

where

$$H^{(\alpha)}_{Q_{S^k}}(S^k) \triangleq \frac{1}{1-\alpha} \log_2 \sum_{\mathbf{s}\in\mathcal{S}^k : Q_{S^k}(\mathbf{s})>0} Q_{S^k}(\mathbf{s})^{\alpha},$$

is the Rényi entropy of Q_{S^k}, and the special case of $\alpha = 1$ should be interpreted as

$$H^{(1)}_{Q_{S^k}}(S^k) \triangleq \lim_{\alpha\to 1} \frac{1}{1-\alpha} \log_2 \sum_{\mathbf{s}\in\mathcal{S}^k : Q_{S^k}(\mathbf{s})>0} Q_{S^k}(\mathbf{s})^{\alpha} = H_{Q_{S^k}}(S^k).$$

The channel capacity for any discrete (information stable [49], [96]) channel $\mathbf{W_{Y|X}}$ is given by

$$\overline{C}(\mathbf{W_{Y|X}}) = \liminf_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} I_{P_{X^n}W_{Y^n|X^n}}(X^n;Y^n).$$

For discrete channels with finite-input finite-output alphabets, the supremum is achievable and can be replaced by maximum. If the channel $\mathbf{W_{Y|X}}$ is an additive noise channel with noise process $\mathbf{P_W}$, then

$$\overline{C}(\mathbf{W_{Y|X}}) = \log_2 B - \overline{H}_{\mathbf{P_Z}}(Z),$$

where $\overline{H}_{\mathbf{P_Z}}(Z)$ is the noise entropy rate.

7.2 A Strong Converse JSCC Theorem

We first prove a strong converse JSC coding theorem for ergodic sources and channels with additive ergodic noise; no Markov assumption for either the source or the channel is needed for this result. We will employ it to prove an upper bound for E_J.

Theorem 7.1 (Strong converse JSCC Theorem) *For a source* $\mathbf{Q_S}$ *and a channel* $\mathbf{W_{Y|X}}$ *with additive noise* $\mathbf{P_Z}$ *such that* $\mathbf{Q_S}$ *and* $\mathbf{P_Z}$ *are ergodic processes, if*

$$\overline{C}(\mathbf{W_{Y|X}}) = \log_2 B - \overline{H}_{\mathbf{P_Z}}(Z) < \tau \overline{H}_{\mathbf{Q}}(S),$$

then

$$\lim_{n\to\infty} P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 1.$$

Proof: Assume $\overline{C}(\mathbf{W_{Y|X}}) = \tau \overline{H}_{\mathbf{Q_S}}(S) - \varepsilon$ ($\varepsilon > 0$). We first recall the fact that for additive channels the channel capacity $\overline{C}(\mathbf{W_{Y|X}})$ is achieved by the uniform input distribution $\widehat{P}_{X^n}(\mathbf{x}) \triangleq 1/B^n$. Furthermore, this uniform input distribution yields a uniform distribution at the ouput

$$\widehat{P}_{Y^n}(\mathbf{y}) \triangleq \sum_{\mathbf{x}\in\mathcal{X}^n} \widehat{P}_{X^n}(\mathbf{x}) W_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \frac{1}{B^n}.$$

Define for some δ ($0 < \delta < \varepsilon$)

$$\widehat{A}_{\mathbf{s}} = \left\{ \mathbf{y} : \log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{\widehat{P}_{Y^n}(\mathbf{y})} \le n\left(\overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S) + \delta\right) \right\}.$$

Since

$$P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 1 - \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in A_{\mathbf{s}}} Q_{S^{\tau n}}(\mathbf{s}) W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})), \tag{7.2}$$

we need to show that

$$\Pr\left(\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in A_{\mathbf{s}}\}\right) = \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in A_{\mathbf{s}}} Q_{S^{\tau n}}(\mathbf{s}) W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$$

vanishes as n goes to infinity. Note that

$$\Pr\left(\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in A_{\mathbf{s}}\}\right) \le \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in A_{\mathbf{s}} \cap \widehat{A}_{\mathbf{s}}\right\}\right) + \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in \widehat{A}_{\mathbf{s}}^c\right\}\right).$$

For the first sum, we have

$$\begin{aligned}
&\Pr\left(\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in A_{\mathbf{s}}\}\right) \\
&= \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y}\in A_{\mathbf{s}}\cap\widehat{A}_{\mathbf{s}}} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\
&\le \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y}\in A_{\mathbf{s}}\cap\widehat{A}_{\mathbf{s}}} \frac{\widehat{P}_{Y^n}(\mathbf{y})}{Q_{S^{\tau n}}(\mathbf{s})} 2^{n\left(\overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S) + \delta\right)} \\
&\le 2^{n\left(\overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S) + \delta\right)} \sum_{\mathbf{s}} \sum_{\mathbf{y}\in A_{\mathbf{s}}} \widehat{P}_{Y^n}(\mathbf{y}) \\
&= 2^{-n(\varepsilon-\delta)}.
\end{aligned} \tag{7.3}$$

For the second sum, we have

$$
\begin{aligned}
&\Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in \widehat{A}_{\mathbf{s}}^c\right\}\right)\\
&= \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \frac{1}{n}\log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{\widehat{P}_{Y^n}(\mathbf{y})} - \left(\overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S)\right) > \delta\right\}\right)\\
&\le \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \left|\frac{1}{n}\log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{\widehat{P}_{Y^n}(\mathbf{y})} - \left(\overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S)\right)\right| > \delta\right\}\right)\\
&= \Pr\Bigg(\Bigg\{(\mathbf{s},\mathbf{y}) : \Bigg|-\frac{1}{n}\log_2 W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) - \frac{1}{n}\log_2 Q_{S^{\tau n}}(\mathbf{s})\\
&\qquad -\overline{H}_{\mathbf{P_Z}}(Z) - \tau\overline{H}_{\mathbf{Q_S}}(S)\Bigg| > \delta\Bigg\}\Bigg)\\
&\le \Pr\left(\left\{\mathbf{s} : \left|-\frac{1}{\tau n}\log_2 Q_{S^{\tau n}}(\mathbf{s}) - \overline{H}_{\mathbf{Q_S}}(S)\right| > \frac{\delta}{2\tau}\right\}\right)\\
&\quad +\Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \left|-\frac{1}{n}\log_2 W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) - \overline{H}_{\mathbf{P_Z}}(Z)\right| > \frac{\delta}{2}\right\}\right)\\
&= \Pr\left(\left\{\mathbf{s} : \left|-\frac{1}{\tau n}\log_2 Q_{S^{\tau n}}(\mathbf{s}) - \overline{H}_{\mathbf{Q_S}}(S)\right| > \frac{\delta}{2\tau}\right\}\right)\\
&\quad +\Pr\left(\left\{\mathbf{z} : \left|-\frac{1}{n}\log_2 P_{Z^n}(\mathbf{z}) - \overline{H}_{\mathbf{P_Z}}(Z)\right| > \frac{\delta}{2}\right\}\right) \qquad (7.4)
\end{aligned}
$$

where the above probabilities are taken under the joint distribution $Q_{S^{\tau n}}(\mathbf{s})W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$, and (7.4) follows from an exchange of RV $\mathbf{z} = \mathbf{y} \ominus f_n(\mathbf{s})$ and the fact that $P_{Z^n}(\mathbf{z}) = W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$. It follows from the well known Shannon-McMillan-Breiman Theorem for ergodic processes [18] that the above probabilities converge to 0 as n goes to infinity. On account of (7.3), (7.4) and (7.2), the proof is complete. ■

7.3 Markov Sources and Artificial Markov Sources

Without loss of generality, we consider first-order Markov sources since any L-th order Markov source can be converted to a first-order Markov source by L-step blocking it (see Section 7.5). For the sake of convenience (since we will apply the following results to both the SEM source and the SEM channel), we use, throughout this section, $\mathbf{P_U} \triangleq \{P_{U^n} \in \mathcal{P}(\mathcal{U}^n)\}_{n=1}^{\infty}$ to denote a first-order SEM source with finite alphabet $\mathcal{U} \triangleq \{1, 2, ..., M\}$, initial distribution

$$p_i \triangleq \Pr\{U_1 = i\}, \quad i \in \mathcal{U}$$

and transition distribution

$$p_{ij} \triangleq \Pr\{U_{k+1} = j | U_k = i\}, \quad i, j \in \mathcal{U},$$

so that the n-tuple probability is given by

$$\begin{aligned} P_{U^n}(i^n) &\triangleq \Pr\{U_1 = i_1, ..., U_n = i_n\} \\ &= p_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}, \quad i_1, ..., i_n \in \mathcal{U}. \end{aligned}$$

Denote the transition matrix by $P \triangleq [p_{ij}]_{M \times M}$, we then set, for any $0 \le \alpha \le 1$,

$$P(\alpha) \triangleq \left[p_{ij}^{\alpha}\right]_{M \times M},$$

which is nonnegative and irreducible $M \times M$ matrix (here we define $0^0 = 0$). The Perron-Frobenius Theorem [82] asserts that the matrix $P(\alpha)$ possesses a maximal positive eigenvalue $\lambda_\alpha(\mathbf{P_U})$ with positive (right) eigenvector $\mathbf{v}(\alpha) = (v_1(\alpha), ..., v_M(\alpha))^t$ such that

$$\sum_i v_i(\alpha) = 1.$$

As in [94], we define the artificial Markov source $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)} \triangleq \left\{\widetilde{P}_{U^n}^{(\alpha)} \in \mathcal{P}(\mathcal{U}^n)\right\}_{n=1}^{\infty}$ with respect to the original source $\mathbf{P_U}$ such that the transition matrix is $\widetilde{P}(\alpha) \triangleq [\widetilde{p}_{ij}(\alpha)]_{M \times M}$, where

$$\widetilde{p}_{ij}(\alpha) \triangleq \frac{p_{ij}^{\alpha} v_j(\alpha)}{\lambda_\alpha(\mathbf{P_U}) v_i(\alpha)}. \tag{7.5}$$

It can be easily verified that $\sum_j \widetilde{p}_{ij}(\alpha) = 1$. We emphasize that the artificial source retains the stochastic characteristics (irreducibility) of the original source because $\widetilde{p}_{ij}(\alpha) = 0$ if and only if $p_{ij} = 0$, and clearly, for all n, the nth marginal of $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}$ is absolutely continuous with respect to the nth marginal of $\mathbf{P_U}$. The entropy rate of the artificial Markov process is hence given by

$$\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}}(S) = -\sum_i \sum_j \pi_i(\alpha) \widetilde{p}_{ij}(\alpha) \log_2 \widetilde{p}_{ij}(\alpha),$$

where $\pi(\alpha) \triangleq (\pi(\alpha)_1, \pi(\alpha)_2, ..., \pi(\alpha)_M)$ is the stationary distribution of the stochastic matrix $\widetilde{P}(\alpha)$. We call the artificial Markov source with initial distribution $\pi(\alpha)$ the *artificial SEM source*. It is known [94, Lemmas 2.1-2.4] that $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}}(S)$ is a continuous and non-increasing function of $\alpha \in [0,1]$. In particular, $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(0)}}(S) = \log_2 \lambda_0(\mathbf{P_U})$ and $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(1)}}(S) = \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}}(S)$. The following lemma illustrates the relation between $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(0)}}(S)$ and the entropy of the DMS with uniform distribution $\left(\frac{1}{M}, ..., \frac{1}{M}\right)$.

Lemma 7.1 $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(0)}}(S) \le \log_2 M$ *with equality if and only if* $P > [0]_{M\times M}$*, i.e.,* $p_{ij} > 0$ *for all* $i, j \in \mathcal{U}$*.*

Proof: Let A be the $M \times M$ matrix with all components equal to 1, i.e., $A \triangleq [1]_{M\times M}$. Clearly, $\mathbf{u} \triangleq [\frac{1}{M}, ..., \frac{1}{M}]$ is the unique normalized positive eigenvector (Perron vector) of A with associated positive eigenvalue M; thus when $P > [0]_{M\times M}$, $\lambda_0(\mathbf{P_U}) = M$. We next show by contradiction that $\lambda_0(\mathbf{P_U}) < M$ if there are zero components in matrix P. We assume that there exist some $p_{ij} = 0$ and $\lambda_0(\mathbf{P_U}) \ge M$. Then

$$\lambda_0(\mathbf{P_U})\mathbf{u} \ge M\mathbf{u} = A\mathbf{u} = A\mathbf{v}(0),$$

where the last equality holds since $\mathbf{u}$ and $\mathbf{v}(0)$ are both normalized vectors. We thus have

$$(A - P(0))\mathbf{v}(0) \le \lambda_0(\mathbf{P_U})(\mathbf{u} - \mathbf{v}(0)).$$

Now summing all the components of the vectors on both sides, we obtain

$$\sum_{i,j} a_{ij} v_j(0) \le 0,$$

where a_{ij} is the (i,j)th component of the matrix $A - P(0)$ such that $a_{ij} = 0$ if $p_{ij} > 0$ and $a_{ij} = 1$ if $p_{ij} = 0$. This contradicts with the fact that all $v_j(0)$'s are positive and thus $\lambda_0(\mathbf{P_U}) < M$ if $\mathbf{P_U}$ has zero components. We also conclude that $P > [0]_{M\times M}$ is the sufficient and necessary condition for $\lambda_0(\mathbf{P_U}) = M$. ■

The following properties regarding the artificial SEM source are important in deriving the (computable) upper and lower bounds for the JSCC exponent of SEM source-channel pairs.

Lemma 7.2 *Let* $\{U_i\}_{i=1}^{\infty}$ *be an SEM source under* $\mathbf{P_U}$ *and* $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}$ $(0 < \alpha \le 1)$*, then*

$$\frac{1}{n}\log_2 \frac{\widetilde{P}_{U^n}^{(\alpha)}(U^n)}{P_{U^n}(U^n)} \longrightarrow \frac{1-\alpha}{\alpha}\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}}(S) - \frac{1}{\alpha}\log_2 \lambda_\alpha(\mathbf{P_U}),$$

almost surely under $\widetilde{P}_{U^n}^{(\alpha)}$ *as* $n \to \infty$.

Proof: Since $\{U_i\}_{i=1}^{\infty}$ is SEM source under $\mathbf{P_U}$ and $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}$, it follows by the Ergodic Theorem [18] that the normalized log-likelihood ratio between $\mathbf{P_U}$ and $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}$ converges to their Kullback-Leibler divergence rate almost surely, i.e.,

$$\frac{1}{n}\log_2\frac{\widetilde{P}_{U^n}^{(\alpha)}(U^n)}{P_{U^n}(U^n)} \longrightarrow D(\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}\|\mathbf{P_U})$$

almost surely under $\widetilde{P}_{U^n}^{(\alpha)}$ as $n \to \infty$, where

$$D(\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}\|\mathbf{P_U}) \triangleq \lim_{n\to\infty}\frac{1}{n}D(\widetilde{P}_{U^n}^{(\alpha)}\|P_{U^n}).$$

Note that for any n we can write

$$\frac{1}{n}D(\widetilde{P}_{U^n}^{(\alpha)}\|P_{U^n}) = -\frac{1}{n}H_{\widetilde{P}_{U^n}^{(\alpha)}}(U^n) - \frac{1}{n}\sum_{i^n}\widetilde{P}_{U^n}^{(\alpha)}(i^n)\log_2 P_{U^n}(i^n), \quad i^n = (i_1\cdots i_n)\in\mathcal{U}^n. \tag{7.6}$$

Recalling that $\mathbf{P_U}$ is described by the initial stationary distribution $\pi = \{\pi_1, \pi_2, ..., \pi_M\}$ and transition matrix $P = [p_{ij}]_{M\times M}$, and that $\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}$ is described by the initial stationary distribution $\pi(\alpha) = (\pi(\alpha)_1, \pi(\alpha)_2, ..., \pi(\alpha)_M)$ and transition matrix $\widetilde{P}(\alpha) \triangleq [\widetilde{p}_{ij}(\alpha)]_{M\times M}$ given by (7.5), we have

$$\begin{aligned}\widetilde{P}_{U^n}^{(\alpha)}(i^n) &= \pi(\alpha)_{i_1}\frac{p_{i_1i_2}^{\alpha}\cdots p_{i_{n-1}i_n}^{\alpha}}{\lambda_\alpha(\mathbf{P_U})^{n-1}}\frac{v_{i_n}(\alpha)}{v_{i_1}(\alpha)}\\ &= \frac{P_{U^n}(i^n)^\alpha}{\lambda_\alpha(\mathbf{P_U})^{n-1}}\frac{\pi(\alpha)_{i_1}v_{i_n}(\alpha)}{\pi_{i_1}^{\alpha}v_{i_1}(\alpha)}\end{aligned} \tag{7.7}$$

for all $i^n \in \mathcal{U}$. Consequently, using (7.6) and (7.7), we have

$$\begin{aligned}\frac{1}{n}D(\widetilde{P}_{U^n}^{(\alpha)}\|P_{U^n}) &= \frac{1-\alpha}{\alpha}\frac{1}{n}\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}}(S) - \frac{1}{\alpha}\frac{n-1}{n}\log_2\lambda_\alpha(\mathbf{P_U})\\ &\quad -\frac{1}{n}\frac{1}{\alpha}\sum_{i_1,i_n}\widetilde{p}_\alpha(i_1,i_n)\log_2\left(\frac{\pi_{i_1}^{\alpha}v_{i_1}(\alpha)}{\pi(\alpha)_{i_1}v_{i_n}(\alpha)}\right).\end{aligned} \tag{7.8}$$

Taking the limit on both sides of (7.8), and noting that the last term approaches 0 since

$$\left|\frac{1}{\alpha}\sum_{i_1,i_n}\widetilde{p}_\alpha(i_1,i_n)\log_2\left(\frac{\pi_{i_1}^{\alpha}v_{i_1}(\alpha)}{\pi(\alpha)_{i_1}v_{i_n}(\alpha)}\right)\right| \le \frac{M^2}{\alpha}\max_{i_1,i_n}\left|\log_2\left(\frac{\pi_{i_1}^{\alpha}v_{i_1}(\alpha)}{\pi(\alpha)_{i_1}v_{i_n}(\alpha)}\right)\right| < +\infty,$$

where π, $\pi(\alpha)$, and $\mathbf{v}(\alpha)$ are all positive for SEM sources (according to the Perron-Frobenius Theorem [82]). We hence obtain

$$D(\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)} \| \mathbf{P}_{\mathbf{U}}) = \frac{1-\alpha}{\alpha} \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\alpha)}}(S) - \frac{1}{\alpha} \log_2 \lambda_\alpha(\mathbf{P}_{\mathbf{U}}).$$

■

Lemma 7.3 [77], [94] *For an SEM source* $\mathbf{P}_{\mathbf{U}}$ *and any* $\rho \geq 0$*, we have*

$$\rho \overline{H}_{\mathbf{P}_{\mathbf{U}}}^{(\frac{1}{1+\rho})}(S) = (1+\rho) \log_2 \lambda_{\frac{1}{1+\rho}}(\mathbf{P}_{\mathbf{U}}),$$

and

$$\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{U}}^{(\frac{1}{1+\rho})}}(S) = \frac{\partial}{\partial \rho}(1+\rho) \log_2 \lambda_{\frac{1}{1+\rho}}(\mathbf{P}_{\mathbf{U}}).$$

Lemma 7.3 follows directly from [77, Lemma 1] and [94, Lemma 2.3]. Note that there is a slight error in the expression of $H(\alpha)$ in [94, Lemma 2.3], where a factor α is missing in the second term of the right-hand side of (2.11).

7.4 Upper and Lower Bounds

7.4.1 A Sphere-Packing Type Upper Bound

We first establish a sphere-packing type upper bound for E_J for SEM source-channel pairs $(\mathbf{Q}_{\mathbf{S}}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$. Before we proceed, we define the following function for an SEM source-channel pair:

$$F(\rho) \triangleq \rho \log_2 B - (1+\rho) \log_2 \left[\lambda^{\tau}_{\frac{1}{1+\rho}}(\mathbf{Q}_{\mathbf{S}}) \lambda_{\frac{1}{1+\rho}}(\mathbf{P}_{\mathbf{Z}}) \right], \quad \rho \geq 0. \tag{7.9}$$

Lemma 7.4 $F(\rho)$ *has the following properties:*

(a) $F(0) = 0$ *and*

$$f(\rho) \triangleq \frac{\partial}{\partial \rho} F(\rho) = \log_2 B - \left(\tau \overline{H}_{\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho})}}(S) + \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{Z}}^{(\frac{1}{1+\rho})}}(Z) \right) \tag{7.10}$$

is continuous non-increasing in ρ.

(b) *$F(\rho)$ is concave in ρ; hence every local maximum (stationary point) of $F(\cdot)$ is the global maximum.*

(c) *$\sup_{\rho\geq 0} F(\rho)$ is positive if and only if $\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$; otherwise $\sup_{\rho\geq 0} F(\rho) = 0$.*

(d) *$\sup_{\rho\geq 0} F(\rho)$ is finite if $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$ and infinite if $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_W}) < B$.*

Remark 7.1 If $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) \geq B$, then $\sup_{\rho\geq 0} F(\rho) = \lim_{\rho\to\infty} F(\rho)$, no matter whether the limit is finite or not.

Proof: We start from (a). $F(0) = 0$ since the largest eigenvalue for any stochastic matrix is 1. (7.10) follows from Lemma 7.3. $f(\rho)$ is continuous non-increasing function since $\overline{H}_{\widetilde{\mathbf{Q}}_\mathbf{S}^{(\frac{1}{1+\rho})}}(S)$ and $\overline{H}_{\widetilde{\mathbf{P}}_\mathbf{Z}^{(\frac{1}{1+\rho})}}(Z)$ are both continuous nondecreasing functions. (b) follows immediately from (a). (c) follows from the concavity of $F(\rho)$ and the facts that $F(0) = 0$ and that $f(0) = \overline{C}(\mathbf{W_{Y|X}}) - \tau\overline{H}_{\mathbf{Q_S}}(S)$. (d) follows from the concavity of $F(\rho)$ and the facts that $F(0) = 0$ and that $\lim_{\rho\to\infty} f(\rho) = \log_2 B - \log_2[\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z})]$. ■

Theorem 7.2 *For an SEM source $\mathbf{Q_S}$ and a discrete channel $\mathbf{W_{Y|X}}$ with additive SEM noise $\mathbf{P_Z}$ such that $\tau\overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ and $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$, the JSCC error exponent $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ satisfies*

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq \max_{\rho\geq 0} F(\rho). \tag{7.11}$$

Remark 7.2 We point out that the condition $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$ holds for most cases of interest. First note that the eigenvalues $\lambda_0(\mathbf{Q_S})$ and $\lambda_0(\mathbf{P_Z})$ are no less than 1. By Lemma 7.1, we have that $\lambda_0(\mathbf{P_Z}) = B$ if the noise transition matrix P_W has positive entries (i.e., $P_W > [0]_{B\times B}$); in that case, the condition $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$ is satisfied if $\lambda_0^\tau(\mathbf{Q_S}) > 1$ (i.e., if the source transition matrix Q is not a deterministic matrix). In fact, when $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) < B$, $\max_{\rho\geq 0} F(\rho) = +\infty$ by Lemma 7.4 (d), and hence it gives a trivial upper bound for E_J. When $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) = B$, we do not have an upper bound for E_J.

Remark 7.3 Using the first identity of Lemma 7.3, the upper bound can be equivalently represented as

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq \max_{\rho \geq 0} \left\{ \rho \left[\log_2 B - \tau \overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{Q_S}}(S) - \overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{P_Z}}(Z) \right] \right\}$$

where $\overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{Q_S}}(S)$ and $\overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{P_Z}}(Z)$ are the Rényi entropy rates of $\mathbf{Q_S}$ and $\mathbf{P_Z}$, respectively. Meanwhile, the upper bound (7.11) holds for any one of the following source-channel pairs: DMS Q and SEM channel $\mathbf{W_{Y|X}}$, SEM source $\mathbf{Q_S}$ and additive DMC W, and DMS source Q and additive DMC $W_{Y|X}$, all with finite alphabets. For example, when the source is DMS with distribution $\mathbf{q} \triangleq \{q_1, q_2, ..., q_M\}$ such that $q_i > 0$ for all $i = 1, 2, ..., M$, the source could be regarded as an SEM source $\mathbf{Q_S}$ with transition matrix

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_M \\ q_1 & q_2 & \cdots & q_M \\ \vdots & \vdots & \vdots & \vdots \\ q_1 & q_2 & \cdots & q_M \end{bmatrix}$$

and initial distribution $\mathbf{q}$. It is easy to verify that for such a Q, the eigenvalue $\lambda_{\frac{1}{1+\rho}}(Q)$ reduces to $\lambda_{\frac{1}{1+\rho}}(\mathbf{Q_S}) = \sum_i q_i^{1/1+\rho}$, which agrees with the results for memoryless systems given Chapter 5. Thus, the above bound is a sphere-packing type upper bound for E_J for SEM source-channel systems.

Proof of Theorem 7.2: Under the assumption

$$\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}}) \text{ and } \lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B,$$

it follows from Lemma 7.4 that $f(0) > 0$ and $\lim_{\rho\to\infty} f(\rho) < 0$. Since $f(\rho)$ is continuous and non-increasing, there must exist some $\rho_o \in (0, +\infty)$ such that $f(\rho_o) + \varepsilon = 0$, where $\varepsilon > 0$ is small enough. For the SEM source $\mathbf{Q_S}$, we introduce an artificial SEM source $\widetilde{\mathbf{Q}}^{(\alpha_o)}_{\mathbf{S}}$ (as described in Section 7.3) such that $\alpha_o \triangleq 1/(1+\rho_o) \in (0,1)$. For the SEM channel $\mathbf{W_{Y|X}}$, we introduce an artificial additive channel $\mathbf{V_{Y|X}}$ for which the corresponding SEM noise is $\widetilde{\mathbf{P}}^{(\alpha_o)}_{\mathbf{Z}}$.

Based on the construction of the artificial SEM source-channel pair $(\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha_o)}, \mathbf{V}_{\mathbf{Y}|\mathbf{X}})$, we define for some δ_1 ($\delta_1 > 0$) the set

$$\begin{aligned}\widetilde{A}_{\mathbf{s}} = \quad & \Bigg\{\mathbf{y} : \log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))\widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})} \\ & \geq -n\left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B + \varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda_{\alpha_o}^{\tau}(\mathbf{Q_S})\lambda_{\alpha_o}(\mathbf{P_Z})\right] + \delta_1\right)\Bigg\},\end{aligned}$$

where we set $\widetilde{A}_{\mathbf{s}} = \emptyset$ for those $\mathbf{s}$ such that $W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s}) = 0$ for some $\mathbf{y} \in \mathcal{Y}^n$. We then have a lower bound for the average probability of error

$$\begin{aligned}P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \quad & \geq \quad \sum_{\mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y}\in A_{\mathbf{s}}^c \cap \widetilde{A}_{\mathbf{s}}} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\ & \geq \quad 2^{-n\left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B+\varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda_{\alpha_o}^{\tau}(\mathbf{Q_S})\lambda_{\alpha_o}(\mathbf{P_Z})\right]+\delta_1\right)} \\ & \qquad \cdot \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in A_{\mathbf{s}}^c\cap\widetilde{A}_{\mathbf{s}}} \widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})), \qquad (7.12)\end{aligned}$$

where the last sum can be lower bounded as follows

$$\begin{aligned}& \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in A_{\mathbf{s}}^c\cap\widetilde{A}_{\mathbf{s}}} \widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\ & \geq \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in A_{\mathbf{s}}^c} \widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) - \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in \widetilde{A}_{\mathbf{s}}^c} \widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})). \quad (7.13)\end{aligned}$$

We point out that the first sum in the right-hand side of (7.13) is exactly the error probability of the JSC system consisting of the artificial SEM source $\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha_o)}$ and the artificial SEM channel $\mathbf{V}_{\mathbf{Y}|\mathbf{X}}$. Since by definition $f(\rho_o) < 0$, which implies

$$\tau \overline{H}_{\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha_o)}}(S) > \log_2 B - \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{Z}}^{(\alpha_o)}}(Z) = \overline{C}(\mathbf{V}_{\mathbf{Y}|\mathbf{X}}),$$

then applying the strong converse JSCC Theorem (Theorem 7.1) to $\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\alpha_o)}$ and $\mathbf{V}_{\mathbf{Y}|\mathbf{X}}$, the first sum in the right-hand side of (7.13) converges to 1 as n goes to infinity. We next show that the second term in the right-hand side of (7.13)

$$\Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in \widetilde{A}_{\mathbf{s}}^c\right\}\right) = \sum_{(\mathbf{s},\mathbf{y}):\mathbf{y}\in\widetilde{A}_{\mathbf{s}}^c} \widetilde{Q}_{S^{\tau n}}^{(\alpha_o)}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$$

vanishes asymptotically.

$$
\begin{aligned}
&\Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \mathbf{y} \in \widetilde{A}^c_{\mathbf{s}}\right\}\right) \\
&= \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \frac{1}{n}\log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))\widetilde{Q}^{(\alpha_o)}_{S^{\tau n}}(\mathbf{s})} \right.\right. \\
&\quad \left.\left. + \left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B + \varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda^{\tau}_{\alpha_o}(\mathbf{Q_S})\lambda_{\alpha_o}(\mathbf{P_Z})\right]\right) < -\delta_1\right\}\right) \\
&\le \Pr\left(\left\{(\mathbf{s},\mathbf{y}) : \left|\frac{1}{n}\log_2 \frac{W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))Q_{S^{\tau n}}(\mathbf{s})}{V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))\widetilde{Q}^{(\alpha_o)}_{S^{\tau n}}(\mathbf{s})} \right.\right.\right. \\
&\quad \left.\left.\left. + \left(\frac{1-\alpha_o}{\alpha_o}(\log_2 B + \varepsilon) - \frac{1}{\alpha_o}\log_2\left[\lambda^{\tau}_{\alpha_o}(\mathbf{Q_S})\lambda_{\alpha_o}(\mathbf{P_Z})\right]\right)\right| > \delta_1\right\}\right) \\
&= \Pr\left(\left\{(\mathbf{s},\mathbf{z}) : \left|\frac{1}{n}\log_2 \frac{\widetilde{P}^{\alpha_o}_{Z^n}(\mathbf{z})}{P_{Z^n}(\mathbf{z})} + \frac{1}{n}\log_2 \frac{\widetilde{Q}^{(\alpha_o)}_{S^{\tau n}}(\mathbf{s})}{Q_{S^{\tau n}}(\mathbf{s})} \right.\right.\right. \\
&\quad - \left[\tau\left(\frac{1-\alpha_o}{\alpha_o}\overline{H}_{\widetilde{\mathbf{Q}}^{(\alpha_o)}_{\mathbf{S}}}(S) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\mathbf{Q_S})\right)\right. \\
&\quad \left.\left.\left.\left. + \frac{1-\alpha_o}{\alpha_o}\overline{H}_{\widetilde{\mathbf{P}}^{(\alpha_o)}_{\mathbf{Z}}}(Z) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\mathbf{P_Z})\right]\right| > \delta_1\right\}\right) \qquad (7.14) \\
&\le \Pr\left(\left\{\mathbf{s} : \left|\frac{1}{\tau n}\log_2 \frac{\widetilde{Q}^{(\alpha_o)}_{S^{\tau n}}(\mathbf{s})}{Q_{S^{\tau n}}(\mathbf{s})} - \left[\frac{1-\alpha_o}{\alpha_o}\overline{H}_{\widetilde{\mathbf{Q}}^{(\alpha_o)}_{\mathbf{S}}}(S) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\mathbf{Q_S})\right]\right| > \frac{\delta_1}{2\tau}\right\}\right) \\
&\quad + \Pr\left(\left\{\mathbf{z} : \left|\frac{1}{n}\log_2 \frac{\widetilde{P}^{(n)}_{W_{\alpha_o}}(\mathbf{z})}{P^{(n)}_{W}(\mathbf{z})} - \left[\frac{1-\alpha_o}{\alpha_o}\overline{H}_{\widetilde{\mathbf{P}}^{(\alpha_o)}_{\mathbf{Z}}}(Z) - \frac{1}{\alpha_o}\log_2\lambda_{\alpha_o}(\mathbf{P_Z})\right]\right| > \frac{\delta_1}{2}\right\}\right), \qquad (7.15)
\end{aligned}
$$

where the probabilities are taken under the joint distribution $\widetilde{Q}^{(\alpha_o)}_{S^{\tau n}}(\mathbf{s})V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$, and (7.14) follows from the facts that $P_{Z^n}(\mathbf{z}) = W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$ and that $\widetilde{P}^{(\alpha_o)}_{Z^n}(\mathbf{z}) = V_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s}))$ for $\mathbf{z} = \mathbf{y} \ominus f_n(\mathbf{s})$.

Applying Lemma 7.2, the above probabilities converge to 0 as $n \to \infty$.[1] On account of (8.15), (7.13) and (7.15) and noting that ε and δ_1 are arbitrary, we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P^{(n)}_e(Q_{S^{\tau n}}, W_{Y^n|X^n}, \tau) \le \frac{1-\alpha_o}{\alpha_o}\log_2 B - \frac{1}{\alpha_o}\log_2\left[\lambda^{\tau}_{\alpha_o}(\mathbf{Q_S})\lambda_{\alpha_o}(\mathbf{P_Z})\right].$$

Finally, replacing α_o by $1/(1+\rho_o)$ in the above right-hand side terms and taking the maximum over ρ_o complete the proof. ∎

[1]Convergence almost surely implies convergence in probability.

7.4.2 Gallager's Lower Bound for Systems with Memory

In Observation 5.1, we showed that Gallager's JSCC lower bound [42, Problem 5.16] may not as good as Csiszár's source-channel random-coding lower bound. However, Gallager's lower bound is more powerful in the sense that it applies for discrete source and channel pairs with arbitrary memory, i.e., for a discrete source $\mathbf{Q_S}$ and a discrete channel $\mathbf{W_{Y|X}}$ with transmission rate τ, we have

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \geq \max_{0\leq\rho\leq 1}[E_o(\rho, \mathbf{Q_S}) - \tau E_s(\rho, \mathbf{W_{Y|X}})]$$

in which

$$E_s(\rho, \mathbf{Q_S}) \triangleq \limsup_{\tau n\to\infty} \frac{(1+\rho)}{\tau n} \log_2 \sum_{\mathbf{s}\in\mathcal{S}^n} Q_{S^{\tau n}}(\mathbf{s})^{\frac{1}{1+\rho}} \tag{7.16}$$

is Gallager's source function for the discrete source $\mathbf{Q_S}$ with arbitrary memory, and

$$E_o(\rho, \mathbf{W_{Y|X}}) \triangleq \liminf_{n\to\infty} \max_{P_{X^n}} \frac{1}{n} E_o(\rho, P_{X^n}) \tag{7.17}$$

with

$$E_o(\rho, P_{X^n}) \triangleq -\log_2 \sum_{\mathbf{y}\in\mathcal{Y}^n} \left(\sum_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x}) W_{Y^n|X^n}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

is Gallager's channel function for the discrete channel $\mathbf{W_{Y|X}}$ with arbitrary memory.

We next specialize Gallager's lower bound for SEM source-channel pairs by using Lemma 7.3. We recall when the channel is symmetric (in the Gallager sense, see Section 5.3.3), which directly applies to channels with additive noise, the maximum in (7.17) is achieved by the uniform distribution: $P_{X^n}(\mathbf{x}) = 1/|\mathcal{X}|^n$ for all $\mathbf{x} \in \mathcal{X}^n$. Thus for our (modulo B) additive noise channels, $E_o(\rho)$ reduces to

$$E_o(\rho, \mathbf{Q_S}) - \tau E_s(\rho, \mathbf{W_{Y|X}}) = \rho \log_2 B - \limsup_{n\to\infty} \frac{(1+\rho)}{n} \log_2 \left(\sum_{\mathbf{z}\in\mathcal{Z}^n} P_{Z^n}(\mathbf{z})^{\frac{1}{1+\rho}} \right). \tag{7.18}$$

It immediately follows by Lemma 7.3 that for our SEM source-channel pair,

$$E_o(\rho, \mathbf{Q_S}) - \tau E_s(\rho, \mathbf{W_{Y|X}}) = \rho \log_2 B - \rho\tau \overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{Q_S}}(S) - \rho \overline{H}^{(\frac{1}{1+\rho})}_{\mathbf{P_Z}}(Z) = F(\rho). \tag{7.19}$$

That is, the SEM source-channel function we defined in (7.9) is exactly the same as the difference of Gallager's channel and source function. In light of Theorem 7.2, we obtain the following regarding the computation of E_J.

Theorem 7.3 *For an SEM source* $\mathbf{Q_S}$ *and an SEM channel* $\mathbf{W_{Y|X}}$ *with noise* $\mathbf{P_Z}$ *such that* $\tau\overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ *and* $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$*,* $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ *is positive and determined exactly by* $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = F(\rho^*)$ *if* $\rho^* \leq 1$*, where* ρ^* *is the smallest positive number satisfying the equation* $f(\rho^*) = 0$*. Otherwise (if* $\rho^* > 1$*), the following bounds hold:*

$$\log_2 B - 2\log_2\left[\lambda_{\frac{1}{2}}^\tau(\mathbf{Q_S})\lambda_{\frac{1}{2}}(\mathbf{P_Z})\right] \leq E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq F(\rho^*).$$

Remark 7.4 If $\tau\overline{H}_{\mathbf{Q_S}}(S) + \overline{H}_{\mathbf{P_Z}}(Z) \geq \log_2 B$, then $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 0$.

Remark 7.5 According to Lemma 7.4 (c) and (d), there must exist a positive and finite ρ^* provided that $\tau\overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ and $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$. Using Lemma 7.4 (a), such ρ^* can be numerically determined.

The proof of Theorem 7.3 directly follows from Theorem 7.2, the comparison of Gallager's lower bound, and Lemma 7.4.

Example 7.1 We consider a system consisting of a binary SEM source $\mathbf{Q_S}$ and a binary SEM channel $\mathbf{W_{Y|X}}$ with transmission rate $t = 1$, both with symmetric transition matrices given by

$$Q = \begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix} \quad \text{and} \quad P_Z = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix},$$

such that $0 < p, q < 1$. The upper and lower bounds for $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ are plotted as a function of parameters p and q in Fig. 7.2. It is observed that for this source-channel pair, the bounds are tight for a large class of (p, q) pairs. Only when p or q is extremely close to 0 or 1, is E_J not exactly known.

7.4.3 Error Exponents for SEM Sources and SEM Channels

In this section we investigate two special cases of Theorem 7.2. One special case is that when the SEM channel is a noiseless channel, and the other is that when the SEM source is uniform source. In the first case, the upper bound for E_J reduces to an upper bound for the SEM source error exponent, and in the second case, Theorem 7.2 reduces to an upper bound to the

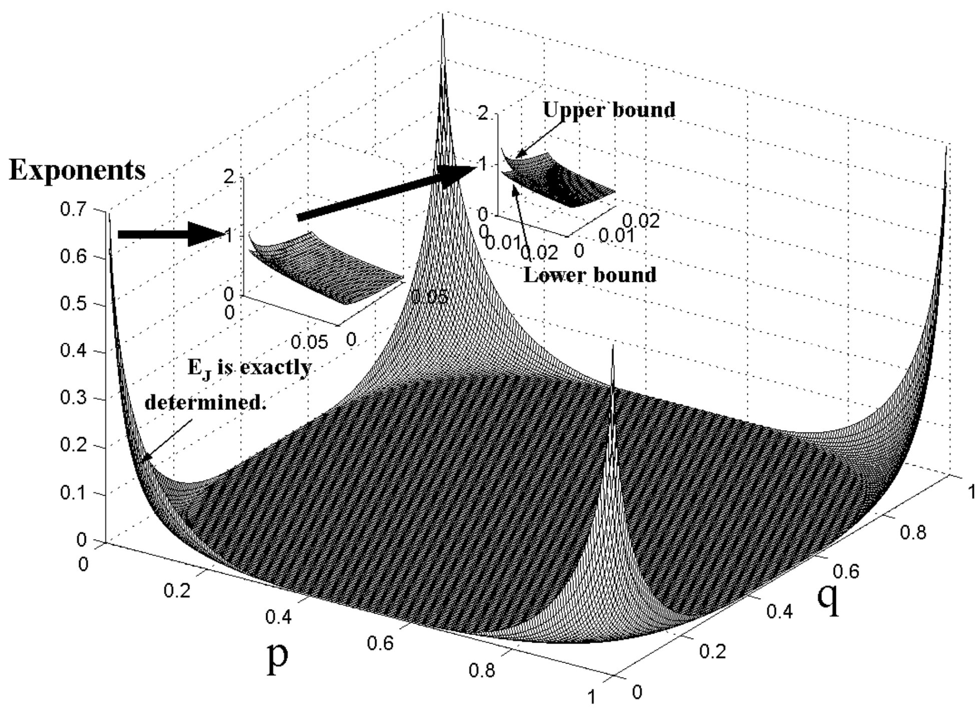

Figure 7.2: The lower and upper bounds of E_J for the binary SEM source and the binary SEM channel of Example 7.1 with $\tau = 1$.

SEM channel error exponent. Note that the SEM source error exponent has been studied in [72] and [94], and our results coincide with their results. The upper bound for the channel error exponent for SEM channels, however, has not been addressed before. We next need to extend the definition of the DMS error exponent for SEM sources; to be more general, we define the error exponent for arbitrarily discrete sources $\mathbf{Q_S} \triangleq \{Q_{S^{\tau n}} \in \mathcal{P}(\mathcal{S}^{\tau n})\}_{\tau n=1}^{\infty}$.

A (k, M_k) block source code for a discrete source $\mathbf{Q_S}$, which is defined the same as the block source code for a DMS, is a pair of mappings

$$f_{sk} : \mathcal{S}^k \longrightarrow \{1, 2, ..., M_k\}$$

and

$$\varphi_{sk} : \{1, 2, ..., M_k\} \longrightarrow \mathcal{S}^k.$$

The code rate is defined by

$$R_k \triangleq \frac{1}{k}\log_2 M_k \quad \text{bits/source symbol.}$$

The probability of erroneously reconstructing the source via the (k, M_k) block source code (f_{sk}, φ_{sk}) is given by

$$P_{se}^{(k)}(\mathbf{Q_S}, R_k) \triangleq \sum_{\mathbf{s}:\varphi_{sk}(f_{sk}(\mathbf{s}))\neq \mathbf{s}} Q_{S^k}(\mathbf{s}). \tag{7.20}$$

Definition 7.2 For any $R > 0$, the source error exponent $e(R, \mathbf{Q_S})$ of the discrete source $\mathbf{Q_S}$ is defined as the supremum of the set of all numbers e for which there exists a sequence of (k, M_k) block codes (f_{sk}, φ_{sk}) with

$$e \leq \liminf_{n\to\infty} -\frac{1}{n}\log P_{se}^{(k)}(\mathbf{Q_S}, R_k) \tag{7.21}$$

and

$$R \geq \limsup_{k\to\infty} R_k. \tag{7.22}$$

The following by-product result regarding the error exponent for SEM sources immediately follow from Theorem 7.2.

Corollary 7.1 [94] *For any rate* $0 < R < \log_2 \lambda_0(\mathbf{Q_S})$*, the source error exponent* $e(R, \mathbf{Q_S})$ *for an SEM source* $\mathbf{Q_S}$ *satisfies*

$$e(R, \mathbf{Q_S}) \leq \overline{e}(R, \mathbf{Q_S}), \tag{7.23}$$

where

$$\overline{e}(R, \mathbf{Q_S}) \triangleq \sup_{\rho\geq 0}[R\rho - (1+\rho)\log_2 \lambda_{\frac{1}{1+\rho}}(\mathbf{Q_S})]. \tag{7.24}$$

In particular, $e(R, \mathbf{Q_S}) = 0$ *for* $0 < R < \overline{H}_{\mathbf{Q_S}}(S)$.

Note that $\log_2 \lambda_0(\mathbf{Q_S}) = \log_2 |S|$ when the source reduces to a DMS (with alphabet $\mathcal{S}$). This upper bound is exactly the same as the one given by Vašek [94]. In fact, he shows that $\overline{e}(R, \mathbf{Q_S})$ is the real source error exponent (also see [25]) for all $R > 0$. We point out that $\overline{e}(R, \mathbf{Q_S})$ can be equivalently expressed in terms of a constrained minimum of Kullback-Leibler divergence [72], as the error exponent for DMS (2.4); also see (7.34) in the Section 7.4.5.

Similarly, before we specialize our bound to SEM channel error exponent, we first define the channel error exponent for an arbitrary discrete channel $\mathbf{W}_{\mathbf{Y}|\mathbf{Z}} \triangleq \{W_{Y^n|X^n} \in \mathcal{P}(\mathcal{X}^n \to \mathcal{Y}^n)\}_{n=1}^{\infty}$.

An (n, M_n) block channel code for a discrete channel $\mathbf{W}_{\mathbf{Y}|\mathbf{X}}$, which is defined the same as the block code for a DMC, is a pair of mappings

$$f_{cn} : \{1, 2, ..., M_n\} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_{cn} : \mathcal{Y}^n \longrightarrow \{1, 2, ..., M_n\}.$$

The code rate is defined as

$$R_n \triangleq \frac{1}{n} \log_2 M_n \quad \text{bits/channel use.}$$

The (average) probability of decoding error for the (f_{cn}, φ_{cn}) code is given by

$$P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_n) \triangleq \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y})=j} W_{Y^n|X^n}(\mathbf{y}|f_{cn}(i)). \tag{7.25}$$

Definition 7.3 For any $R > 0$, the channel error exponent $E(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$ of the discrete channel $\mathbf{W}_{\mathbf{Y}|\mathbf{X}}$ is defined as the supremum of the set of all numbers E for which there exists a sequence of (n, M_n) block codes (f_{cn}, φ_{cn}) with

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \ln P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_n)$$

and

$$R \leq \liminf_{n\to\infty} R_n.$$

The following by-product results regarding the error exponent for SEM channels immediately follow from Theorem 7.2.

Corollary 7.2 *For any rate* $\log_2(B/\lambda_0(\mathbf{P_Z})) < R < +\infty$*, the channel error exponent* $E(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$ *for an SEM channel* $\mathbf{W}_{\mathbf{Y}|\mathbf{X}}$ *with additive noise* $\mathbf{P_Z}$ *satisfies*

$$E(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}) \leq E_{sp}(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}), \tag{7.26}$$

where

$$E_{sp}(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}) \triangleq \sup_{\rho \geq 0} \left\{ \rho(\log_2 B - R) - (1+\rho) \log_2 \lambda_{\frac{1}{1+\rho}}(\mathbf{P}_{\mathbf{Z}}) \right\}. \quad (7.27)$$

In particular, $E(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}) = 0$ *for* $\overline{C}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}) \leq R < +\infty$.

When the SEM channel reduces to an additive noise DMC, $\log_2(B/\lambda_0(\mathbf{P}_{\mathbf{Z}})) = R_\infty$. Note that the usual case (when the transition matrix is positive) is that $\log_2(B/\lambda_0(\mathbf{P}_{\mathbf{Z}})) = 0$ (see Lemma 7.1). It can be shown that $E_{sp}(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$ is positive, non-increasing and convex, and hence strictly decreasing in R. Comparing with Gallager's random-coding lower bound for $E(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$ [42, Theorem 5.6.1] (when specialized for SEM channels) given by

$$E_r(R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}) \triangleq \max_{0 \leq \rho \leq 1} \left\{ \rho(\log_2 B - R) - (1+\rho) \log_2 \lambda_{\frac{1}{1+\rho}}(\mathbf{P}_{\mathbf{Z}}) \right\}, \quad (7.28)$$

and applying the results of Section 7.3, we note that the upper and lower bounds are tight if $R \geq R_{cr}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}})$, where

$$R_{cr} \triangleq \log_2 B - \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{Z}}^{(\frac{1}{2})}}(Z)$$

is the critical rate of the SEM channel. Thus, the channel error exponent for SEM channel is determined exactly for $R \geq R_{cr}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}})$.

Example 7.2 Consider a binary SEM channel $\mathbf{W}_{\mathbf{Y}|\mathbf{X}}$ with noise process

$$\mathbf{P}_{\mathbf{Z}} = \{P_{Z^n} \in \mathcal{P}(\mathcal{Z}^n)\}_{n=1}^{\infty}$$

whose transition matrix is given by

$$P_Z = \begin{bmatrix} \varepsilon + (1-\varepsilon)(1-p) & (1-\varepsilon)p \\ (1-\varepsilon)(1-p) & \varepsilon + (1-\varepsilon)p \end{bmatrix},$$

where $0 \leq \varepsilon < 1$ is the noise correlation coefficient and $0 < p < 1$. It is easy to see that the stationary distribution of the noise process is $\pi = [1-p, p]$. Note also that the SEM channel reduces to a (memoryless) BSC with crossover probability p when we choose $\varepsilon = 0$. In Fig. 7.3, we plot the upper and lower bounds for the SEM channel error exponent (7.27) and (7.28) for $p = 0.01$ and $\varepsilon = 0.5$, and for $p = 0.01$ and $\varepsilon = 0.9$. We also plot the upper and lower bounds for the BSC error exponent with crossover probability $p = 0.01$ for comparison.

It is seen that for the SEM channel with parameters $p = 0.01$ and $\varepsilon = 0.5$, the channel exponent is determined exactly between $R_{cr}(\mathbf{W_{Y|X}}) = 0.39$ and $C(\mathbf{W_{Y|X}}) = 0.95$; for the SEM channel with parameters $p = 0.01$ and high correlation $\varepsilon = 0.9$, the channel exponent is determined exactly between $R_{cr}(\mathbf{W_{Y|X}}) = 0.57$ and $C(\mathbf{W_{Y|X}}) = 0.98$; while for the BSC with $p = 0.01$, the channel exponent is determined exactly between $R_{cr}(W_{Y|X}) = 0.55$ and $C(W_{Y|X}) = 0.92$. For rates close to capacity, since the channel capacity of the SEM channel is generally larger than the capacity of BSC, the corresponding channel error exponent of SEM channel is larger; this is the case for example, when $0.75 < R < 0.95$, for the SEM channel exponent with $p = 0.01$ and $\varepsilon = 0.5$. However, for middle rates, when R is between 0.2 and 0.75, the BSC error exponent beats the SEM channel error exponent with $p = 0.01$ and $\varepsilon = 0.5$ (since the lower bound of the BSC error exponent is above the upper bound of the SEM channel exponent). In Fig. 7.4, we plot the upper and lower bounds for the SEM error exponent vs the noise correlation coefficient ε. When $p = 0.01$ and $R = 0.4$, we see that the upper and lower bounds coincide for $\epsilon \in [0.27, 0.65]$, and hence exactly determine the exponent $E(R, \mathbf{W_{Y|X}})$. When $p = 0.01$ and $R = 0.5$, the SEM channel error exponent is determined for $\epsilon \in [0.05, 0.83]$. When $p = 0.01$ and $R = 0.5$, the SEM channel error exponent is determined for $\epsilon \in [0, 0.93]$. This example demonstrates that the channel noise memory does not necessarily increase the channel error exponent and the critical rate (as seen in Figs. 7.3 and 7.4). However, we also stress that since in general the SEM capacity is larger than the capacity of the BSC with the same parameter p, there might be a considerable gain of error exponent at a rate below the SEM capacity. For instance, we plot the upper and lower bounds for the SEM channel error exponent for $p = 0.1$ and $\varepsilon = 0.8$, and the BSC error exponent with crossover probability $p = 0.1$. As shown in Fig. 7.5, in this case, the capacity of the SEM channel ($C(\mathbf{W_{Y|X}}) = 0.79$) is noticeably larger than the capacity of the BSC ($C(W_{Y|X}) = 0.52$). Thus, as expected, the SEM error exponent considerably outperforms the BSC error exponent for a wide range of rates below the SEM capacity.

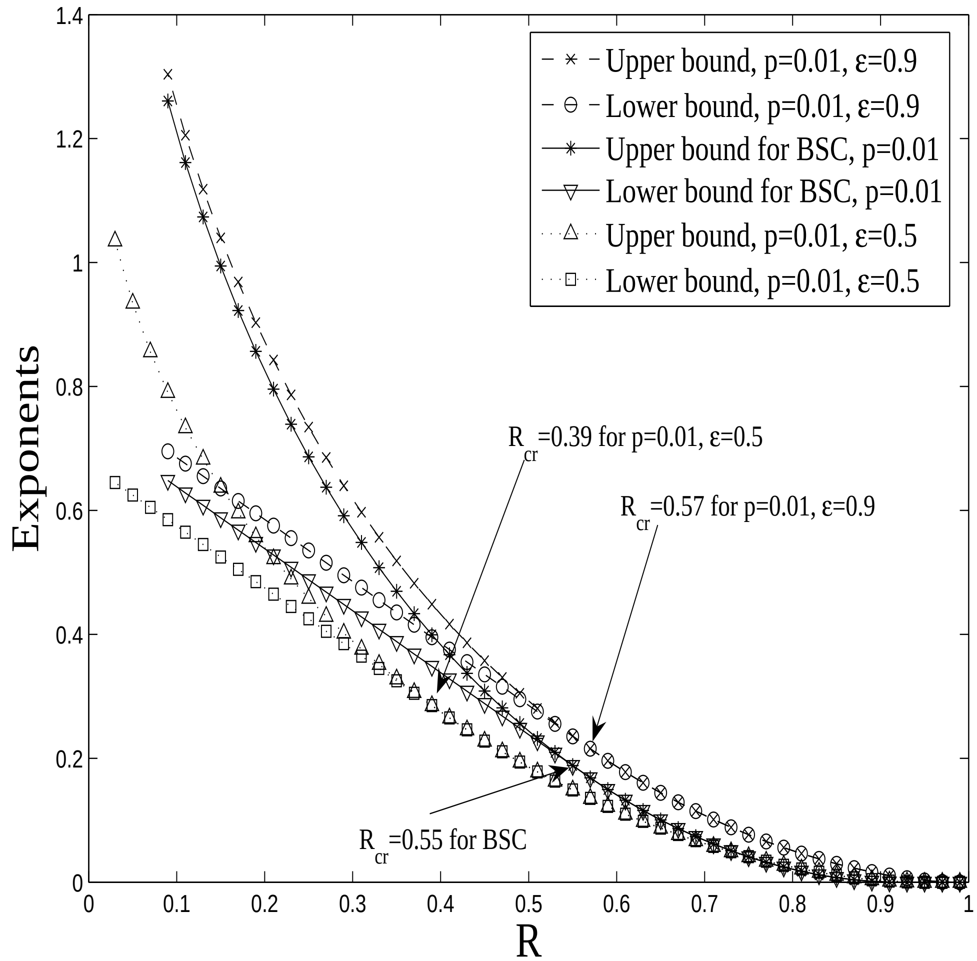

Figure 7.3: The upper and lower bounds for the channel error exponents of Example 7.2 in terms of rates R for $p = 0.01$.

7.4.4 Equivalent Bounds

One may next ask if the lower and upper bounds for the SEM source-channel pair enjoy a form that is similar to Csiszár's bounds for DMS-DMC pairs, $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ given by (5.5) and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ given by (5.6), which are expressed as the minimum of the sum of the source error exponent and the lower/upper bound of the channel error exponent. The answer is indeed affirmative, as given in the following theorem.

Theorem 7.4 *Let* $\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ *and* $\lambda_0^{\tau}(\mathbf{Q_S})\lambda_0(\mathbf{P_Z}) > B$. *The following equiva-*

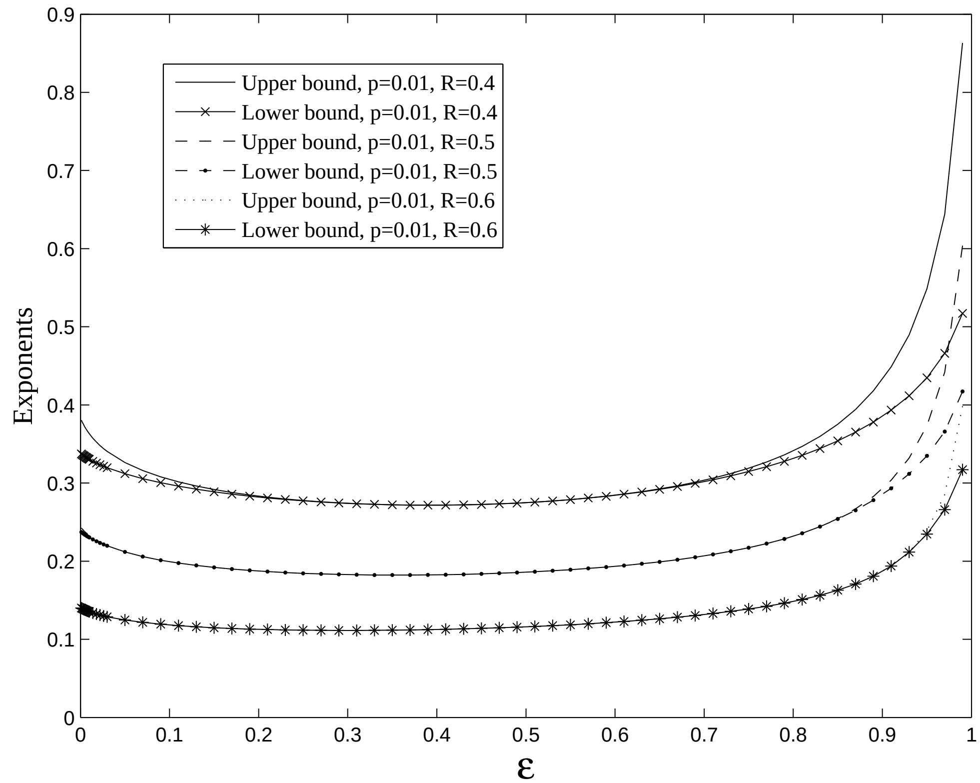

Figure 7.4: The upper and lower bounds for the channel error exponents of Example 7.2 in terms of ε.

lent representations hold

$$\max_{\rho\geq 0} F(\rho) = \min_{\log_2(B/\lambda_0(\mathbf{P_Z}))<R<\tau\log_2(\lambda_0(\mathbf{Q_S}))} \left[\tau e\left(\frac{R}{\tau},\mathbf{Q_S}\right)+E_{sp}(R,\mathbf{W_{Y|X}})\right], \quad (7.29)$$

$$\max_{0\leq\rho\leq 1} F(\rho) = \min_{0<R<\tau\log_2(\lambda_0(\mathbf{Q_S}))} \left[\tau e\left(\frac{R}{\tau},\mathbf{Q_S}\right)+E_r(R,\mathbf{W_{Y|X}})\right]. \quad (7.30)$$

where $F(\rho)$ *is defined in (7.9),* $e(R,\mathbf{Q_S})=\overline{e}(R,\mathbf{Q_S})$ *is given in Corollary 7.1,* $E_{sp}(R,\mathbf{W_{Y|X}})$ *and* $E_r(R,\mathbf{W_{Y|X}})$ *are given in Corollary 7.2 and (7.28).*

Remark 7.6 The assumption $\lambda_0^{\tau}(\mathbf{Q_S})\lambda_0(\mathbf{P_Z})>B$ ensures that the right-hand side of (7.29) is finite and the minimum is attained by some $R\in(\log_2(B/\lambda_0(\mathbf{P_Z})),\tau\log_2\lambda_0(\mathbf{Q_S}))$.

Theorem 7.4 is proved in a similar manner as Theorem 5.1 based on Fenchel duality theorem. Note that here the equivalent expressions for these bounds can also be

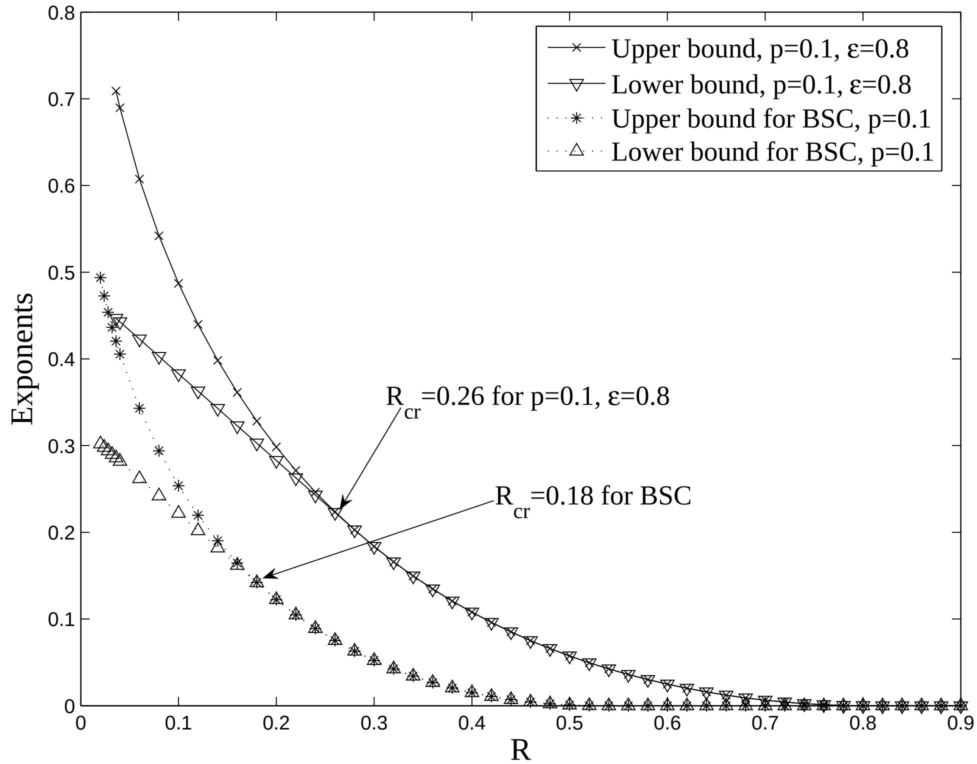

Figure 7.5: The upper and lower bounds for the channel error exponents of Example 7.2 in terms of rates R for $p = 0.1$.

proved via the technique of Lagrange multipliers, since the functions $\log_2 \lambda_{1/(1+\rho)}(\mathbf{Q_S})$ and $\log_2 \lambda_{1/(1+\rho)}(\mathbf{P_Z})$ are differentiable functions of ρ and their derivatives admit closed-form expressions (recall Lemma 7.3). When the source $\mathbf{Q_S}$ and channel $\mathbf{W_{Y|X}}$ are discrete memoryless, the right-hand side of (7.29) and (7.30) reduce to Csiszár's lower and upper bounds $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$ and $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$.

We point out that the parametric expressions of these bounds (the left-hand side of (7.29) and (7.30)) facilitate the computation of E_J, while the bounds in Csiszár's form (the right-hand side of (7.29) and (7.30)) are instrumental for the comparison of JSC and tandem coding exponents.

7.4.5 Markov Types and A Conceptual Upper Bound

Recall that Csiszár originally establishes the upper bound for E_J for a DMS-DMC pair $(Q_S, W_{Y|X})$ by the exact source and channel exponents $e(R, Q_S)$ and $E(R, W_{Y|X})$,

$$E_J(Q_S, W_{Y|X}, \tau) \leq \min_R \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X})\right] \tag{7.31}$$

and Csiszár's source-channel sphere-packing bound $\overline{E}_{Jsp}$ is obtained by replacing the channel error exponent $E(R, W_{Y|X})$ by its sphere-packing upper bound $E_{sp}(R, W_{Y|X})$. In Section 6.1.3, we proved that the JSCC error exponent with feedback can be upper bounded by

$$E_{Jfb}(Q_S, W_{Y|X}, \tau) \leq \min_R \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E_{fb}(R, W_{Y|X})\right] \tag{7.32}$$

where $E_{fb}(R, W_{Y|X})$ is the channel error exponent for a DMC with feedback. We also pointed out in Observation 6.1 that this kind of conceptual upper bound, expressed in terms of source and channel error exponent, holds for many discrete systems, as long as we can partition the source space by polynomial numbers of type classes, and we can rewrite the averaged probability of error for each type class as a channel coding probability error. Now for the SEM source-channel pairs, this is still true, i.e., E_J is upper bounded by the minimum of the sum of the SEM source exponent $e(R, \mathbf{Q_S})$ and the SEM channel exponent $E(R, \mathbf{W_{Y|X}})$.

In the following we prove this conceptual bound for SEM system by introducing Markov types. Like (7.31) and (7.32), the bound in terms of $e(R, \mathbf{Q_S})$ and $E(R, \mathbf{W_{Y|X}})$, though tighter than the sphere-packing type bound (7.29), is not computable in general, since the behavior of the SEM channel error exponent $E(R, \mathbf{W_{Y|X}})$ is unknown for rates smaller than the critical rate $R_{cr}(\mathbf{W_{Y|X}})$. In Chapter 10, we shall use this bound to prove that the JSCC error exponent can at most double the tandem coding exponent.

We first set up some notations and basic facts regarding Markov types adopted from [34] and [72]. Given a source sequence $\mathbf{s} = (s_1, s_2, ..., s_k) \in \mathcal{S}^k$ ($|\mathcal{S}| = M$), let $k_{ij}(\mathbf{s})$ be the number of transitions from $i \in \mathcal{S}$ to $j \in \mathcal{S}$ in $\mathbf{s}$ with the cyclic convention that s_1 follows s_k. We denote the matrix

$$\left[\frac{k_{ij}(\mathbf{s})}{k}\right]_{M\times M}$$

by $\Phi^{(k)}(\mathbf{s})$ and call it the Markov type (empirical matrix) of $\mathbf{s}$, where $\sum_{i,j} k_{ij}(\mathbf{s}) = k$ and it is easily seen that $\sum_j k_{ij} = \sum_j k_{ji}$ for all i. In other words, the (k-length) sequence $\mathbf{s}$ of type P (which is an $M \times M$ matrix) has the empirical matrix $\Phi^{(k)}(\mathbf{s})$ which is equal to P. The set of all types of k-length sequences will be denoted by $\mathcal{E}_k$. Next we introduce a class of matrices that includes $\mathcal{E}_k$ for all k as a dense subset. Let

$$\mathcal{E} = \left\{ P : P = [p_{ij}]_{M\times M}, \sum_{i,j} p_{ij} = 1, \quad \text{and} \quad p_{ij} \geq 0, \quad \sum_j p_{ij} = \sum_j p_{ji} \quad \text{for all } i \right\}.$$

Note that $\mathcal{E}_k \to \mathcal{E}$ as $k \to \infty$ in the sense that for any $P \in \mathcal{E}$, there exists a sequence of $\{\Phi^{(k)}\} \in \mathcal{E}_k$, such that $\Phi^{(k)} \to P$ uniformly.

For $P \in \mathcal{E}$ and any $M \times M$ transition (stochastic) matrix $Q = [q_{ij}]_{M\times M}$ (such that $\sum_j q_{ij} = 1$ for all i), define

$$H_c(P) \triangleq -\sum_{i,j} p_{ij} \log \frac{p_{ij}}{\sum_j p_{ij}}$$

be the conditional entropy of P and

$$D_c(P \parallel Q) \triangleq \sum_{i,j} p_{ij} \log \frac{p_{ij}}{q_{ij} \sum_j p_{ij}}$$

be the conditional divergence of P over Q. Let $P \in \mathcal{E}_k$ be a Markov type, and let

$$\mathbb{T}_P = \left\{ \mathbf{s} \in \mathcal{S}^k : \Phi^{(k)}(\mathbf{s}) = P \right\}$$

be a Markov type class. We define

$$\mathcal{M}_P(i,j) \triangleq \{\mathbf{s} = (s_1, s_2, ..., s_k) \in \mathbb{T}_P : s_1 = i, s_k = j\}.$$

Clearly, $\mathcal{M}_P(i,j)$ partitions the entire type class $\mathbb{T}_P$ over $(i,j) \in \mathcal{S} \times \mathcal{S}$, and all sequences in $\mathcal{M}_P(i,j)$ are equiprobable under $Q_{S^k}(\cdot)$.

Lemma 7.5 [34] Let $\mathbf{Q_S}$ be a first-order finite-alphabet irreducible Markov source with transition matrix $Q = [q_{ij}]_{M\times M}$ and arbitrary initial distribution $\mathbf{q} > 0$. Let $\alpha \triangleq \min_i q_i$. Then we have the following bounds.

(1) For any $i, j \in \mathcal{S}$ and $P \in \mathcal{E}_k$ such that $\mathcal{M}_P(i,j) \neq \emptyset$, $|\mathcal{M}_P(i,j)| \geq k^{-M}(k+1)^{-M^2} 2^{kH_c(P)}$.

(2) $Q_{S^k}(\mathbb{T}_P) \geq k^{-M}(k+1)^{-M^2} \alpha 2^{-kD_c(P\|Q)}$.

Remark 7.7 Remark that in [34], the authors assume both irreducibility and aperiodicity for the Markov source $\mathbf{Q_S}$ and also derive an upper bound for the probability of type classes $Q_{S^k}(\mathbb{T}_P)$. Here we only need the lower bound above for $Q_{S^k}(\mathbb{T}_P)$; thus the aperiodicity assumption is not required.

Note also that M and α are quantities independent of k, and that for SEM sources, the stationary distribution (which is the initial distribution) is unique and positive.

Theorem 7.5 *For an SEM source* $\mathbf{Q_S}$ *and a discrete channel* $\mathbf{W_{Y|X}}$ *with additive SEM noise* $\mathbf{P_W}$ *such that* $\tau\overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$*, the JSCC error exponent* $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ *satisfies*

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq \inf_{\tau\overline{H}_{\mathbf{Q_S}}(S) \leq R \leq \tau \log_2 \lambda_0(\mathbf{Q_S})} \left[\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right) + E(R, \mathbf{W_{Y|X}})\right] \tag{7.33}$$

Proof: We know from [34] that the source error exponent for the SEM source admits the following divergence form

$$e\left(\frac{R}{t}, \mathbf{Q_S}\right) = \min_{P \in \mathcal{E}: H_c(\mathbb{T}_P) \geq R/t} D_c(P \parallel Q) = \min_{P \in \mathcal{E}: H_c(\mathbb{T}_P) = R/t} D_c(P \parallel Q), \tag{7.34}$$

which is an equivalent representation of $\overline{e}(R, \mathbf{Q_S})$ given in Corollary 7.1 (see [72]), where the second equality of (7.34) follows from the strict monotonicity of $e(R, \mathbf{Q_S})$ in the interval $[\overline{H}_{\mathbf{Q_S}}(S), \log_2 \lambda_0(\mathbf{Q_S})]$. Thus, we can write

$$\begin{aligned} &\inf_{\tau\overline{H}_{\mathbf{Q_S}}(S) \leq R \leq \tau \log_2 \lambda_0(\mathbf{Q_S})} \left[\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right) + E(R, \mathbf{W_{Y|X}})\right] \\ &= \inf_{P \in \mathcal{E}} \left[\tau D_c(P \parallel Q) + E\left(\tau H_c(P), \mathbf{W_{Y|X}}\right)\right]. \end{aligned} \tag{7.35}$$

We assume that the above is finite (the upper bound is trivial if it is infinity) and the infimum actually becomes a minimum. Let the minimum be achieved by a matrix (joint distribution) $P^* \in \mathcal{E}$, then there must exist a sequence of Markov types $\left\{\widehat{P}_S \in \mathcal{E}_{\tau n}\right\}_{n=n_o}^{\infty}$ such that $\widehat{P} \to P^*$ uniformly. Next rewrite the probability of error given in (7.1) as a sum of probabilities of types and lower bound it by

$$
\begin{aligned}
& P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \\
&= \sum_{P \in \mathcal{E}_{\tau n}} \sum_{\mathbf{s} \in \mathbb{T}_P} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\
&\geq \sum_{\mathbf{s} \in \mathbb{T}_{\widehat{P}}} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\
&= \sum_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} \sum_{\mathbf{s} \in \mathcal{M}_{\widehat{P}}(i,j)} Q_{S^{\tau n}}(\mathbf{s}) \sum_{\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\
&= \sum_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} \left(\sum_{\mathbf{s}' \in \mathcal{M}_{\widehat{P}}(i,j)} Q_{S^{\tau n}}(\mathbf{s}') \right) \\
&\qquad \sum_{\mathbf{s} \in \mathcal{M}(i,j)} \frac{Q_{S^{\tau n}}(\mathbf{s})}{\sum_{\mathbf{s}' \in \mathcal{M}_{\widehat{P}}(i,j)} Q_{S^{\tau n}}(\mathbf{s}')} \sum_{\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})) \\
&= \sum_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} \sum_{\mathbf{s}' \in \mathcal{M}_{\widehat{P}}(i,j)} Q_{S^{\tau n}}(\mathbf{s}') P_e(\mathcal{M}_{\widehat{P}}(i,j)) \\
&\geq \sum_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} \sum_{\mathbf{s}' \in \mathcal{M}_{\widehat{P}}(i,j)} Q_{S^{\tau n}}(\mathbf{s}') \min_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} P_e(\mathcal{M}_P(i,j)) \\
&= Q_{S^{\tau n}}(\mathbb{T}_{\widehat{P}}) \min_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} P_e(\mathcal{M}_{\widehat{P}}(i,j)), \qquad (7.36)
\end{aligned}
$$

where

$$
P_e(\mathcal{M}_{\widehat{P}}(i,j)) \triangleq \frac{1}{|\mathcal{M}_{\widehat{P}}(i,j)|} \sum_{\mathbf{s} \in \mathcal{M}_{\widehat{P}}(i,j)} \sum_{\mathbf{y} \in A_{\mathbf{s}}^c} W_{Y^n|X^n}(\mathbf{y}|f_n(\mathbf{s})).
$$

We note that $P_e(\mathcal{M}_{\widehat{P}}(i,j))$ is actually the (average) probability of error of the n-block *channel* code (f_n, φ_n) with message set (source) $\mathcal{M}_{\widehat{P}}(i,j)$ and channel $\mathbf{W_{Y|X}}$. Now setting $R_n(i,j) = \frac{1}{n} \log_2 |\mathcal{M}_{\widehat{P}}(i,j)|$, by the definition of the channel error exponent for SEM channels (Definition 7.3) and Lemma 7.5

$$
\liminf_{n \to \infty} -\frac{1}{n} \log_2 P_e(\mathcal{M}_{\widehat{P}}(i,j)) \leq E\left(\liminf_{n \to \infty} R_n(i,j), \mathbf{W_{Y|X}} \right) = E\left(\tau H_c(\widehat{P}), \mathbf{W_{Y|X}} \right)
$$

for any sequence of JSC codes (f_n, φ_n) and for any $(i,j) \in \mathcal{S} \times \mathcal{S}$ such that $\mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset$. It then follows from (7.36) and Lemma 7.5 again that

$$
\begin{aligned}
& \liminf_{n \to \infty} -\frac{1}{n} \log_2 P_e^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \\
&\leq \liminf_{n \to \infty} \left[-\frac{1}{n} \log_2 Q_{S^{\tau n}}(\mathbb{T}_{\widehat{P}}) - \frac{1}{n} \log_2 \min_{(i,j) \in \mathcal{S} \times \mathcal{S}: \mathcal{M}_{\widehat{P}}(i,j) \neq \emptyset} P_e(\mathcal{M}_{\widehat{P}}(i,j)) \right]
\end{aligned}
$$

$$\leq \limsup_{n\to\infty} -\frac{1}{n}\log_2 Q_{S^{\tau n}}(\mathbb{T}_{\widehat{P}}) + \liminf_{n\to\infty} -\frac{1}{n}\log_2 \min_{(i,j)\in\mathcal{S}\times\mathcal{S}:\mathcal{M}_{\widehat{P}}(i,j)\neq\emptyset} P_e(\mathcal{M}_{\widehat{P}}(i,j))$$
$$\leq \tau D_c(P^* \parallel Q) + E\left(\tau H_c(P^*), \mathbf{W}_{\mathbf{Y}|\mathbf{X}}\right).$$

Since the above bound holds for any sequence of JSC codes, we complete the proof of Theorem 7.5. ■

7.5 Systems with Arbitrary Markovian Orders

Suppose that the SEM source $\{U_i\}_{n=1}^{\infty}$ with alphabet $\mathcal{U}$ has a Markovian order $K_s \geq 1$. Define process $\{S_i\}_{n=1}^{\infty}$ obtained by K_s-step blocking the Markov source, i.e.,

$$S_n \triangleq (U_n, U_{n+1}, ..., U_{n+K_s-1}).$$

Then

$$\Pr(S_n = j_n | S_{n-1} = j_{n-1}, ..., S_1 = j_1) = \Pr(S_n = j_n | S_{n-1} = j_{n-1}), \quad j_1, ..., j_n \in \mathcal{S} = \mathcal{U}^{K_s}$$

and the source $\mathbf{Q_S} = \{Q_{S^n} \in \mathcal{P}(\mathcal{S}^n)\}_{n=1}^{\infty}$ is a first order SEM source with $|\mathcal{U}|^{K_s}$ states. Therefore, all the results in this paper can be readily extended to SEM systems with arbitrary order by converting the K_s-th order SEM source to a first order SEM source of larger alphabet. Also, if the additive SEM noise $\mathbf{P_Z}$ of the channel has Markovian order $K_c \geq 1$, we can similarly convert it to a first order SEM noise with expanded alphabet. In Section 10.3 we will compare the JSCC error exponent with the tandem coding error exponent for a Markovian system consisting of an SEM source (of order $K_s = 1$) and the queue based channel (QBC) [101] with memory $K_c = 2$ (see Example 10.4).

7.6 Conclusion

In this chapter, we established a computable upper bound for the JSCC error exponent E_J for SEM source-channel systems. As special cases, the upper bound to E_J leads to an upper bound for the SEM source (channel) error exponent. We next examined Gallager's lower bound for E_J for the Markovian systems. The lower/upper bound can be expressed in terms of a maximum of the difference source and channel functions, and equivalently, can be expressed in terms of the minimum of the sum of the SEM source error exponent and the lower/upper bound of the SEM channel exponent. It was shown that E_J can be exactly determined by the two bounds for a large class of SEM source-channel pairs. We next established a conceptual upper bound for E_J in terms of SEM source and channel error exponents by introducing Markov types. This upper bound will be applied in Chapter 10 to compare the JSCC error exponent with the tandem coding error exponent.

Chapter 8

JSCC Excess Distortion Exponent for Memoryless Continuous-Alphabet Systems

In this chapter, we address the JSCC excess distortion exponent for a communication system consisting of a (stationary continuous) memoryless source Q_S with a distortion measure and a (stationary continuous) memoryless channel $W_{Y|X}$ with an input cost constraint. Specifically, we first focus on the memoryless Gaussian system and then extend our results to other continuous source-channel pairs such as the Laplacian-source Gaussian-channel pair, and a certain class of source-channel pairs when the distortion is a metric.

We first define the JSCC excess distortion exponent and formulate our problem in Section 8.1. In Section 8.2, we establish upper and lower bounds for the JSCC excess distortion exponent for Gaussian systems. For a Gaussian communication system consisting of an MGS Q_S with the squared-error distortion and an MGC $W_{Y|X}$ with additive noise P_Z and the power input constraint, we show that the JSCC excess distortion exponent $E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$ with transmission rate τ, under a distortion threshold Δ and power constraint $\mathcal{E}$, is upper bounded by the minimum of the sum of the MGS exponent $\tau F_G(R/\tau, Q_S, \Delta)$ defined in (2.46) and (2.47) and the sphere-packing upper bound of the

Gaussian channel error exponent $E_{sp}(R, W_{Y|X}, \mathcal{E})$ given by (2.51). The proof of the upper bound relies on a strong converse JSCC theorem and the judicious construction of an auxiliary MGS and an auxiliary MGC to lower bound the probability of excess distortion. We also establish a lower bound for $E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$. As a matter of fact, we derive the lower bound for MGSs and general continuous memoryless channels with an input cost constraint. To prove the lower bound, we employ a concatenated "quantization – lossless JSCC" scheme as in [7], use the type covering lemma for Gaussian-type classes (Lemma 3.6), and then bound the probability of error for the lossless JSCC part, which involves a memoryless source with a countably infinite alphabet and the memoryless continuous channel, by using a modified version of Gallager's random-coding bound for the JSCC error exponent for DMS-DMC pairs [42, Problem 5.16] (the modification is made to allow for input cost constrained channels with countably-infinite input alphabets and continuous output alphabets). This lower bound is expressed by the maximum of the difference of Gallager's constrained-input channel function $E_0(W_{Y|X}, \mathcal{E}, \rho)$ given in (2.40) and the source function $\tau E(Q_S, \Delta, \rho)$ given in (4.7). Note that when the channel is an MGC with an input power constraint, a computable but somewhat looser lower bound is obtained by replacing $E_0(W_{Y|X}, \mathcal{E}, \rho)$ by Gallager's Gaussian-input channel function $\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho)$ given by (2.54). Also we remind that the source function $E(Q_S, \Delta, \rho)$ for the MGS is equal to the guessing exponent [6] and admits an explicit analytic form (4.8).

As in our previous chapters for discrete systems (Chapters 5 and 7), we derive equivalent expressions for the lower and upper bounds by applying Fenchel duality theorem. We show that the upper bound, though proved in the form of a minimum of the sum of source and channel exponents, can also be represented as a (dual) maximum of the difference of Gallager's channel function $\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho)$ and the source function $\tau E(Q_S, \Delta, \rho)$. Analogously, the lower bound, which is established in Gallager's form, can also be represented in Csiszár's form, as the minimum of the sum of the source exponent and the lower bound of the channel exponent. In this regard, our upper and lower bounds are natural extension of Csiszár's upper and lower bounds from the case of (finite alphabet) discrete memoryless systems to the case of memoryless Gaussian systems. We then compare the upper and lower bounds using their equivalent forms and derive explicit analytical conditions for which the two bounds coincide.

We next observe that upper and lower bounds for $E_J^{\Delta,\mathcal{E}}$ can also be proved for memoryless Laplacian sources (MLSs) under the magnitude-error distortion measure. In Section 8.3, using a similar approach, we establish upper and lower bounds for the JSCC excess distortion exponent for the lossy transmission of MLSs over MGCs.

In Section 8.4, we considerably modify our approach in light of the result of [55] to prove a lower bound for some continuous source-channel pairs when the distortion is a metric. We show that the lower bound for MGSs and continuous memoryless channels, expressed by the maximum of the difference of source and channel functions, still holds for a continuous source-channel pair if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E}\exp[td(s,s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, where the expectation is taken over the source distribution defined on $\mathbb{R}$ (see Theorem 8.7). Although the condition does not hold for both MGSs with the squared-error distortion and MLSs with the magnitude-error distortion, it holds for generalized MGSs with parameters (α, σ) under the distortion $d(x,y) = |x-y|^p$, $p < \alpha$, and $p \leq 1$. Finally, we draw a conclusion in Section 8.5.

8.1 Definitions and System Description

8.1.1 Notation

Since only continuous systems are treated in this chapter, we assume that the source and channel alphabets are all real space, i.e., $\mathcal{S} = \mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}$. The source distribution Q_S is a valid pdf on $\mathcal{S}$, and the channel transition distribution $W_{Y|X}$ is a valid conditional pdf on $\mathcal{X} \times \mathcal{Y}$. We next introduce some new notation for this chapter. $o(n)$ serves as a generic notation for a vanishing quantity with respect to n such that $\lim_{n\to\infty} o(n)/n = 0$. $\zeta(\epsilon)$ serves as a generic notation for a vanishing quantity with respect to ϵ such that $\lim_{\epsilon\to 0} \zeta(\epsilon) = 0$.

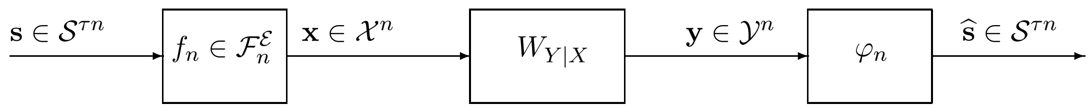

Figure 8.1: JSCC system consisting of a memoryless source and a memoryless channel with continuous alphabets.

8.1.2 JSCC System and JSCC Excess Distortion Exponent

Given a source distribution measure $d(\cdot,\cdot) : \mathcal{S} \times \mathcal{S} \to [0,\infty)$ and a channel input function $g(\cdot) : \mathcal{X} \to [0,\infty)$, a JSC code $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$ with blocklength n and transmission rate τ (source symbols/channel use) for the memoryless source Q_S, and the memoryless channel $W_{Y|X}$ with input cost constraint $\mathcal{E}$ is a pair of mappings (see Fig. 8.1):

$$f_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{X}^n$$

and

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n},$$

where $f_n \in \mathcal{F}_n^{\mathcal{E}}$, and

$$\mathcal{F}_n^{\mathcal{E}} \triangleq \left\{ f_n : \frac{1}{n}\sum_{i=1}^{n} g(x_i) \le \mathcal{E} \quad \text{for all} \quad \mathbf{x} = f_n(\mathbf{s}) \right\}. \tag{8.1}$$

Here $\mathbf{s} \in \mathcal{S}^{\tau n}$ is the transmitted source message and $\mathbf{x} = f_n(\mathbf{s}) \in \mathcal{X}^n$ is the corresponding n-length codeword. The conditional pdf of receiving $\mathbf{y} \in \mathcal{Y}^n$ at the channel output given that the message $\mathbf{s}$ is transmitted is given by

$$W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})) = \prod_{i=1}^{n} W_{Y|X}(y_i|x_i).$$

The probability of failing to decode the JSC code $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$ within a prescribed distortion level $\Delta > 0$ is called the probability of excess distortion and defined by

$$P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \triangleq \int_{\mathcal{S}^{\tau n}} Q_S^{(\tau n)}(\mathbf{s}) \int_{\mathbf{y}: d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta} W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{y} d\mathbf{s}.$$

Definition 8.1 The JSCC excess distortion exponent $E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$ for the above memoryless source Q_S and memoryless channel $W_{Y|X}$ is defined as the supremum of the set of all numbers E for which there exists a sequence of source-channel codes $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$ with blocklength n such that

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \ln P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, t).$$

When there is no possibility of confusion, throughout the sequel the JSCC excess distortion exponent $E_J(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$ will be written as $E_J^{\Delta,\mathcal{E}}$.

8.2 JSCC Excess Distortion Exponent for Gaussian Systems

We now focus on the communication system consisting of an MGS and an MGC ($\mathcal{S} = \mathcal{X} = \mathcal{Y} = \mathbb{R}$) with squared-error distortion measure and input power constraint. We establish an upper and a lower bound for the JSCC excess distortion exponent for the Gaussian system in Sections 8.2.2 and 8.2.4. The bounds will be evaluated in Section 8.2.5,

8.2.1 A Strong Converse (Lossy) JSCC Theorem

We first derive a strong converse JSCC theorem under the probability of excess distortion criterion for the Gaussian system. We use later this result to obtain an upper bound for the excess distortion exponent $E_J^{\Delta,\mathcal{E}}$.

Theorem 8.1 *(Strong Converse JSCC Theorem) For an MGS Q_S and an MGC $W_{Y|X}$, if $\tau R(Q_S, \Delta) > C(W_{Y|X}, \mathcal{E})$, then*

$$\lim_{n\to\infty} P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) = 1$$

for any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$.

Proof: Assume that $C(W_{Y|X}, \mathcal{E}) = \tau R(Q_S, \Delta) - \varepsilon$, where ε is a positive number. For some δ $(0 < \delta < \varepsilon)$ define

$$\widetilde{A} = \left\{ (\mathbf{s}, \mathbf{y}) : \ln \frac{W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})) P^{*(\tau n)}_{S'}(\varphi_n(\mathbf{y}))}{P^{*(n)}_{Y}(\mathbf{y}) P^{*(\tau n)}_{S'|S}((\varphi_n(\mathbf{y}))|\mathbf{s})} \leq n \left(C(W_{Y|X}, \mathcal{E}) - \tau R(Q_S, \Delta) + \delta \right) \right\},$$

where $P^{*(\tau n)}_{S'|S}$ and $P^{*(\tau n)}_{S'}$ are the $\tau n-$dimensional product distributions corresponding to $P^*_{S'|S}$ and $P^*_{S'}$ given in (2.43) and (2.44) respectively, and $P^{*(n)}_{Y}$ is the $n-$dimensional product distribution corresponding to P^*_Y given in (2.50). Recalling that

$$P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) = 1 - \Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\}\right), \tag{8.2}$$

where the probability is with respect to the joint distribution $Q_S^{(\tau n)}(\cdot)W_{Y|X}^{(n)}(\cdot|\cdot)$, it suffices to show that the probability $\Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\}\right)$ approaches 0 asymptotically for any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$. We decompose it as follows

$$\begin{aligned}
&\Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\}\right) \\
&= \Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\} \bigcap \widetilde{A}\right) + \Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\} \bigcap \widetilde{A}^c\right).
\end{aligned} \tag{8.3}$$

For the first probability in (8.3), we can bound it by using the property of set $\widetilde{A}$

$$\begin{aligned}
&\Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s}, \varphi_n(\mathbf{y})\right) \leq \Delta\right\} \bigcap \widetilde{A}\right) \\
&= \int_{\left\{(\mathbf{s},\mathbf{y}): d^{(\tau n)}(\mathbf{s}, \varphi_n(\mathbf{y})) \leq \Delta\right\} \bigcap \widetilde{A}} Q_S^{(\tau n)}(\mathbf{s}) W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})) d\mathbf{s} d\mathbf{y} \\
&\leq \int_{\left\{(\mathbf{s},\mathbf{y}): d^{(\tau n)}(\mathbf{s}, \varphi_n(\mathbf{y})) \leq \Delta\right\} \bigcap \widetilde{A}} e^{n\left(C(W_{Y|X}, \mathcal{E}) - \tau R(Q_S, \Delta) + \delta\right)} \\
&\qquad \cdot Q_S^{(\tau n)}(\mathbf{s}) \frac{P^{*(n)}_{Y}(\mathbf{y}) P^{*(\tau n)}_{S'|S}((\varphi_n(\mathbf{y}))|\mathbf{s})}{P^{*(\tau n)}_{S'}(\varphi_n(\mathbf{y}))} d\mathbf{s} d\mathbf{y} \\
&\leq e^{-n(\varepsilon - \delta)} \int_{\mathcal{Y}^n} \frac{P^{*(n)}_{Y}(\mathbf{y})}{P^{*(\tau n)}_{S'}(\varphi_n(\mathbf{y}))} \underbrace{\int_{\mathbf{s}: d^{(\tau n)}(\mathbf{s}, \varphi(\mathbf{y})) \leq \Delta} Q_S^{(\tau n)}(\mathbf{s}) P^{*(\tau n)}_{S'|S}(\varphi_n(\mathbf{y})|\mathbf{s}) d\mathbf{s}}_{\leq P^{*(\tau n)}_{S'}(\varphi_n(\mathbf{y}))} d\mathbf{y} \\
&\leq e^{-n(\varepsilon - \delta)} \int_{\mathcal{Y}^n} P^{*(n)}_{Y}(\mathbf{y}) d\mathbf{y} \\
&= e^{-n(\varepsilon - \delta)}.
\end{aligned} \tag{8.4}$$

It remains to bound the second probability in (8.3). Using the expressions of the pdf's, we have

$$\begin{aligned}
\frac{1}{n} \ln \frac{W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s})) P^{*(\tau n)}_{S'}(\varphi_n(\mathbf{y}))}{P^{*(n)}_{Y}(\mathbf{y}) P^{*(\tau n)}_{S'|S}((\varphi_n(\mathbf{y}))|\mathbf{s})} &= C(W_{Y|X}, \mathcal{E}) + \frac{\mathbf{y}^T \mathbf{y}}{2n(\mathcal{E} + \sigma_Z^2)} - \frac{\mathbf{z}^T \mathbf{z}}{2n\sigma_Z^2} \\
&- \tau R(Q_S, \Delta) + \frac{\tau d^{(\tau n)}(\varphi_n(\mathbf{y}), \mathbf{s})}{2\Delta} - \frac{\mathbf{s}^T \mathbf{s}}{2n\sigma_S^2}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s},\varphi_n(\mathbf{y})\right)\le\Delta\right\}\bigcap\widetilde{A}^c\right)\\
=\;&\Pr\left(\left\{d^{(\tau n)}\left(\mathbf{s},\varphi_n(\mathbf{y})\right)\le\Delta\right\}\bigcap\left\{\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}\right.\right.\\
&\left.\left.+\frac{\tau d^{(\tau n)}(\varphi_n(\mathbf{y}),\mathbf{s})}{2\Delta}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right\}\right)\\
\le\;&\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}+\frac{\tau}{2}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right)\\
\le\;&\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right)+\Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2}-1<-\frac{2\delta}{3}\right)+\Pr\left(\frac{\mathbf{s}^T\mathbf{s}}{n\sigma_S^2}-\tau<-\frac{2\delta}{3}\right)\\
&. \qquad (8.5)
\end{aligned}$$

It suffices to show

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right)=0, \tag{8.6}$$

$$\lim_{n\to\infty}\Pr\left(-\frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2}+1<-\frac{2\delta}{3}\right)=0, \tag{8.7}$$

and

$$\lim_{n\to\infty}\Pr\left(-\frac{\mathbf{s}^T\mathbf{s}}{n\sigma_S^2}+\tau<-\frac{2\delta}{3}\right)=0. \tag{8.8}$$

Clearly, (8.7) and (8.8) follow by the weak law of large numbers (WLLN), noting that $\mathbf{s}$ and $\mathbf{z}$ are memoryless sequences. To derive (8.6), we write, as in the proof of [70, Lemma 4])

$$\begin{aligned}
&\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{n(\mathcal{E}+\sigma_Z^2)}-1>\frac{2\delta}{3}\right)\\
=\;&\Pr\left(\frac{\mathbf{x}^T\mathbf{x}}{n}+\frac{\mathbf{z}^T\mathbf{z}}{n}+\frac{2\mathbf{x}^T\mathbf{z}}{n}-(\mathcal{E}+\sigma_Z^2)>\frac{2\delta}{3}(\mathcal{E}+\sigma_Z^2)\right)\\
\le\;&\Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n}+\frac{2\mathbf{x}^T\mathbf{z}}{n}-\sigma_Z^2>\frac{2\delta}{3}(\mathcal{E}+\sigma_Z^2)\right)\\
\le\;&\Pr\left(\frac{\mathbf{z}^T\mathbf{z}}{n}-\sigma_Z^2>\frac{\delta}{3}(\mathcal{E}+\sigma_Z^2)\right)+\Pr\left(\frac{2\mathbf{x}^T\mathbf{z}}{n}>\frac{\delta}{3}(\mathcal{E}+\sigma_Z^2)\right),
\end{aligned} \tag{8.9}$$

where the first inequality follows from the power constraint (8.1), the first probability in (8.9) converges to zero as $n\to\infty$ by the WLLN and the second probability in (8.9) converges to zero as $n\to\infty$ by the WLLN, the fact the $\mathbf{z}$'s have zero mean, and the independence of $\mathbf{x}$ and $\mathbf{z}$. Thus, (8.6), (8.7) and (8.8) yield

$$\lim_{n\to\infty}\Pr\left(\frac{\mathbf{y}^T\mathbf{y}}{2n(\mathcal{E}+\sigma_Z^2)}-\frac{\mathbf{z}^T\mathbf{z}}{2n\sigma_Z^2}+\frac{\tau}{2}-\frac{\mathbf{s}^T\mathbf{s}}{2n\sigma_S^2}>\delta\right)=0. \tag{8.10}$$

On account of (8.4), (8.10) and (8.2), we complete the proof. ■

Note that the above theorem also holds for a slightly wider class of MGCs with scaled inputs, described by $Y_i = bX_i + Z_i$ (X_i and Z_i are independent from each other), and with transition pdf

$$W_{Y|X}(y|x) = P_Z(y - bx) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{(y-bx)^2}{2\sigma_Z^2}},$$

where b is a nonzero constant. We next apply this result to prove the upper bound of $E_J^{\Delta,\mathcal{E}}$. It follows from Theorem 8.1 that the JSCC excess distortion exponent is 0 if the source rate-distortion function is larger than the channel capacity, i.e., $\tau R(Q_S, \Delta) > C(W_{Y|X}, \mathcal{E})$. We thus confine our attention to the case of $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$ in the following theorem.

8.2.2 The Upper Bound

Theorem 8.2 *For an MGS Q_S and an MGC $W_{Y|X}$ with $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$, the JSCC excess distortion exponent satisfies*

$$E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \leq \overline{E}_{Jsp}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau), \tag{8.11}$$

where

$$\overline{E}_{Jsp}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \triangleq \min_{\tau R(Q_S,\Delta) \leq R \leq C(W_{Y|X},\mathcal{E})} \left[\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R, W_{Y|X}, \mathcal{E}) \right], \tag{8.12}$$

where (R, Q_S, Δ) is the MGS exponent given in (2.46) and (2.47) and $E_{sp}(R, W_{Y|X}, \mathcal{E})$ is the sphere-packing bound of the channel error exponent for an MGC $W_{Y|X}$ given in (2.51).

Proof: For any sufficiently small $\varepsilon > 0$, fix an $R \in [\tau R(Q_S, \Delta) + \varepsilon, C(W_{Y|X}, \mathcal{E})]$. Define an auxiliary MGS for this R with alphabet $\mathcal{S} = \mathbb{R}$ and distribution $\widetilde{Q}_S \sim \mathcal{N}(0, \widetilde{\sigma}_S^2)$, where $\widetilde{\sigma}_S^2 \triangleq \Delta e^{2R/\tau}$, so that the rate-distortion function of $\widetilde{Q}_S$ is given by

$$R(\widetilde{Q}_S, \Delta) = \frac{1}{2} \ln \max\left\{\frac{\widetilde{\sigma}_S^2}{\Delta}, 1\right\} = \frac{R}{\tau}.$$

Also, it can be easily verified that the Kullback-Leibler divergence between the auxiliary MGS $\widetilde{Q}_S$ and the original source Q_S is

$$D(\widetilde{Q}_S \parallel Q_S) = \frac{1}{2}\left(\frac{\widetilde{\sigma}_S^2}{\sigma_S^2} - \ln\frac{\widetilde{\sigma}_S^2}{\sigma_S^2} - 1\right) = F_G\left(\frac{R}{\tau}, Q_S, \Delta\right).$$

Next we define for $R' \triangleq R - \frac{\varepsilon}{2} > 0$ an auxiliary MGC with scaled inputs $\widetilde{W}_{Y|X}$ associated with the original MGC $W_{Y|X}$ with the alphabets $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and transition pdf

$$\widetilde{W}_{Y|X}(y|x) \triangleq \frac{1}{\sqrt{2\pi\widetilde{\sigma}_Z^2}} e^{-\frac{(y+ax)^2}{2\widetilde{\sigma}_Z^2}}$$

where the parameter a is uniquely determined by β' ($\beta' = e^{2R'}$) and SNR as follows

$$a \triangleq \frac{-\text{SNR}(\beta'-1) - \sqrt{\text{SNR}^2(\beta'-1)^2 + 4\text{SNR}\beta'(\beta'-1)}}{2\text{SNR}\beta'} < 0, \tag{8.13}$$

and

$$\widetilde{\sigma}_Z^2 \triangleq \frac{a^2\mathcal{E}}{\beta'-1}. \tag{8.14}$$

It can be verified that the capacity of the MGC $\widetilde{W}$ is given by

$$C(\widetilde{W}_{Y|X}, \mathcal{E}) = \sup_{P_X:\mathbb{E}X^2 \le \mathcal{E}} I(X;Y) = \frac{1}{2}\ln\left(1 + \frac{a^2\mathcal{E}}{\widetilde{\sigma}_Z^2}\right) = R',$$

where the supremum is achieved by the Gaussian distribution $P_X = P_X^*$ given in (2.49).

For some $\delta > 0$, define the set

$$\widehat{A} \triangleq \left\{(\mathbf{s},\mathbf{y}) : \ln\frac{\widetilde{Q}_S^{(\tau n)}(\mathbf{s})\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))}{Q_S^{(\tau n)}(\mathbf{s})W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))} \le n\left(\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R', W_{Y|X}, \mathcal{E}) + \delta\right)\right\}.$$

Consequently, we can use $\widehat{A}$ to lower bound the probability of excess distortion of any sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, \tau)$,

$$\begin{aligned}
&P_\Delta^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\
&\ge \int_{\left\{(\mathbf{s},\mathbf{y}):d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta\right\}\cap\widehat{A}} Q_S^{(\tau n)}(\mathbf{s})W_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
&\ge e^{-n\left(\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R', W_{Y|X}, \mathcal{E}) + \delta\right)} \\
&\quad \int_{\left\{(\mathbf{s},\mathbf{y}):d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta\right\}\cap\widehat{A}} \widetilde{Q}_S^{(\tau n)}(\mathbf{s})\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y},
\end{aligned} \tag{8.15}$$

and the last integration can be decomposed as

$$\begin{aligned}
&\int_{\{(\mathbf{s},\mathbf{y}):d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta\}\cap\widehat{A}} \widetilde{Q}_S^{(\tau n)}(\mathbf{s})\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
&\geq \int_{(\mathbf{s},\mathbf{y}):d^{(\tau n)}(\mathbf{s},\varphi_n(\mathbf{y}))>\Delta)} \widetilde{Q}_S^{(\tau n)}(\mathbf{s})\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
&\quad - \int_{\widehat{A}^c} \widetilde{Q}_S^{(\tau n)}(\mathbf{s})\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|f_n(\mathbf{s}))d\mathbf{s}d\mathbf{y} \\
&= P_\Delta^{(n)}(\widetilde{Q}_S,\widetilde{W}_{Y|X},\mathcal{E},\tau) - \Pr\left(\widehat{A}^c\right),
\end{aligned} \tag{8.16}$$

where the probabilities are with respect to the joint distribution $\widetilde{Q}_S^{(\tau n)}(\cdot)\widetilde{W}_{Y|X}^{(n)}(\cdot|\cdot)$. Note that the first term in the right-hand side of (8.16) is exactly the probability of excess distortion for the joint source-channel system consisting of the auxiliary MGS $\widetilde{Q}_S$ and the auxiliary MGC $\widetilde{W}_{Y|X}$ with transmission τ, and, according to our setting, with

$$\tau R(\widetilde{Q}_S,\Delta) = R > R' = C(\widetilde{W}_{Y|X},\mathcal{E}).$$

Thus, this quantity converges to 1 as n goes to infinity according to the strong converse JSCC theorem. It remains to show that the second term in the right-hand side of (8.16) vanishes asymptotically. Note that

$$\begin{aligned}
\Pr\left(\widehat{A}^c\right) &\leq \Pr\left(\frac{1}{\tau n}\ln\frac{\widetilde{Q}_S^{(\tau n)}(\mathbf{s})}{Q_S^{(\tau n)}(\mathbf{s})} > F_G\left(\frac{R}{\tau},Q_S,\Delta\right)+\frac{\delta}{2\tau}\right) \\
&\quad +\Pr\left(\frac{1}{n}\ln\frac{\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})}{W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})} > E_{sp}(R',W_{Y|X},\mathcal{E})+\frac{\delta}{2}\right).
\end{aligned} \tag{8.17}$$

It follows by the WLLN that as $n\to\infty$,

$$\frac{\widetilde{Q}_S^{(\tau n)}(\mathbf{s})}{Q_S^{(\tau n)}(\mathbf{s})} \longrightarrow \mathbb{E}_{\widetilde{Q}_S}\left[\ln\frac{\widetilde{Q}_S(s)}{Q_S(s)}\right] = F_G\left(\frac{R}{\tau},Q_S,\Delta\right) \qquad \text{in Prob.,}$$

which implies that

$$\lim_{n\to\infty}\Pr\left(\frac{1}{\tau n}\ln\frac{\widetilde{Q}_S^{(\tau n)}(\mathbf{s})}{Q_S^{(\tau n)}(\mathbf{s})} > F_G\left(\frac{R}{\tau},Q_S,\Delta\right)+\frac{\delta}{2\tau}\right) = 0. \tag{8.18}$$

For the second term of (8.17), setting $\mathbf{z}=\mathbf{y}+a\mathbf{x}$, we can write

$$\frac{1}{n}\ln\frac{\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})}{W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})} = \frac{1}{2}\left[\ln\frac{\sigma_Z^2}{\widetilde{\sigma}_Z^2} - \frac{\mathbf{z}^T\mathbf{z}}{n\widetilde{\sigma}_Z^2} + \frac{\mathbf{z}^T\mathbf{z}}{n\sigma_Z^2} - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} + \frac{(a+1)^2\mathbf{x}^T\mathbf{x}}{n\sigma_Z^2}\right].$$

On the other hand, recalling that a is given in (8.13) and $\widetilde{\sigma}_Z^2$ is given in (8.14), and noting that

$$\begin{aligned}\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} &= \frac{\mathrm{SNR}(\beta'-1)+2\beta'+\sqrt{\mathrm{SNR}^2(\beta'-1)^2+4\mathrm{SNR}\beta'(\beta'-1)}}{2\beta'^2} \\ &= \frac{4\beta'^2}{2\beta'^2[\mathrm{SNR}(\beta'-1)+2\beta'-\sqrt{\mathrm{SNR}^2(\beta'-1)^2+4\mathrm{SNR}\beta'(\beta'-1)}]} \\ &= \frac{2}{2\beta'+\mathrm{SNR}(\beta'-1)\left[1-\sqrt{1+\frac{4\beta'}{\mathrm{SNR}(\beta'-1)}}\right]},\end{aligned}$$

where $\beta' = e^{2R'}$, we see that

$$\begin{aligned}&\frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] \\ &= \frac{\mathrm{SNR}}{4\beta'}\left[(\beta'+1)-(\beta'-1)\sqrt{1+\frac{4\beta'}{\mathrm{SNR}(\beta'-1)}}\right] \\ &\quad +\frac{1}{2}\ln\left\{\beta' - \frac{\mathrm{SNR}(\beta'-1)}{2}\left[\sqrt{1+\frac{4\beta'}{\mathrm{SNR}(\beta'-1)}}-1\right]\right\},\end{aligned}$$

which is exactly the sphere-packing bound $E_{sp}(R', W_{Y|X}, \mathcal{E})$. Therefore, it suffices to show that

$$\begin{aligned}&\Pr\left(\frac{1}{n}\ln\frac{\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})}{W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})} > \frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] + \frac{\delta}{2}\right) \\ = \ &\Pr\left[\left(\frac{1}{\sigma_Z^2} - \frac{1}{\widetilde{\sigma}_Z^2}\right)\left(\frac{\mathbf{z}^T\mathbf{z}}{n} - \widetilde{\sigma}_Z^2\right) - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\left(\frac{\mathbf{x}^T\mathbf{x}}{n} - \mathcal{E}\right) > \delta\right]\end{aligned}$$

converges to 0 as n goes to infinity. This is true (as before) since the above probability is less than

$$\Pr\left[\left(\frac{1}{\sigma_Z^2} - \frac{1}{\widetilde{\sigma}_Z^2}\right)\left(\frac{\mathbf{z}^T\mathbf{z}}{n} - \widetilde{\sigma}_Z^2\right) - \frac{2(a+1)\mathbf{x}^T\mathbf{z}}{n\sigma_Z^2} > \delta\right] \tag{8.19}$$

by the power constraint (8.1), and $\mathbf{z}^T\mathbf{z}/n \to \widetilde{\sigma}_Z^2$ and $\mathbf{x}^T\mathbf{z}/n \to 0$ in probability 1. This yields

$$\lim_{n\to\infty}\Pr\left(\frac{1}{n}\ln\frac{\widetilde{W}_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})}{W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})} \le \frac{1}{2}\left[\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} - \ln\frac{\widetilde{\sigma}_Z^2}{\sigma_Z^2} + \frac{(a+1)^2}{\sigma_Z^2}\mathcal{E} - 1\right] + \frac{\delta}{2}\right) = 0. \tag{8.20}$$

On account of (8.15), (8.16), (8.18) and (8.20), we obtain

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_\Delta^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \le \tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}\left(R - \frac{\varepsilon}{2}, W_{Y|X}, \mathcal{E}\right) + \delta.$$

Since the above inequality holds for any rate R in the region $[\tau R(Q_S, \Delta) + \varepsilon, C(W_{Y|X}, \mathcal{E})]$ and δ and ε can be arbitrarily small, we obtain that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n} \ln P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\
&\leq \min_{\tau R(Q_S,\Delta)\leq R\leq C(W_{Y|X},\mathcal{E})} \left[\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R, W_{Y|X}, \mathcal{E})\right]. \qquad (8.21)
\end{aligned}$$

■

Since the MGS exponent $\tau F_G(R/\tau, Q_S, \Delta)$ is convex increasing for $R \geq \tau R(Q_S, \Delta)$ and the sphere-packing bound $E_{sp}(R, W_{Y|X}, \mathcal{E})$ is convex decreasing in $R \leq C(W_{Y|X}, \mathcal{E})$, their sum is also convex and there exists a global minimum in the interval $[\tau R(Q_S, \Delta), C(W_{Y|X}, \mathcal{E})]$ for the upper bound given in (8.11). For $R \in [\tau R(Q_S, \Delta), C(W_{Y|X}, \mathcal{E})]$, setting

$$\tau \frac{\partial F_G\left(\frac{R}{\tau}, Q_S, \Delta\right)}{\partial R} + \frac{\partial E_{sp}(R, W_{Y|X}, \mathcal{E})}{\partial R} = 0,$$

which gives (cf. Lemma 2.1)

$$\frac{\beta^{\frac{1}{\tau}}}{\text{SDR}} = \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}}\right), \qquad (8.22)$$

where $\text{SDR} \triangleq \sigma_S^2/\Delta$ is called the source-to-distortion ratio (i.e., the source variance to distortion threshold ratio), and $\beta = e^{2R}$. Thus, the minimum of the upper bound is achieved by the R which is the (unique) root of (8.22).

8.2.3 Gallager's Lower Bound for Lossless JSCC Error Exponent

In this section, we modify Gallager's upper bound for the error probability of JSCC for discrete memoryless systems, so that it is applicable to a JSCC system consisting of a DMS and a continuous memoryless channel with cost constraint $\mathcal{E}$. In the next section we shall apply this auxiliary bound to a system consisting of an MGS and a memoryless channel.

A JSC code $(\widetilde{f}_n, \widetilde{\varphi}_n)$ [107] for a DMS P_C and a continuous MC with transition pdf $W_{Y|X}$ is a pair of mappings $\widetilde{f}_n : \mathcal{C} \longrightarrow \mathcal{X}^n$ and $\widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \mathcal{C}$, where $\mathcal{C} \subseteq \mathcal{S}^{\tau n}$. That is, each source message $\mathbf{s} \in \mathcal{C}$ with pmf $P_C(\mathbf{s})$ is encoded as blocks $\mathbf{x} = \widetilde{f}_n(\mathbf{s})$ of symbols from $\mathcal{X}$ of length n, transmitted, received as blocks $\mathbf{y}$ of symbols from $\mathcal{Y}$ of length n and decoded as source symbol $\widetilde{\varphi}_n(\mathbf{y}) \in \mathcal{S}$. Denote the codebook for the codewords be $\mathcal{B} \triangleq \{\mathbf{x} = \widetilde{f}(\mathbf{s})\}$. The probability of decoding error is

$$P_e^{(n)}(P_C, W_{Y|X}) = P_e^{(n)}(P_C, W_{Y|X}, \mathcal{B}) \triangleq \sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s}) \int_{\mathbf{y}\in\mathcal{Y}^n} W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})\, \mathbb{1}\{\widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{s}\} d\mathbf{y}.$$

We next recast Gallager's random-coding bound for the JSCC probability of error [42, Problem 5.16] for DMS's and continuous MC's and we show the following bound.

Proposition 8.1 For each $n \geq 1$, given pdf P_{X^n} defined on $\mathcal{X}^n \subseteq \mathbb{R}^n$, there exists a sequence of JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n)$ such that for any $0 \leq \rho \leq 1$ the probability of error is upper bounded by

$$P_e^{(n)}(P_C, W) \leq \left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho} \int_{\mathbf{y}\in\mathcal{Y}^n} \left[\int_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x}) W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{x}. \quad (8.23)$$

Proof: The bound is shown analogously to [42, Problem 5.16] based on a random-coding argument. Consider the following random encoder: for each source message $\mathbf{s}$, we independently generate a codeword $\mathbf{x}^n$, which are $\mathbb{R}^n$-valued vectors, according to pdf P_{X^n}. So the codebook $\Pr(\mathcal{B})$ is generated with pdf $\Pr(\mathcal{B}) = \prod_{\widehat{\mathbf{x}}\in\mathcal{B}} P_{X^n}(\widehat{\mathbf{x}})$. Consider a maximum a-posteriori probability (MAP) decoder, which, given $\mathbf{y}$, chooses the source message $\mathbf{s}$ that maximizes $P_C(\mathbf{s}) W_{Y|X}^{(n)}\left(\mathbf{y}|\widetilde{f}_n(\mathbf{s})\right)$. In the following, we will bound the averaged probability of error over all possible codebooks $\mathcal{B}$, under the MAP decoding rule by

$$\begin{aligned}
\overline{P_e^{(n)}}(P_C, W, \mathcal{B}) &= \int_{\mathcal{B}} \Pr(\mathcal{B}) P_e^{(n)}(P_C, W_{Y|X}, \mathcal{B}) d\mathcal{B} \\
&\leq \left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho} \int_{\mathbf{y}\in\mathcal{Y}^n} \left[\int_{\mathbf{x}\in\mathcal{X}^n} P_{X^n}(\mathbf{x}) W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{x}.
\end{aligned} \quad (8.24)$$

Then from (8.24) we can conclude that, there must exist a sequence of JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n)$ such that (8.23) is valid.

It remains to show (8.24). Given source message $\mathbf{s} \in \mathcal{C}$, codeword $\mathbf{x} = \widetilde{f}_n(\mathbf{s})$, and received $\mathbf{y}$, define the event for an $\mathbf{s}' \neq \mathbf{s}$ by

$$E_{\mathbf{s}'} : P_C(\mathbf{s}') W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x}')) \geq P_C(\mathbf{s}) W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})),$$

where $\mathbf{x}' = \widetilde{f}_n(\mathbf{s}')$. Thus, when message $\mathbf{s} \in \mathcal{C}$ is sent, an error can occur only if $E_{\mathbf{s}'}$ occurs for some $\mathbf{s}' \neq \mathbf{s}$. This yields the following upper bound

$$\Pr\left(\{\widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{s}\} | \,\mathbf{s}, \mathbf{x}, \mathbf{y}\right) \leq \Pr\left(\bigcup_{\mathbf{s}' \neq \mathbf{s}} E_{\mathbf{s}'}\right) \leq \sum_{\mathbf{s}' \neq \mathbf{s}} \Pr\left(E_{\mathbf{s}'}\right) \leq \left[\sum_{\mathbf{s}' \neq \mathbf{s}} \Pr\left(E_{\mathbf{s}'}\right)\right]^{\rho} \tag{8.25}$$

for any $\rho \in [0,1]$. On the other hand, since the codebook $\mathcal{B}$ is generated according to a pdf $\Pr(\mathcal{B}) = \prod_{\mathbf{x}' \in \mathcal{B}} P_{X^n}(\mathbf{x}')$, by the definition of $E_{\mathbf{s}'}$, we have

$$\begin{aligned}
\Pr\left(E_{\mathbf{s}'}\right) &= \int_{\mathcal{B}} \Pr(\mathcal{B}) \mathbb{1}\left\{P_C(\mathbf{s}') W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}'\right)\right) \geq P_C(\mathbf{s}) W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}\right)\right)\right\} d\mathcal{B} \\
&= \int_{\mathbf{x}' \in \mathcal{X}^n} P_{X^n}(\mathbf{x}') \mathbb{1}\left\{P_C(\mathbf{s}') W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}'\right)\right) \geq P_C(\mathbf{s}) W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}\right)\right)\right\} d\mathbf{x} \\
&\leq \int_{\mathbf{x}' \in \mathcal{B}} P_{X^n}(\mathbf{x}') \left(\frac{P_C(\mathbf{s}') W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}'\right))}{P_C(\mathbf{s}) W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}\right))}\right)^{\frac{1}{1+\rho}} d\mathbf{x}.
\end{aligned} \tag{8.26}$$

Plugging (8.26) into (8.25), we obtain

$$\begin{aligned}
\overline{P_e^{(n)}}(P_C, W, \mathcal{B}) &= \int_{\mathcal{B}} \Pr(\mathcal{B}) \sum_{\mathbf{s} \in \mathcal{C}} P_C(\mathbf{s}) \int_{\mathbf{y} \in \mathcal{Y}^n} W_{Y|X}^{(n)}\left(\mathbf{y}|\mathbf{x}\right) \Pr\left(\{\widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{s}\} | \,\mathbf{s}, \mathbf{x}, \mathbf{y}\right) d\mathcal{B} \\
&= \left[\sum_{\mathbf{s} \in \mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho} \int_{\mathbf{y} \in \mathcal{Y}^n} \left[\int_{\mathbf{x} \in \mathcal{X}^n} P_{X^n}(\mathbf{x}) W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{x}.
\end{aligned}$$

■

Next, we need a small modification of (8.23) for the DMS P_C and the MC $W_{Y|X}$ to incorporate the channel input cost constraint. Let P_X^* be an arbitrary pdf of the channel input on $\mathcal{X}$ satisfying $\mathbb{E}g(X) \leq \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$ (these restrictions are made to make the term $\left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho}$ in (8.27) grow sub-exponentially with respect to n) and let $P{*}_X^{(n)}$ be the corresponding n-dimensional pdf on sequences of n channel inputs, i.e., the product pdf of P_X^*. We then adopt the technique of Gallager [42, Chapter 7], by setting $P_{X^n}(\mathbf{x}) = \kappa^{-1}\Phi(\mathbf{x}) P{*}_X^{(n)}(\mathbf{x})$, where

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \text{if } n\mathcal{E} - \eta \leq \sum_{i=1}^n g(x_i) \leq n\mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

in which $\eta > 0$ is arbitrary, and $\kappa = \int_{\mathbf{x}} P{*}_X^{(n)}(\mathbf{x})\Phi(\mathbf{x}) d\mathbf{x}$ is a normalizing constant. Thus, P_{X^n} is a valid probability density that satisfies the constraint (8.1). We thus have, for any

$r \geq 0$,

$$P_{X^n}(\mathbf{x}) \leq \kappa^{-1} e^{r\eta} P^{*(n)}_X(\mathbf{x}) e^{r[\sum_{i=1}^n g(x_i) - n\mathcal{E}]}.$$

Substituting the above into (8.23) for the memoryless channel $W_{Y|X}$, changing the summation to integration, and denoting the probability of error under constraint $\mathcal{E}$ by $P_e^{(n)}(Q_S, W_{Y|X}, \mathcal{E})$, we have

$$P_e^{(n)}(P_C, W_{Y|X}, \mathcal{E}) \leq \left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho} \left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right]^{1+\rho}$$

$$\times \int_{\mathbf{y}\in\mathcal{Y}^n} \left[\int_{\mathbf{x}\in\mathcal{X}^n} P^{*(n)}_X(\mathbf{x}) e^{r[\sum_{i=1}^n g(x_i) - n\mathcal{E}]} W^{(n)}_{Y|X}(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} d\mathbf{x}\right]^{1+\rho} d\mathbf{y}. \quad (8.27)$$

We remark that $\left[\frac{e^{r\eta}}{\kappa}\right]^{1+\rho}$ grows with n as $n^{(1+\rho)/2}$ and does not affect the exponential dependence of the bound on n [41], [42, pp. 326–333]. Thus, applying the upper bound for the DMS P_C and the memoryless channel W with cost constraint (8.1), and noting that P_X^* is an arbitrary pdf satisfying $\mathbb{E}g(X) \leq \mathcal{E}$ and $\mathbb{E}g(X)^3 < \infty$, we obtain

$$P_e^{(n)}(P_C, W_{Y|X}, \mathcal{E}) \leq \exp\left\{-n \max_{0\leq\rho\leq 1}\left[E_o(W_{Y|X}, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P_C)\right] + o(n)\right\}, \quad (8.28)$$

where $E_o(W_{Y|X}, \mathcal{E}, \rho)$ is the Gallager's constraint channel function given by (2.40), $o(n)$ has the form $c_1 \ln n + c_2$ for some constants c_1 and c_2, and $E_s^{(n)}(\rho, Q_S)$ is Gallager's source function

$$E_s^{(n)}(\rho, P_C) \triangleq \frac{1+\rho}{n} \ln\left[\sum_{\mathbf{s}\in\mathcal{C}} P_C(\mathbf{s})^{\frac{1}{1+\rho}}\right].$$

8.2.4 The Lower Bound

Theorem 8.3 *For an MGS Q_S and a continuous memoryless channel $W_{Y|X}$ with a cost constraint $\mathcal{E}$ at the channel input, the JSCC excess distortion exponent satisfies*

$$E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \geq E_{RC}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau), \quad (8.29)$$

where

$$E_{RC}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \triangleq \max_{0\leq\rho\leq 1}[E_o(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)], \quad (8.30)$$

where $E_o(W_{Y|X}, \mathcal{E}, \rho)$ is Gallager's constrained channel function given by (2.40) and $E(Q_S, \Delta, \rho)$ is the source function for the MGS Q_S given by (4.8). Furthermore, if $W_{Y|X}$ is an MGC, we have

$$E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \geq \underline{E}_{Jr}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau), \tag{8.31}$$

where

$$\underline{E}_{Jr}^{\Delta,\mathcal{E}}(Q_S, W, t) \triangleq \max_{0\leq\rho\leq 1}[\widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho) - tE(Q_S, \Delta, \rho)], \tag{8.32}$$

where $\widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)$ is Gallager's Gaussian-input channel function given by (2.54).

Proof: Fix $\tau > 0$. In the sequel we let $k = \tau n$ and assume that k (and hence n) is sufficiently large. For a given $\epsilon \in (0, \Delta)$ small enough, we construct a sequence of Gaussian-type classes $\mathbb{T}_i \triangleq \mathbb{T}^\epsilon(\sigma^2(i))$ by $\sigma^2(i) = \Delta + (2i-1)\epsilon$, $i = 1, 2, \cdots$. That is,

$$\begin{aligned} \mathbb{T}_i &\triangleq \left\{\mathbf{s} : \left|\mathbf{s}^T\mathbf{s} - k(\Delta + (2i-1)\epsilon)\right| \leq k\epsilon\right\} \\ &= \left\{\mathbf{s} : k(\Delta + (2i-2)\epsilon) \leq \mathbf{s}^T\mathbf{s} \leq k(\Delta + 2i\epsilon)\right\}, \qquad i = 1, 2, \cdots. \end{aligned} \tag{8.33}$$

Also, we define the set $\mathbb{T}_0 \triangleq \{\mathbf{s} : \mathbf{s}^T\mathbf{s} \leq k\Delta\}$ such that all these type classes $(\mathbb{T}_1, \mathbb{T}_2, \cdots)$ together with $\mathbb{T}_0$ partition the whole space $\mathbb{R}^k$. For this special set $\mathbb{T}_0$, we shall use the trivial bound $Q_S^{(k)}(\mathbb{T}_0) \leq 1$ and by definition $\mathbb{T}_0$ is covered by the ball $B(\mathbf{0}, \Delta)$; thus, we say that $\mathbb{T}_0$ satisfies the type covering lemma (Lemma 3.6) in the sense that there exists a set $\mathcal{C} \triangleq \{\mathbf{0}\}$ of size $|\mathcal{C}| = 1 \leq \exp\{k[R(\widehat{Q}_S, \Delta)]\}$ such that every $\mathbf{s} \in \mathbb{T}_0$ is covered by the the ball of size Δ, where we let $\widehat{Q}_S \sim \mathcal{N}(0, \Delta)$ and hence $R(\widehat{Q}_S, \Delta) = 0$.

Based on the above setup, we claim that, first, by Lemma 3.28, for all $i = 1, 2, \cdots$, the probability of $\mathbb{T}_i$ under the k-dimensional Gaussian pdf $Q_S^{(k)}$, denoted by $Q_S^{(k)}(\mathbb{T}_i)$, decays exponentially at the rate of $D(Q_S^{(i)} \parallel Q_S) + \widetilde{\zeta}_1(\epsilon)$ in k, where $Q_S^{(i)}$ is a zero-mean Gaussian source with variance $\sigma^2(i) = \Delta + (2i-1)\epsilon$, and

$$\widetilde{\zeta}_1(\epsilon) = -\frac{\epsilon}{\sigma_S^2} - \ln\left(1 + \frac{\epsilon}{\Delta}\right) \geq \zeta_1(\epsilon) \tag{8.34}$$

is a vanishing term independent of i. Second, the type covering lemma is applicable for all $\mathbb{T}_i$, $i = 1, 2, \cdots$. Note that when $\sigma^2(i) > \Delta$, $\zeta_2(\epsilon)$ in the type covering lemma (Lemma 3.6) can be bounded by

$$\zeta_2(\epsilon) \le \widetilde{\zeta}_2(\epsilon) \triangleq \frac{1}{2} \ln \frac{\Delta}{(\sqrt{\Delta} - \epsilon)^2 - 5\epsilon\Delta} + 2\epsilon + 2\ln[1 + \epsilon + 4\sqrt{\Delta\epsilon}] \tag{8.35}$$

and is also independent of i. In the sequel, we will denote, without loss of generality, that all these vanishing terms $\widetilde{\zeta}_1(\epsilon)$ and $\widetilde{\zeta}_2(\epsilon)$ by $\zeta(\epsilon)$.

We next employ a concatenated "quantization – lossless JSCC" scheme [7] to show the existence of a sequence of JSC codes for the source-channel pair (Q_S, W) such that its probability of excess distortion is upper bounded by

$$\exp[-nE_{RC}(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau) + o(n)]$$

for n sufficiently large.

First Stage Coding: Δ-admissible Quantization.

It follows from the above setup and the type covering lemma (Lemma 3.6) that for each $\mathbb{T}_i$ $(i = 1, 2, \cdots)$, there exists a code $\mathcal{C}_i = \{\mathbf{c}^{(i)}\}$ with codebook size $|\mathcal{C}_i| \le \exp\{k[R(Q_S^{(i)}, \Delta) + \zeta(\epsilon)] + o(k)\}$ that covers $\mathbb{T}_i$. Recall that we also have, trivially, that a code $\mathcal{C}_0 = \{\mathbf{0}\}$ with $|\mathcal{C}_0| = 1$ which covers $\mathbb{T}_0$. Therefore, we can employ a Δ-admissible quantizer via the sets $\mathcal{C}_i$, $i = 0, 1, 2, ...$ as follows:

$$f_{\Delta,k} : \mathbb{R}^k \longrightarrow \bigcup_{i=0}^{\infty} \mathcal{C}_i$$

such that for every $\mathbf{s} \in \mathbb{R}^k$, the output of $f_{\Delta,k}$ with respect to $\mathbf{s}$ has a distortion less than Δ. We denote the DMS at the output of $f_{\Delta,k}$ by P with alphabet $\bigcup_{i=0}^{\infty} \mathcal{C}_i$ and pmf

$$P(\mathbf{c}^{(i)}) = \int_{\mathbf{s} \in \mathbb{T}_i : f_{\Delta,k}(\mathbf{s}) = \mathbf{c}^{(i)}} Q_S^{(k)}(\mathbf{s}) d\mathbf{s}, \quad \forall \quad \mathbf{c}^{(i)} \in \mathcal{C}_i, \quad i = 0, 1, 2, ...$$

Second Stage Coding and Decoding: Lossless JSCC with Power Constraint $\mathcal{E}$.

For the DMS P and the continuous memoryless channel $W_{Y|X}$, a pair of (asymptotically) lossless JSC code

$$\widetilde{f}_n : \bigcup_{i=0}^{\infty} \mathcal{C}_i \longrightarrow \mathcal{X}^n \quad \text{and} \quad \widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \bigcup_{i=0}^{\infty} \mathcal{C}_i$$

is applied, where the encoder is subject to a cost constraint $\mathcal{E}$, i.e., $\widetilde{f}_n \in \mathcal{F}_n^{\mathcal{E}}$. Note that the decoder $\widetilde{\varphi}_n$ creates an approximation $\widehat{\mathbf{c}} = \widetilde{\varphi}_n(\mathbf{y})$ of $\mathbf{c}^{(i)}$ based upon the sequence $\mathbf{y}$ received at the channel output. According to a modified version of Gallager's JSCC random-coding bound (derived in the last section), there exists a sequence of lossless JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n, \mathcal{E})$ with bounded probability of error

$$\begin{aligned} P_e^{(n)}(P, W_{Y|X}, \mathcal{E}) &\triangleq \Pr(\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}) \\ &= \sum_{\bigcup_{i=0}^{\infty} \mathcal{C}_i} P(\mathbf{c}^{(i)}) \int_{\mathbf{y}:\widetilde{\varphi}_n(\mathbf{y}) \neq \mathbf{c}^{(i)}} W_{Y|X}^{(n)}\left(\mathbf{y} \,\middle|\, \widetilde{f}_n(\mathbf{c}^{(i)})\right) d\mathbf{y} \\ &\leq \exp\left\{-n \max_{0\leq\rho\leq 1}\left[E_o(W_{Y|X}, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P)\right] + o(n)\right\}, \end{aligned}$$

where

$$E_0(W_{Y|X}, \mathcal{E}, \rho) = \sup_{P_X:\mathbb{E}g(X)\leq\mathcal{E}} \max_{r\geq 0} E_0(\rho, r, W, P_X)$$

is Gallager's constrained channel function given in (2.40) and $E_s^{(n)}(\rho, P)$ is Gallager's source function here given by

$$E_s^{(n)}(\rho, P) = \frac{1+\rho}{k} \ln\left\{\sum_{i=0}^{\infty} \sum_{\mathbf{c}^{(i)}\in\mathcal{C}_i} P(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}}\right\}.$$

Probability of Excess Distortion.

According to the Δ-admissible quantization rule, if the distortion between the source message $\mathbf{s}$ and the reproduced sequence $\widehat{\mathbf{c}}$ is larger than Δ, then we must have $\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}$. This implies that

$$\begin{aligned} &P_\Delta^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\ &= \Pr\left(d^{(k)}(\widehat{\mathbf{c}}, \mathbf{s}) > \Delta\right) \\ &\leq \Pr\left(\widehat{\mathbf{c}} \neq \mathbf{c}^{(i)}\right) \\ &\leq \exp\left\{-n \max_{0\leq\rho\leq 1}\left[E_o(W_{Y|X}, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P)\right] + o(n)\right\}. \qquad (8.36) \end{aligned}$$

Next we bound $E_s^{(n)}(\rho, P)$ in terms of Q_S for k (also n) sufficiently large and when ϵ goes to zero (when N goes to infinity). Rewrite

$$\begin{aligned}
E_s^{(n)}(\rho, P) &= \frac{1+\rho}{k} \ln \left\{ \sum_{i=0}^{\infty} \sum_{\mathbf{c} \in \mathcal{C}_i} \left[Q_S^{(k)}(\mathbb{T}_i) P_{S^k}^{(i)}(\mathbf{c}^{(i)}) \right]^{\frac{1}{1+\rho}} \right\} \\
&= \frac{1+\rho}{k} \ln \left\{ \sum_{i=0}^{\infty} Q_S^{(k)}(\mathbb{T}_i)^{\frac{1}{1+\rho}} \sum_{\mathbf{c} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \right\} \\
&\le \frac{1+\rho}{k} \ln \left\{ 1 + \sum_{i=1}^{\infty} Q_S^{(k)}(\mathbb{T}_i)^{\frac{1}{1+\rho}} \sum_{\mathbf{c} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \right\}
\end{aligned}$$

where

$$P_{S^k}^{(i)}(\mathbf{c}^{(i)}) \triangleq \frac{P(\mathbf{c}^{(i)})}{Q_S^{(k)}(\mathbb{T}_i)}$$

is the normalized probability over $\mathbb{T}_i$ for each $i = 0, 1, ...$ By Jensen's inequality and the type covering lemma, the sum over each $\mathcal{C}_i$ can be bounded by

$$\sum_{\mathbf{c}^{(i)} \in \mathcal{C}_i} P_{S^k}^{(i)}(\mathbf{c}^{(i)})^{\frac{1}{1+\rho}} \le |\mathcal{C}_i|^{\frac{\rho}{1+\rho}} \le \exp \left\{ \frac{\rho}{1+\rho} [kR(Q_S^{(i)}, \Delta) + \zeta(\epsilon)] + o(k) \right\}$$

for k sufficiently large and ϵ sufficiently small (N sufficiently large). Recalling that

$$Q_S^{(k)}(\mathbb{T}_i) \le \exp\{-k[D(Q_S^{(i)} \parallel Q_S) + \zeta(\epsilon)]\},$$

we have

$$E_s^{(n)}(\rho, P) \le \frac{1+\rho}{k} \ln \left\{ 1 + \sum_{i=1}^{\infty} \exp \left[\frac{k}{1+\rho} \left(\rho R(Q_S^{(i)}, \Delta) - D(Q_S^{(i)} \parallel Q_S) + \zeta(\epsilon) \right) + o(k) \right] \right\} \tag{8.37}$$

for k sufficiently large and ϵ sufficiently small (N sufficiently large). Recall that $P_S^{(i)}$ denotes the Gaussian source with mean zero and variance $\sigma^2(i) = \Delta + (2i-1)\epsilon$. Consequently, using the fact [6] that if the exponential rate of each term, as a function of i, is of the form $U_i = \ln(Ai + B) - Ci$, where A, B, and C are positive reals, then the term with the largest exponent dominates the exponential behavior of the summation, i.e.,

$$\lim_{k \to \infty} \frac{1}{k} \ln \left\{ 1 + \sum_{i=1}^{\infty} \exp\left[k(\ln(Ai + B) - Ci) + o(k)\right] \right\} = \max_{i \ge 1} [\ln(Ai + B) - Ci], \tag{8.38}$$

we obtain

$$\lim_{n\to\infty} E_s^{(n)}(\rho, P) \leq \tau \max_{i\geq 1}[\rho R(Q_S^{(i)}, \Delta) - D(Q_S^{(i)} \parallel Q_S) + \zeta(\epsilon)]. \tag{8.39}$$

Note also that the sequence $\left\{\rho R(Q_S^{(i)}, \Delta) - D(Q_S^{(i)} \parallel Q_S)\right\}_{i=1}^{\infty}$ is non-increasing after some finite i, which means the maximum of (8.39) is achieved for some finite $\sigma^2(i)$. Letting ϵ go to zero, it follows by the continuity of $R(Q_S^{(i)}, \Delta)$ and $D(Q_S^{(i)} \parallel Q_S)$ as functions of $\sigma^2(i)$ that

$$\begin{aligned}
&\lim_{\epsilon\to 0} \sup_{\sigma^2(i)} [\rho R(Q_S^{(i)}, \Delta) - D(Q_S^{(i)} \parallel Q_S) + \zeta(\epsilon)] \\
&= \max[\rho R(\widetilde{Q}_S, \Delta) - D(\widetilde{Q}_S \parallel Q_S)]
\end{aligned} \tag{8.40}$$

where the maximum in (8.40) is taken over all the MGS $\widetilde{Q}_S$ with mean zero and variance $\sigma^2 > \Delta$. Therefore, we have

$$\lim_{n\to\infty} E_s^{(n)}(\rho, P) \leq \left\{0, \frac{\tau}{2}\left[\rho \ln \frac{\sigma_S^2}{\Delta} + (1+\rho)\ln(1+\rho) - \rho\right]\right\} = \tau E(Q_S, \Delta, \rho). \tag{8.41}$$

Finally, on account of (8.36) and (8.41), we may claim that, there exists a sequence of JSC codes $(f_n, \varphi_n, \Delta, \mathcal{E}, t)$, where $f_n = \widetilde{f}_n \circ f_{\Delta,k}$ and $\varphi_n = \widetilde{\varphi}_n$, such that for n sufficiently large,

$$P_{\Delta}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \leq \exp\left\{-n \max_{0\leq\rho\leq 1}\left[E_o(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)\right] + o(n)\right\},$$

by which we establish the lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$ given in (8.30). Furthermore, when $W_{Y|X}$ is an MGC, the bound (8.31) holds trivially since $\widetilde{E}_o(W_{Y|X}, \mathcal{E}, \rho)$ is a lower bound of $E_o(W_{Y|X}, \mathcal{E}, \rho)$. ■

8.2.5 Tightness of the Upper and Lower Bounds

Applying Fenchel duality theorem in Chapter 4 to our source and channel functions $E(Q_S, \Delta, \rho)$ and $\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho)$ with respect to their Fenchel transforms in Lemmas 4.4, 4.5 and 4.6, we obtain the following equivalent bounds.

Theorem 8.4 *Let* $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$. *Then*

$$\begin{aligned}
\overline{E}^{\Delta,\mathcal{E}}_{Jsp}(Q_S, W, t) &= \max_{0\leq\rho<\infty} [\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)] \\
&= \min_{R\geq 0} \left[\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{sp}(R, W_{Y|X}, \mathcal{E})\right], \qquad (8.42) \\
\underline{E}^{\Delta,\mathcal{E}}_{Jr}(Q_S, W, t) &= \max_{0\leq\rho\leq 1} [\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)] \\
&= \min_{R\geq 0} \left[\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right) + E_{\dagger}(R, W_{Y|X}, \mathcal{E})\right]. \qquad (8.43)
\end{aligned}$$

The proof of the theorem follows in a similar manner as Theorem 5.1. We next provide a necessary and sufficient condition under which $\overline{E}^{\Delta,\mathcal{E}}_{Jsp} = \underline{E}^{\Delta,\mathcal{E}}_{Jr}$ for the MGS-MGC pair.

Theorem 8.5 *Let* $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$. *The upper and lower bounds for* $E^{\Delta,\mathcal{E}}_J$ *given in Theorem 8.2 and (8.31) of Theorem 8.3 are equal if and only if*

$$2(2SDR)^\tau - \frac{2(2SDR)^\tau}{2(2SDR)^\tau - 1} \geq SNR. \qquad (8.44)$$

Remark 8.1 For $\tau R(Q_S, \Delta) \geq C(W_{Y|X}, \mathcal{E})$, $E^{\Delta,\mathcal{E}}_J(Q_S, W_{Y|X}, \tau) = 0$.

Proof: By comparing (8.42) and (8.43) we observe that the two bounds are identical if and only if the minimum of (8.42) (or (8.43)) is achieved at a rate no less than the channel critical rate, i.e.,

$$R_m \geq R_{cr}(W_{Y|X}) = \frac{1}{2}\ln\left[\frac{1}{2} + \frac{\text{SNR}}{4} + \frac{1}{2}\sqrt{1 + \frac{\text{SNR}^2}{4}}\right]$$

where R_m is the solution of (8.22). Let

$$f(R) \triangleq \frac{\beta^{\frac{1}{\tau}}}{\text{SDR}} - \frac{\text{SNR}}{2\beta}\left(1 + \sqrt{1 + \frac{4\beta}{\text{SNR}(\beta - 1)}}\right),$$

which is a strictly increasing function of R (refer to (2.53)), where $\beta = e^{2R}$. In order to ensure that the root of $f(R)$, R_m, is no less than $R_{cr}(W_{Y|X})$, we only need $f(R_{cr}(W_{Y|X})) \leq 0$. This reduces to the condition (8.44). ■

In Fig. 8.2, we partition the SDR-SNR plane into three parts for transmission rate $\tau = 0.5$, 1, 1.5 and 2: in Region **A** (including the boundary between **A** and **B**) $\tau R(Q_S, \Delta) \geq$

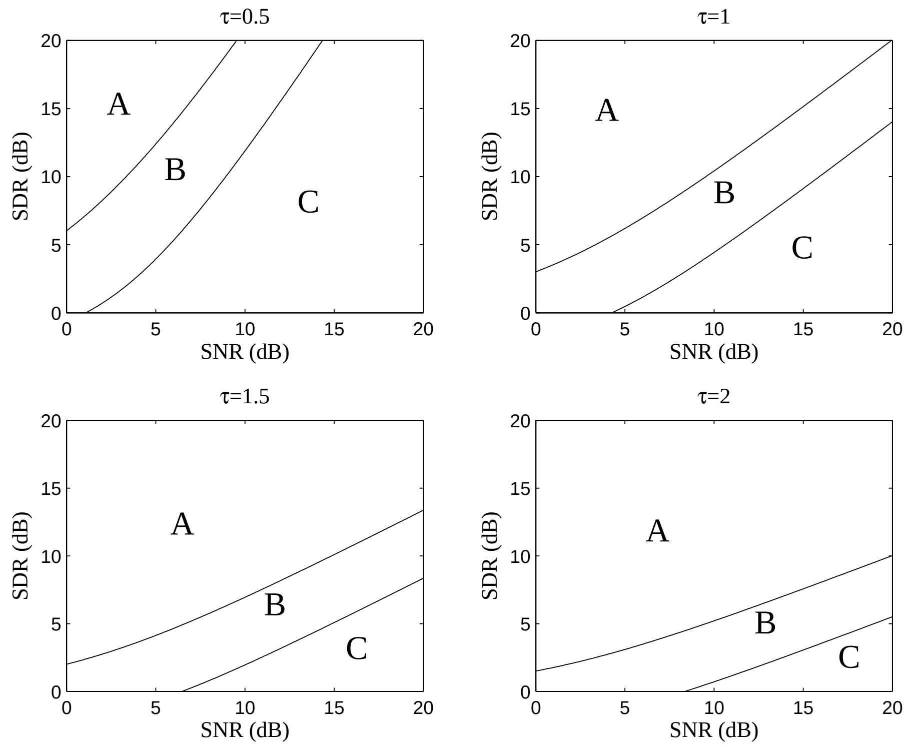

Figure 8.2: MGS-MGC source-channel pair: the regions for SNR and SDR pairs (both in dB) for different τ. In Region **A** (including the boundary between **A** and **B**) $E_J^{\Delta,\mathcal{E}} = 0$; in Region **B** (including the boundary between **B** and **C**), $E_J^{\Delta,\mathcal{E}}$ is determined exactly; and in region **C**, $E_J^{\Delta,\mathcal{E}} > 0$ is bounded by (8.11) and (8.31).

$C(W_{Y|X}, \mathcal{E})$ and $E_J^{\Delta,\mathcal{E}} = 0$; in Region **B** (including the boundary between **B** and **C**), $\overline{E}_{Jsp}^{\Delta,\mathcal{E}} = \underline{E}_{Jr}^{\Delta,\mathcal{E}}$ and hence $E_J^{\Delta,\mathcal{E}}$ is determined exactly; and in Region **C**, $E_J > 0$ is bounded by $\overline{E}_{Jsp}^{\Delta,\mathcal{E}}$ and $\underline{E}_{Jr}^{\Delta,\mathcal{E}}$. Fig. 8.3 shows the two bounds $\overline{E}_{Jsp}^{\Delta,\mathcal{E}}$ and $\underline{E}_{Jr}^{\Delta,\mathcal{E}}$ for different SDR-SNR pairs and transmission rate $\tau = 1$. We observe from the two figures that the two bounds coincide for a large class SDR-SNR pairs.

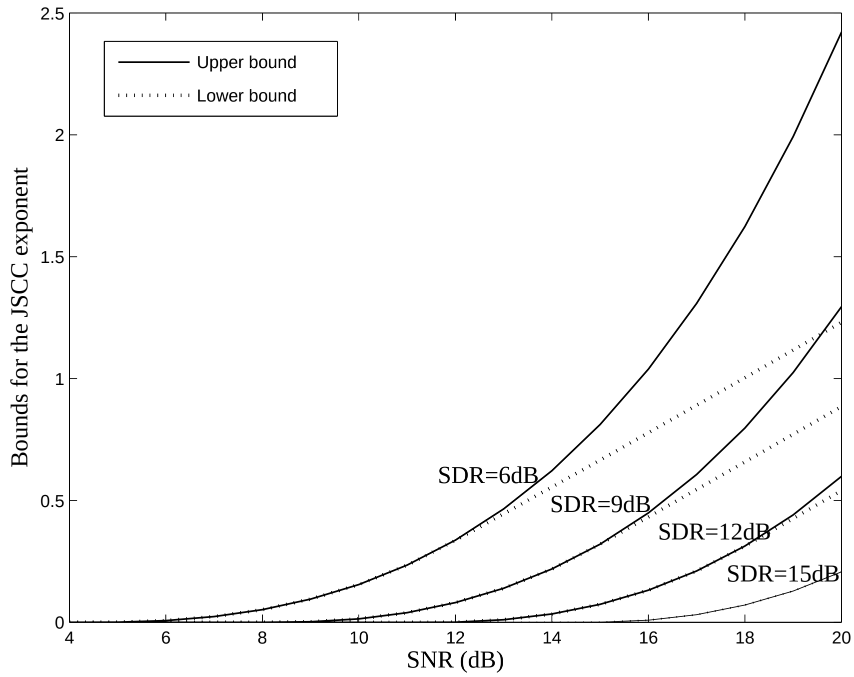

Figure 8.3: MGS-MGC source-channel pair: the upper and lower bounds for $E_J^{\Delta,\mathcal{E}}$ with $\tau = 1$.

8.3 Laplacian Sources with the Magnitude-Error Distortion over MGCs

In image coding applications, the Laplacian distribution is well known to provide a good model to approximate the statistics of transform coefficients such as discrete cosine and wavelet transform coefficients [78, 91]. Thus, it is of interest to study the theoretical performance for the lossy transmission of MLSs, say, over an MGC.

Consider an MLS Q_S with alphabet $\mathcal{S} = \mathbb{R}$, mean zero, variance $2\alpha^2$, and pdf

$$Q_S(s) = \frac{1}{2\alpha} \exp\left\{-\frac{|s|}{\alpha}\right\}, \quad s \in \mathcal{S},$$

denoted by $Q_S \backsim \mathcal{L}(0, \alpha)$. Note that for $Q_S \backsim \mathcal{L}(0, \alpha)$, $\mathbb{E}|s| = \alpha$. We assume that the distortion measure is the magnitude-error distortion given by $d(s, s') \triangleq |s - s'|$ for any

$s, s' \in \mathbb{R}$. The pdf for k-tuple source symbols is hence given by

$$Q_S^{(k)}(\mathbf{s}) = \left[\frac{1}{2\alpha}\right]^k \exp\left\{-\frac{\sum_{i=1}^k |s_i|}{\alpha}\right\}, \quad \mathbf{s} \in \mathcal{S}^k$$

and the distortion for any $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^k$ is hence given by

$$d^{(k)}(\mathbf{s}, \mathbf{s}') = \frac{1}{k}\sum_i |s_i - s_i'|.$$

For $Q_S \backsim \mathcal{L}(0, \alpha)$, the differential entropy[1] and the rate-distortion function (under the magnitude-error distortion measure) are respectively given by

$$H_{Q_S}(S) = 1 + \ln(2\alpha)$$

and

$$R(Q_S, \Delta) = \max\left\{0, \ln\frac{\alpha}{\Delta}\right\}.$$

Recall that the Kullback-Leibler divergence between two MLS $\widehat{Q}_S \backsim \mathcal{L}(0, \widehat{\alpha})$ and $Q_S \backsim \mathcal{L}(0, \alpha)$ is equal to

$$D(\widehat{Q}_S \parallel Q_S) = \frac{\widehat{\alpha}}{\alpha} - \ln\frac{\widehat{\alpha}}{\alpha} - 1.$$

The Laplacian sources enjoys the follows properties.

Lemma 8.1 *Let P_S be an arbitrary pdf on $\mathcal{S} = \mathbb{R}$ such that $\mathbb{E}_{P_S}|s| = \alpha < \infty$. Consider two MLS $Q_S \backsim \mathcal{L}(0, \alpha)$ and $\widetilde{Q}_S \backsim \mathcal{L}(0, \widetilde{\alpha})$. Then*

(a) $H_{P_S}(S) \le H_{Q_S}(S)$ with equality if and only if $P_S = Q_S$;

(b) $D(P_S \parallel \widetilde{Q}_S) \ge D(Q_S \parallel \widetilde{Q}_S)$ with equality if and only if $P_S = Q_S$;

(c) $R(P_S, \Delta) \le R(Q_S, \Delta)$ for any $\Delta > 0$ with equality if $P_S = Q_S$ ('only if' holds when $\Delta \le \alpha$).

[1]The differential entropy of a source with alphabet $\mathcal{S} \subseteq \mathbb{R}$ and a pdf Q_S is defined by (e.g., [29]) $H_{Q_S}(S) = -\int_{\mathcal{S}} Q_S(s) \log_2 Q_S(s) ds$.

Proof: (a) follows from

$$\begin{aligned} 0 \le D(P_S \parallel Q_S) &= -H_{P_S}(S) - \int P_S(s) \ln Q_S(s) ds \\ &= -H_{P_S}(S) + (1 + \ln(2\alpha)) \\ &= -H_{P_S}(S) + H_{Q_S}(S). \end{aligned} \tag{8.45}$$

From the above we note that $\mathbb{E}_{P_S}|s| < \infty$ also implies that $P_S \ll Q_S$. Similarly, we write

$$D(P_S \parallel \widetilde{Q}_S) = -H_{P_S}(S) + \int P_S(s) \ln \widetilde{Q}_S(s) ds$$

and

$$D(Q_S \parallel \widetilde{Q}_S) = -H_{Q_S}(S) + \int Q_S(s) \ln \widetilde{Q}_S(s) ds.$$

Noting that $\int P_S(s) \ln \widetilde{Q}_S(s) ds = \int Q_S(s) \ln \widetilde{Q}_S(s) ds$, (b) immediately follows from (a). Next we prove (c). Without loss of generality, we assume P_S has mean zero. Recalling that

$$R(P_S, \Delta) = \inf_{P_{S'|S}: \mathbb{E}|S'-S| \le \Delta} I(S; S'),$$

where S and S' are RV's in $\mathbb{R}$. So for arbitrary conditional density $P(S'|S)$ satisfying the constraint, we have $R(P_S, \Delta) \le I(S; S')$. We first assume $\alpha \ge \Delta$, and consider the test channel

$$S' = \left(1 - \frac{\Delta}{\alpha}\right)^2 S + \text{sgn}(S)|W|,$$

where $\text{sgn}(S)$ is equal to 1 if $S \ge 0$ and is equal to -1 if $S < 0$, W is a dummy Laplacian RV with mean 0 and $\mathbb{E}|W| = (1 - \Delta/\alpha)\Delta$, and S is independent of W. We thus have

$$\begin{aligned} \mathbb{E}|S'| &= \mathbb{E}\left|\left(1 - \frac{\Delta}{\alpha}\right)^2 S + \text{sgn}(S)|W|\right| \\ &= \left(1 - \frac{\Delta}{\alpha}\right)^2 \mathbb{E}|S| + \mathbb{E}|W| \\ &= \left(1 - \frac{\Delta}{\alpha}\right)^2 \alpha + \left(1 - \frac{\Delta}{\alpha}\right) \Delta \\ &= \alpha - \Delta, \end{aligned} \tag{8.46}$$

and

$$\begin{aligned}
\mathbb{E}|S'-S| &= \mathbb{E}\left|\left(1-\frac{\Delta}{\alpha}\right)^2 S+\mathrm{sgn}(S)|W|-S\right| \\
&= \left|\frac{\Delta}{\alpha}\left(2-\frac{\Delta}{\alpha}\right)\mathbb{E}|S|-\mathbb{E}|W|\right| \\
&= \left|\frac{\Delta}{\alpha}\left(2-\frac{\Delta}{\alpha}\right)\alpha-\left(1-\frac{\Delta}{\alpha}\right)\Delta\right| \\
&= \Delta, \qquad (8.47)
\end{aligned}$$

Now for the choice of S', we have

$$\begin{aligned}
R(P_S,\Delta) &\le I(S;S') \\
&= H(S')-H(S'|S) \\
&\overset{(1)}{=} H(S')-H(\mathrm{sgn}(S)|W||S) \\
&= H(S')-H(|W||S)-\ln|\mathrm{sgn}(S)| \\
&\overset{(2)}{=} H(S')-H(|W|) \\
&\overset{(3)}{=} H(S')-H(W) \\
&\overset{(4)}{\le} \ln[2e(\alpha-\Delta)]-\ln[2e(1-\Delta/\alpha)\Delta] \\
&= \ln\frac{\alpha}{\Delta},
\end{aligned}$$

where the mutual information and entropies are taken under the joint distribution $P_S P_{S'|S}$, (1) holds since the differential entropy is invariant under translation, (2) holds since S is independent of W, (3) holds since Laplace distribution is symmetric, and (4) follows from the (a), noting that the equality in (4) is achieved if and only if $P_S \backsim \mathcal{L}(0,\alpha)$.

For $\Delta > \alpha$, let S' satisfy $\Pr(S'=0)=1$ and $S' \perp S$. Then $\mathbb{E}|S-S'| \le \mathbb{E}|S|+\mathbb{E}|S'| = \alpha < \Delta$. For this choice of S', $R(P_S,\Delta) \le I(S;S') = 0$ implies that $R(P_S,\Delta)=0$. ∎

Due to the striking similarity between the Laplacian source and the Gaussian source, the results in this above (especially regarding the bounds for $E_J(Q_S,W,\Delta,\mathcal{E},t)$) can be easily extended to a system composed by an MLS under the magnitude-error distortion measure and an MGC.

We remark that to employ the two-stage coding scheme as in the proof of Theorem 8.3 to derive a lower bound for transmitting the MLS over the MGC with magnitude distortion measure, we need to employ the type covering lemma (Lemma 3.7) for Laplacian-type classes.

For an MLS $Q_S \sim \mathcal{L}(0, \alpha)$ and distortion threshold Δ, we define the MLS exponent function

$$F_L(R, Q_S, \Delta) = \begin{cases} \frac{e^R \Delta}{\alpha} - \ln \frac{e^R \Delta}{\alpha} - 1 & \text{if} \quad R > \ln \frac{\alpha}{\Delta}, \\ 0 & \text{otherwise.} \end{cases} \tag{8.48}$$

if $\alpha > \Delta$; otherwise (if $\alpha \leq \Delta$), let

$$F_L(R, Q_S, \Delta) = \begin{cases} \frac{e^R \Delta}{\alpha} - \ln \frac{e^R \Delta}{\alpha} - 1 & \text{if} \quad R > 0, \\ \frac{\Delta}{\alpha} - \ln \frac{\Delta}{\alpha} - 1 & \text{otherwise.} \end{cases} \tag{8.49}$$

Consequently, by using Lemma 3.7, similar versions of Theorems 8.2 and 8.3 can be deduced by replacing the MGS by an MLS and we obtain the following results.

Theorem 8.6 *For the MLS Q_S and the MGC $W_{Y|X}$ with transmission rate τ,*

$$E_J(Q_S, W, \Delta, \mathcal{E}, \tau) \leq \min_{R \geq 0} \left[\tau F_L \left(\frac{R}{\tau}, Q_S, \Delta \right) + E_{sp}(R, W_{Y|X}, \mathcal{E}) \right]$$

and

$$E_J(Q_S, W, \Delta, \mathcal{E}, \tau) \geq \min_{R \geq 0} \left[\tau F_L \left(\frac{R}{\tau}, Q_S, \Delta \right) + E_{\dagger}(R, W_{Y|X}, \mathcal{E}) \right],$$

where $E_{sp}(R, W_{Y|X}, \mathcal{E})$ and $E_{\dagger}(R, W_{Y|X}, \mathcal{E})$ are given by (2.51) and (2.56) respectively.

It can be easily shown (by setting the channel to be a noiseless channel) that $F_L(R, Q_S, \Delta)$ $(R > 0)$ determines the lossy source excess distortion exponent for the MLS, i.e,

$$e_{\Delta}(R, P_S) = F_L(R, Q_S, \Delta),$$

for any $R > 0$. Meanwhile, like the MGS exponent with squared-error distortion, we can show that the MLS excess distortion exponent with magnitude-error distortion can also be expressed in Marton's form, as shown in the following lemma.

Lemma 8.2 *For the MLS $Q_S \sim \mathcal{L}(0, \alpha)$ and distortion threshold Δ,*

$$F_L(R, Q_S, \Delta) = \inf_{P_S \in \mathcal{P}(\mathcal{S}): R(P_S, \Delta) > R} D(P_S \parallel Q_S), \quad R > 0, \tag{8.50}$$

where the infimum is taken over all distribution P_S defined on $\mathbb{R}$.

Proof: We only need to consider the pdf P_S defined on $\mathbb{R}$ with $P_S \ll Q_S$ and $R(P_S, \Delta) \geq R$. Suppose the expectation $\mathbb{E}|s|$ under P_S is equal to γ. According to Lemma 8.1, the Laplacian pdf $Q_S^* \backsim \mathcal{L}(0, \gamma)$ satisfies $D(Q_S^* \parallel Q_S) \leq D(P_S \parallel Q_S)$ and $R(Q_S^*, \Delta) \geq R(P_S, \Delta) \geq R$. Therefore,

$$\begin{aligned}\inf_{P_S: R(P_S, \Delta) \geq R} D(P_S \parallel Q_S) &= \inf_{Q_S^* \backsim \mathcal{L}(0,\gamma): R(Q_S^*, \Delta) \geq R} D(Q_S^* \parallel Q_S) \\ &= \begin{cases} \frac{\gamma^*}{\alpha} - \ln \frac{\gamma^*}{\alpha} - 1 & \text{for } R > R(Q_S, \Delta), \\ 0 & \text{otherwise,} \end{cases}\end{aligned}$$

where γ^* is determined by $R = \ln(\gamma^*/\Delta)$. This is exactly the exponent $F_L(R, Q_S, \Delta)$ given by (8.48). ∎

8.4 Memoryless Systems with a Metric Source Distortion

In this section we consider the transmission of a class of continuous memoryless sources with alphabet $\mathcal{S} = \mathbb{R}$ over continuous memoryless channels when the source distortion function is a metric such that for $s, s' \in \mathcal{S}$ (1) $d(s, s') \geq 0$ with equality if and only if $s = s'$; (2) $d(s, s') = d(s', s)$; (3) the triangle inequality holds, i.e., for any $s_1, s_2, s_3 \in \mathcal{S}$, $d(s_1, s_2) + d(s_2, s_3) \geq d(s_1, s_3)$. We still assume that for any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}^k$,

$$d^{(k)}(\mathbf{s}, \mathbf{s}') \triangleq \frac{1}{k} \sum_{i=1}^{k} d(s_i, s_i').$$

Theorem 8.7 *For the continuous memoryless source Q_S with a distortion being a metric and the continuous memoryless channel $W_{Y|X}$ with a cost constraint $\mathcal{E}$ at the channel input, if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E}\exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, the JSCC excess distortion exponent satisfies*

$$E_J^{\Delta, \mathcal{E}}(Q_S, W_{Y|X}, \tau) \geq \max_{0 \leq \rho < 1} [E_0(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)], \tag{8.51}$$

where $E_o(W_{Y|X}, \mathcal{E}, \rho)$ *is Gallager's constrained channel function given by (2.40) and* $E(Q_S, \Delta, \rho)$ *is the source function for* Q_S *given by (4.7). Furthermore, if* $W_{Y|X}$ *is an MGC, we have*

$$E_J^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \geq \max_{0\leq\rho<1} [\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho) - \tau E(Q_S, \Delta, \rho)], \tag{8.52}$$

where $\widetilde{E}_0(W_{Y|X}, \mathcal{E}, \rho)$ *is Gallager's Gaussian-input channel function given by (2.54).*

Observation 8.1 Although Theorem 8.7 does not apply to MGSs under the squared-error distortion (which is not a metric) and MLSs under the magnitude-error distortion (which does not satisfy the finiteness condition), it applies to MGSs under the magnitude-error distortion, and more generally, it applies to generalized MGSs with parameters (α, σ) under the distortion function $d(s, s') \triangleq |s - s'|^p$ for any $s, s' \in \mathbb{R}$, whenever $0 < p \leq 1$ and $p < \alpha$; see the following example.

Example 8.1 The Gaussian and Laplacian distributions belong to the class of generalized Gaussian distributions, which are widely used in image coding applications. It is well known that the distribution of image subband coefficients is well approximated by the generalized Gaussian distribution [26, 91]. A generalized MGS Q_S with parameters (α, σ) has alphabet $\mathcal{S} = \mathbb{R}$, mean zero, variance σ^2, and pdf

$$Q_S(s) = \frac{\alpha\eta(\alpha, \sigma)}{2\Gamma(1/\alpha)} \exp\left\{-(\eta(\alpha, \sigma)|s|)^\alpha\right\}, \quad s \in \mathcal{S},$$

where $\Gamma(\cdot)$ is the Gamma function and

$$\eta(\alpha, \sigma) \triangleq \frac{1}{\sigma}\left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right)^{\frac{1}{2}} \quad \alpha > 0.$$

Note that the pdf reduces to the Gaussian and Laplacian pdf's for $\alpha = 2$ and 1, respectively. When $0 < p \leq 1$, the distortion $d(s, s') \triangleq |s - s'|^p$ is a metric. If we choose $s_o = 0$, then $\mathbb{E}\exp[\tau d(s, s_o)]$ would have the form

$$\mathbb{E}\exp[td(s, s_o)] = \int_{-\infty}^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)}ds = 2\int_0^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)}ds$$

where $A > 0$, $B > 0$, and C are independent of s. Clearly, the above integral is finite for any $Ct \geq 0$. If $Ct < 0$, and $\alpha > p$ is provided, the integral can be bounded by

$$\int_0^{+\infty} Ae^{-B|s|^p(|s|^{\alpha-p}+Ct)}ds \leq \int_0^x Ae^{-BCt|s|^p}ds + \int_x^{+\infty} Ae^{-B|s|^\alpha}ds$$

which is also finite, where $x > 0$ satisfies $x^{\alpha-p} + Ct = 0$.

For general *continuous* memoryless sources, unfortunately, we do not have counterparts to the type class and the type covering results of Lemmas 3.6 and 3.7 (for MGSs and MLSs, respectively). Hence, to establish the lower bound for the JSCC excess distortion exponent, we need to modify the proof of Theorem 8.3. We will use a different approach based on the technique introduced in [55] and the type covering lemma (Lemma 3.5) for finite alphabet DMS's.

Proof of Theorem 8.7:

Since the lower bound (8.52) immediately follows from (8.51), we only show the existence of a sequence of JSC codes for the source-channel pair (Q_S, W) such that its probability of excess distortion is upper bounded by

$$\exp\left\{-n \max_{0\le\rho<1}[E_0(W_{Y|X}, \mathcal{E}, \rho) - tE(Q_S, \Delta, \rho)] + o(n)\right\}$$

for n sufficiently large. We shall employ a concatenated "scalar discretization - vector quantization - lossless JSCC" scheme as shown in Fig. 8.4. Throughout the proof, we let $k = \tau n$, where $t > 0$ is finite, and set $0 < \epsilon < \Delta$ and $0 < \delta < \Delta - \epsilon$.

First Stage Coding: ϵ-Neighborhood Scalar Quantization

As in [55], we approximate the continuous memoryless source Q_S by a DMS $\widetilde{P}_{\widetilde{S}}$ with countably infinite alphabet $\widetilde{S}$ via an ϵ-neighborhood scalar quantization scheme. In particular, for any given $0 < \epsilon < \Delta$, there exists a countable set $\widetilde{S} = \{s_i, i = 1, 2, ...\} \subseteq \mathbb{R}$ with corresponding mutually disjoint subsets $\mathcal{S}_i \subseteq \{s \in \mathbb{R} : d(s_i, s) \le \epsilon\}$, $i = 1, 2, ...$, such that $\bigcup_{i=1}^{\infty} \mathcal{S}_i = \mathbb{R}$. Specifically, the subsets $\{\mathcal{S}_i\}$ partition $\mathbb{R}$; for example, a specific partition could be $\mathcal{S}_1 = \{s \in \mathbb{R} : d(s_1, s) \le \epsilon\}$ and

$$\mathcal{S}_i = \{s \in \mathbb{R} : d(s_i, s) \le \epsilon \quad \text{and} \quad d(s_j, s) > \epsilon \quad \text{for any } j < i\}$$

for $i \ge 2$. Consequently, we can employ a scalar quantizer $f_\epsilon : \mathcal{S} \longrightarrow \widetilde{S}$ to discretize the original memoryless source Q_S, such that $f_\epsilon(s) = s_i$ if $s \in \mathcal{S}_i$. Therefore, the first stage

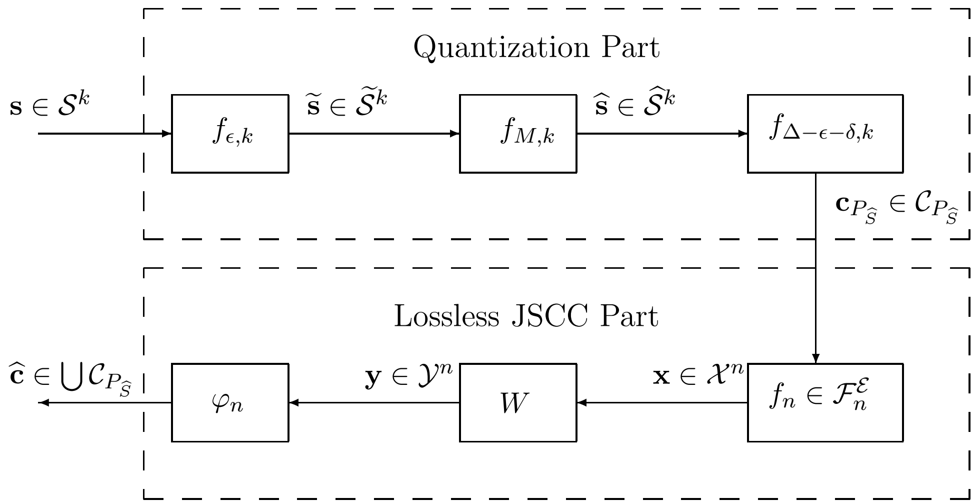

Figure 8.4: "Quantization plus lossless JSCC" scheme.

coding can be described as a mapping:

$$f_{\epsilon,k} : \mathcal{S}^k \longrightarrow \widetilde{\mathcal{S}}^k$$

where $f_{\epsilon,k}(\mathbf{s}) = (f_\epsilon(s_1), f_\epsilon(s_2), ..., f_\epsilon(s_k))$. We denote the source obtained at the output of $f_{\epsilon,k}$ by $\widetilde{P}_{\widetilde{S}}$ with alphabet $\widetilde{\mathcal{S}}$ and pmf

$$\widetilde{P}_{\widetilde{S}}(s_i) = \int_{s \in \mathcal{S}_i} Q_S(s) ds, \quad s_i \in \widetilde{\mathcal{S}}.$$

Lemma 8.3 For any $\epsilon > 0$ and $\rho > 0$, $E\left(\widetilde{P}_{\widetilde{S}}, \Delta + \epsilon, \rho\right) \leq E(Q_S, \Delta, \rho)$.

Proof: The proof is similar to the one of [55, Proposition 3], where the authors show that the rate-reliability function of the original source is bounded by that of its discretized version. Note that

$$E(P_S, \Delta, \rho) = \sup_{Q_S} [\rho R(Q_S, \Delta) - D(Q_S \parallel P_S)]$$

where the supremum is taken over all the distributions Q_S's defined on $\mathcal{S} = \mathbb{R}$ such that $R(Q_S, \Delta)$ and $D(Q_S \parallel P_S)$ are well-defined and finite, and similarly,

$$E\left(\widetilde{P}_{\widetilde{S}}, \Delta + \epsilon, \rho\right) = \sup_{\widetilde{Q}_{\widetilde{S}}} [\rho R(\widetilde{Q}_{\widetilde{S}}, \Delta + \epsilon) - D(\widetilde{Q}_{\widetilde{S}} \parallel \widetilde{P}_{\widetilde{S}})],$$

where the supremum is taken over all the pmf's $\widetilde{Q}_{\widetilde{S}}$'s defined on $\widetilde{\mathcal{S}}$ such that $R(\widetilde{Q}_{\widetilde{S}}, \Delta+\epsilon)$ and $D(\widetilde{Q}_{\widetilde{S}} \parallel \widetilde{P}_{\widetilde{S}})$ are finite. Now for any given $\widetilde{Q}_{\widetilde{S}}$ on $\widetilde{\mathcal{S}}$ which is absolutely continuous with respect to $\widetilde{P}_{\widetilde{S}}$, let $p_i = \widetilde{P}_{\widetilde{S}}(s = s_i)$ and $q_i = \widetilde{Q}_{\widetilde{S}}(s = s_i)$, $i = 1, 2, ...$ We then construct a pdf Q_S on $\mathbb{R}$ by

$$Q_S(s) \quad = \quad \begin{cases} P_S(s)\frac{q_i}{p_i}, & s \in \mathcal{S}_i, \quad p_i \neq 0 \quad i = 1, 2, ..., \\ 0, & s \in \mathcal{S}_i, \quad p_i = 0, \quad i = 1, 2, ... \end{cases}$$

It has been shown in the proof of [56, Proposition 3] that for such Q_S

$$D(Q_S \parallel P_S) = D(\widetilde{Q}_{\widetilde{S}} \parallel \widetilde{P}_{\widetilde{S}}) \quad \text{and} \quad R(\widetilde{Q}_{\widetilde{S}}, \Delta+\epsilon) \leq R(Q_S, \Delta).$$

Since the above holds for all $\widetilde{Q}_{\widetilde{S}}$, it then follows that

$$\sup_{Q_S}[\rho R(Q_S, \Delta) - D(Q_S \parallel P_S)] \geq \sup_{\widetilde{Q}_{\widetilde{S}}}[\rho R(\widetilde{Q}_{\widetilde{S}}, \Delta+\epsilon) - D(\widetilde{Q}_{\widetilde{S}} \parallel \widetilde{P}_{\widetilde{S}})].$$

■

Second Stage Coding: Truncating Source Alphabet

We next truncate the alphabet $\widetilde{S}$ to obtain a finite-alphabet source. Without loss of generality, assuming that $\widetilde{\mathcal{S}} = \{s_1, s_2, ...\}$ such that

$$\widetilde{P}_{\widetilde{S}}(s_1) \geq \widetilde{P}_{\widetilde{S}}(s_2) \geq \widetilde{P}_{\widetilde{S}}(s_3) \geq \cdots,$$

then for M sufficiently large, we take $\widehat{\mathcal{S}}$ be the set of the first M elements, i.e., $\widehat{\mathcal{S}} = \{s_1, s_2, ..., s_M\}$. For $s \in \widetilde{\mathcal{S}} = \{s_1, s_2, ...\}$ define function

$$f_M(s) = \begin{cases} s & \text{if } s \in \widehat{\mathcal{S}}, \\ s_1 & \text{otherwise.} \end{cases}$$

Then the second stage coding is a mapping:

$$f_{M,k} : \widetilde{\mathcal{S}}^k \longrightarrow \widehat{\mathcal{S}}^k$$

where $f_{M,k}(\mathbf{s}) = (f_M(s_1), f_M(s_2), ..., f_M(s_k))$. We denote the finite-alphabet DMS at the output of $f_{M,k}$ by $\widehat{P}_{\widehat{S}}$ with alphabet $\widehat{\mathcal{S}}$ and pmf

$$\widehat{P}_{\widehat{S}}(s) = \sum_{s_i \in \widetilde{\mathcal{S}}: f_M(s_i) = s} \widetilde{P}_{\widetilde{S}}(s_i) \qquad s \in \widehat{\mathcal{S}}.$$

If we let M go to infinity, $\widehat{P}_{\widehat{S}} \to \widetilde{P}_{\widetilde{S}}$, i.e., the statistics of $\widehat{P}_{\widehat{S}}$ approaches that of $\widetilde{P}_{\widetilde{S}}$. Furthermore, we have the following results.

Lemma 8.4 *For any $\delta > 0$ and $\rho > 0$, $E\left(\widehat{P}_{\widehat{S}}, \Delta + \delta, \rho\right) \leq E(\widetilde{P}_{\widetilde{S}}, \Delta, \rho)$ for M large enough.*

The proof of this lemma is similar to that of Lemma 8.3 and is omitted; readers may also refer to [55].

Lemma 8.5 [55, Lemma 1] *For any δ such that*

$$\mathbb{E}d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\},$$

if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E}\exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, then

$$\lim_{k\to\infty} -\frac{1}{k} \ln Pr\left\{d^{(k)}\left[f_{\epsilon,k}(\boldsymbol{s}), f_{M,k}(f_{\epsilon,k}(\boldsymbol{s}))\right] > \delta\right\} = r(M)$$

such that $r(M) \to \infty$ as $M \to \infty$, where the expectations are taken under P_S, and the probability is taken under $Q_S^{(k)}$.

Remark 8.2 Note also that $\mathbb{E}d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] \to 0$ as $M \to \infty$. Equivalently, Lemma 8.5 states that for any $0 < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\}$ and $r > 0$,

$$\lim_{k\to\infty} -\frac{1}{k} \ln \Pr\left\{d^{(k)}\left[f_{\epsilon,k}(\mathbf{s}), f_{M,k}(f_{\epsilon,k}(\mathbf{s}))\right] > \delta\right\} \geq r$$

for M sufficiently large.

Third Stage Coding: $(\Delta - \epsilon - \delta)$-Admissible Quantization

Consider transmitting the DMS $\widehat{P}_{\widehat{S}}$ over the continuous memoryless channel $W_{Y|X}$. Since $\widehat{P}_{\widehat{S}}$ has a finite alphabet $\{s_1, s_2, ..., s_M\}$, we now can employ a similar method as used in the proof of Theorem 8.3. Now we partition the k-dimensional source space $\widehat{\mathcal{S}}^k$ by a sequence of type classes $\left\{\mathbb{T}_{P_{\widehat{S}}} : P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})\right\}$.

Let δ be a number satisfying $0 < \delta < \sup\{d\left[f_\epsilon(s), f_M(f_\epsilon(s))\right] : s \in \mathbb{R}\}$. Setting "$\Delta$" in the type covering lemma (Lemma 3.5) to be $\Delta - \epsilon - \delta$, we can employ a $(\Delta - \epsilon - \delta)$-admissible quantizer via the sets $\mathcal{C}_{P_{\widehat{S}}}$ as follows:

$$f_{\Delta-\epsilon-\delta,k} : \widehat{\mathcal{S}}^k \longrightarrow \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$$

such that for every $\mathbf{s} \in \widehat{\mathcal{S}}^k$, the output of $f_{\Delta-\epsilon-\delta,k}$ with respect to $\mathbf{s}$ has a distortion less that $\Delta-\epsilon-\delta$ and each $|\mathcal{C}_{P_{\widehat{S}}}|$ is bounded by $\exp\{k[R(P_{\widehat{S}}, \Delta-\epsilon-\delta)+\mu]\}$ for sufficiently large k. We denote the finite DMS at the output of $f_{\Delta-\epsilon-\delta,k}$ by P with alphabet $\bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$ and pmf

$$P(\mathbf{c}_{P_{\widehat{S}}}) = \sum_{\mathbf{s} \in \mathbb{T}_{P_{\widehat{S}}} : f_{\Delta-\epsilon-\delta,k}(s) = \mathbf{c}_{P_{\widehat{S}}}} \widehat{P}_{\widehat{S}^k}(\mathbf{s}), \qquad \mathbf{c}_{P_{\widehat{S}}} \in \mathcal{C}_{P_{\widehat{S}}}, \quad P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}}).$$

Fourth Stage Coding and Decoding: Lossless JSCC with Cost Constraint $\mathcal{E}$

For the DMS P and the continuous memoryless channel $W_{Y|X}$, a pair of (asymptotically) lossless JSC code

$$\widetilde{f}_n : \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}} \longrightarrow \mathcal{X}^n \qquad \text{and} \qquad \widetilde{\varphi}_n : \mathcal{Y}^n \longrightarrow \bigcup_{P_{\widehat{S}} \in \mathcal{P}_k(\widehat{\mathcal{S}})} \mathcal{C}_{P_{\widehat{S}}}$$

is applied, where the encoder is subject to a cost constraint $\mathcal{E}$, i.e., $f_n \in \mathcal{F}_n^{\mathcal{E}}$. Note that the decoder φ_n creates an approximation $\widehat{\mathbf{c}} = \varphi_n(\mathbf{y})$ of $\mathbf{c}_{P_{\widehat{S}}}$ based on the sequence $\mathbf{y}$. According to a modified version of Gallager's JSCC random-coding bound (which is derived in Appendix 8.2.3), there exists a sequence of lossless JSC codes $(\widetilde{f}_n, \widetilde{\varphi}_n, \mathcal{E})$ with bounded probability of error

$$\begin{aligned} P_e^{(n)}(P, W_{Y|X}, \mathcal{E}) &\triangleq \Pr(\widehat{\mathbf{c}} \neq \mathbf{c}_{P_{\widehat{S}}}) \\ &\leq \exp\left\{-n \max_{0\leq\rho\leq 1} \left[E_o(W_{Y|X}, \mathcal{E}, \rho) - E_s^{(n)}(\rho, P)\right] + o(n)\right\}. \end{aligned}$$

Analysis of the Probability of Excess Distortion

For the sake of simplicity, let (see Fig. 8.4)

$$\begin{aligned} \widetilde{\mathbf{s}} &= f_{\epsilon,k}(\mathbf{s}) \\ \widehat{\mathbf{s}} &= f_{M,k}(\widetilde{\mathbf{s}}) \in \mathbb{T}_{P_{\widehat{S}}} \end{aligned}$$

$$\begin{aligned} \mathbf{c}_{P_{\widehat{S}}} &= f_{\Delta-\epsilon-\delta,k}(\widehat{\mathbf{s}}) \\ \mathbf{x} &= f_n(\mathbf{c}_{P_{\widehat{S}}}) \\ \widehat{\mathbf{c}} &= \varphi_n(\mathbf{y}). \end{aligned}$$

Since

$$d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) \le d^{(k)}(\mathbf{s},\widetilde{\mathbf{s}}) + d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) \le \epsilon + d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}),$$

we have

$$\begin{aligned} & \Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) \\ \le\ & \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta-\epsilon) \\ \le\ & \Pr\left(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) + d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta-\epsilon, d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) < \delta\right) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \\ \le\ & \Pr\left(d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta-\epsilon-\delta\right) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta), \end{aligned}$$

where the probabilities are taken under the joint distribution $Q_S^{(k)}(\cdot)W_{Y|X}^{(n)}(\cdot|\cdot)$. According to the $(\Delta-\epsilon-\delta)$-admissible quantization rule, $d^{(k)}(\widehat{\mathbf{s}},\widehat{\mathbf{c}}) > \Delta-\epsilon-\delta$ implies that $\mathbf{c}_{P_{\widehat{S}}} \ne \widehat{\mathbf{c}}$, therefore, we can further bound

$$\begin{aligned} & \Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) \\ & < \Pr(\mathbf{c}_{P_{\widehat{S}}} \ne \widehat{\mathbf{c}}) + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \\ & \le \exp\left\{-n\left[\max_{0\le\rho\le1}\left[E_o(W_{Y|X},\mathcal{E},\rho) - E_s^{(n)}(\rho,P)\right] + o(n)\right]\right\} + \Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \end{aligned}$$

for k sufficiently large. It follows from Lemma 8.5 (also see the remark after it) that

$$\lim_{k\to\infty} -\frac{1}{k}\ln\Pr(d^{(k)}(\widetilde{\mathbf{s}},\widehat{\mathbf{s}}) \ge \delta) \to \infty$$

as $M\to\infty$. When we take the sum of two exponential functions that both converge to 0, the one with a smaller convergence rate would dominate the exponential behavior of the sum. Therefore, for sufficiently large M which only depends on δ, noting that $k=\tau n$, we have

$$\liminf_{n\to\infty} -\frac{1}{n}\ln\Pr(d^{(k)}(\mathbf{s},\widehat{\mathbf{c}}) > \Delta) \ge \liminf_{n\to\infty} \max_{0\le\rho\le1}\left[E_o(W_{Y|X},\mathcal{E},\rho) - E_s^{(n)}(\rho,P)\right]. \quad (8.53)$$

Consequently, it can be shown by using the method of types (in a similar manner as the proof of Theorem 8.3) that for M sufficiently large

$$\lim_{n\to\infty} E_s^{(n)}(\rho, P) \le \tau E(\widehat{P}_{\widehat{S}}, \Delta - \epsilon - \delta, \rho).$$

Using Lemmas 8.4 and 8.3 successively, we can approximate $E(\widehat{P}_{\widehat{S}}, \Delta - \epsilon - \delta, \rho)$ by

$$\begin{aligned} \lim_{n\to\infty} E_s^{(n)}(\rho, P) &\le \tau E(\widetilde{P}_{\widetilde{S}}, \Delta - \epsilon - 2\delta, \rho) \\ &\le \tau E(Q_S, \Delta - 2\epsilon - 2\delta, \rho). \end{aligned} \tag{8.54}$$

Finally, substituting (8.54) back into (8.53), and letting $\epsilon \to 0$ and $\delta \to 0$, we complete the proof of Theorem 8.7. ■

8.5 Conclusion

In this chapter, we investigated the JSCC excess distortion exponent $E_J^{\Delta,\mathcal{E}}$ for memoryless communication systems with continuous alphabets. For the Gaussian system with the squared-error source distortion measure and a power channel input constraint, we derived upper and lower bounds for the excess distortion exponent. The bounds extend the earlier results for discrete systems (Chapters 5 and 7) in such a way that the lower/upper bound can be expressed in terms of the sum of source and channel exponents. They can also be expressed in equivalent parametric forms as differences of source and channel functions. We then extended these bounds to Laplacian-Gaussian source-channel pairs with the magnitude-error distortion. By employing a different technique, we established a lower bound (of similar parametric form) for $E_J^{\Delta,\mathcal{E}}$ for a class of memoryless source-channel pairs under a metric distortion measure and some finiteness condition. For the Gaussian system, a sufficient and necessary condition for which the two bounds of $E_J^{\Delta,\mathcal{E}}$ coincide was provided. It was observed that the two bounds coincide for lots of source-channel parameters, thus exactly determining $E_J^{\Delta,\mathcal{E}}$.

Chapter 9

Multi-Terminal Systems: Asymmetric 2-User Discrete Memoryless Systems

In the previous chapters, we investigated the JSCC reliability function for discrete and continuous single-user systems. It is of natural interest to study the JSCC error exponent for multi-terminal source-channel systems.

In this chapter, we address the asymmetric 2-user source-channel coding system depicted in Fig. 9.1. Two discrete memoryless correlated source messages $(\mathbf{s}, \mathbf{l}) \in \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n}$ drawn from a joint distribution $Q_{SL} \in \mathcal{S} \times \mathcal{L}$, consisting of a common source message $\mathbf{s}$ and a private source message $\mathbf{l}$ of length τn, are transmitted over a discrete memoryless asymmetric communication channel described by $W_{YZ|UX} \in \mathcal{P}(\mathcal{Y} \times \mathcal{Z} | \mathcal{U} \times \mathcal{X})$ with block codes of length n, where $\tau > 0$ (measured in source symbol/channel use) is the overall transmission rate. The common source can be accessed by both encoders, but the private source can only be observed by one encoder (say, Encoder 1). In this set-up, the goal is to send the common information to both receivers, and send the private information to only one receiver (say, Decoder 1).

It is worthwhile to point out that the asymmetric 2-user system can be specialized to

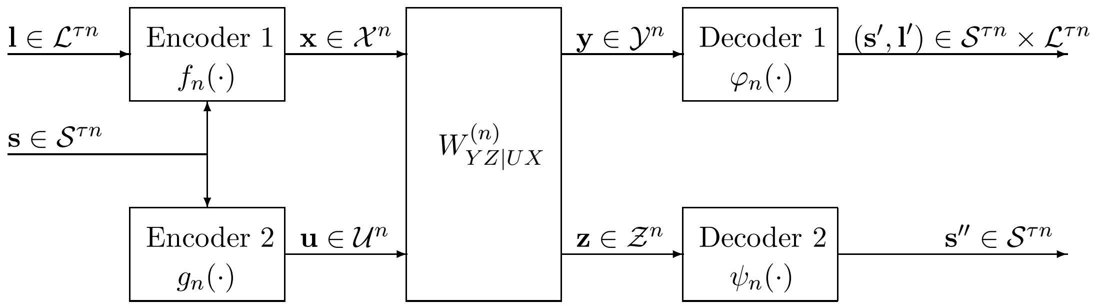

Figure 9.1: Transmitting two CS over the asymmetric 2-user communication channel.

the following two classical asymmetric multi-terminal scenarios.

(a) The CS-AMAC system: If we remove Decoder 2 from Fig. 9.1, and let $|\mathcal{Z}| = 1$, then the channel reduces to a multiple-access channel $W_{Y|UX}$, and the coding problem reduces to transmitting two correlated sources (CS) over an asymmetric multiple-access channel (AMAC) with one receiver.

(b) The CS-ABC system: If we remove Encoder 2 from Fig. 9.1, and let $|\mathcal{U}| = 1$, then the channel reduces to a broadcast channel $W_{YZ|X}$, and the coding problem reduces to transmitting two CS over an asymmetric broadcast channel (ABC) with one transmitter.

The sufficient and necessary condition for the reliable transmission of CS over the AMAC – i.e., the JSCC theorem for the CS-AMAC system – has been derived with single letter characterization in [20]. For the CS-ABC system, neither the sufficient nor the necessary condition is known to the best of our knowledge. In this chapter, we study a refined version of the JSCC theorem for the general asymmetric 2-user system (depicted in Fig. 9.1), by investigating the achievable JSCC error exponent pair (for two receivers) as well as the system JSCC error exponent, i.e., the largest convergence rate of asymptotic exponential decay of the system (overall) probability of erroneous transmission. We also apply our results to the CS-AMAS and CS-ABC systems.

At this point we pause to mention some related works in the literature on the multi-terminal JSCC of CS. The JSCC theorem for transmitting two CS over a (symmetric) multiple access channel (each encoder can only access one source) has been studied in [1, 28, 35, 56, 58, 76], and the JSCC theorem for transmitting two CS over a (symmetric) broadcast channel (each decoder needs to reconstruct one source) has been addressed in [27, 48]. These works focus on the case when the overall transmission rate τ is 1 and establish some sufficient and/or necessary conditions for which the sources can be reliably transmitted over the channel. However, for both (symmetric) systems, no matter whether the transmission rate τ is 1 or not, the tight sufficient and necessary condition (JSCC theorem) with single-letter characterization is still unknown.

In Section 9.1 we formally describe the 2-user source-channel system and define achievable error exponents and the system JSCC error exponent. The idea of superposition encoding for the 2-user asymmetric system is next introduced in Section 9.2.

By employing the joint type packing lemma and generalized maximum mutual information decoders, we establish in Section 9.3 achievable exponential upper bounds for the probabilities of erroneous transmission over an augmented 2-user channel $W_{YZ|TUX}$ for a given triple of n-length sequences $(\mathbf{t}, \mathbf{u}, \mathbf{x})$; see Proposition 9.1. Here, the augmented channel $W_{YZ|TUX}$ is induced from the original 2-user channel $W_{YZ|UX}$ by adding an auxiliary random variable (RV) T such that T, (UX), and (YZ), form a Markov chain in this order. We introduce the RV T because we will employ superposition encoding which maps a source message pair $(\mathbf{s}, \mathbf{l})$ to a codeword triplet $(\mathbf{t}, \mathbf{u}, \mathbf{x})$, where $\mathbf{t}$ is the auxiliary superposition codeword. For the asymmetric 2-user system, since one of the encoders has full access to both sources, it knows the output of the other encoder. By properly designing the two (superposition) encoders, we apply Proposition 9.1 to establish a universally achievable error exponent pair for the two receivers (namely, the pair of exponents can be achieved by a sequence of source-channel codes independently of the statistics of the source and the channel); this generalizes Körner and Sgarro's exponent pair for ABC coding (with uniformly distributed message sets) [59]. We also employ Proposition 9.1 to establish a lower bound for the system JSCC error exponent.

Note that one consequence of our results is a sufficient condition (forward part) for the JSCC theorem. In Section 9.4, we use Fano's inequality to prove a necessary condition (converse part) which coincides with the sufficient condition, and hence completes the JSCC theorem. In Section 9.5 we demonstrate that the separation principle holds for the 2-user system, i.e., there exists a separate source and channel coding system which can achieve optimality from the point of view of reliable transmissibility.

Using an approach analogous to Theorem 6.1 in Chapter 6 and Theorem 7.5 in Chapter 7, we also obtain an upper bound for the system JSCC error exponent in Section 9.6. As applications, we then specialize these results to the CS-AMAC and CS-ABC systems in Section 9.7. The computation of the lower and upper bounds for the system JSCC error exponent is partially studied for the CS-AMAC system when the channel admits a symmetric conditional distribution. Finally, we state our conclusions in Section 9.9.

9.1 System Description

Let $W_{YZ|UX} \in \mathcal{P}(\mathcal{Y} \times \mathcal{Z}|\mathcal{U} \times \mathcal{X})$ be a 2-user discrete memoryless channel with finite input alphabet $\mathcal{U}\times\mathcal{X}$, finite output alphabet $\mathcal{Y}\times\mathcal{Z}$, and a transition distribution $W_{YZ|UX}(y, z|u, x)$ such that the n-tuple transition probability is

$$W^{(n)}_{YZ|UX}(\mathbf{y}, \mathbf{z}|\mathbf{u}, \mathbf{x}) = \prod_{i=1}^{n} W_{YZ|UX}(y_i, z_i|u_i, x_i),$$

where $u \in \mathcal{U}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, $\mathbf{u} \triangleq (u_1, ..., u_n) \in \mathcal{U}^n$, $\mathbf{x} \triangleq (x_1, ..., , x_n) \in \mathcal{X}^n$, $\mathbf{y} \triangleq (y_1, ..., y_n) \in \mathcal{Y}^n$, and $\mathbf{z} \triangleq (z_1, ..., z_n) \in \mathcal{Z}^n$. Denote the marginal transition distributions of $W_{YZ|UX}$ at its Y-output (respectively Z-output) by $W_{Y|UX} \triangleq \sum_Z W_{YZ|UX}$ (respectively $W_{Z|UX} \triangleq \sum_Y W_{YZ|UX}$). The marginal distributions of $W^{(n)}_{YZ|UX}$ are denoted by $W^{(n)}_{Y|UX}$ and $W^{(n)}_{Z|UX}$, respectively.

Consider two discrete memoryless CS with a generic joint distribution Q_{SL} defined on the finite alphabet $\mathcal{S} \times \mathcal{L}$ such that the k-tuple joint distribution is

$$Q^{(k)}_{SL}(\mathbf{s}, \mathbf{l}) = \prod_{i=1}^{k} Q_{SL}(s_i, l_i),$$

where $(s, l) \in \mathcal{S} \times \mathcal{L}$, and $(\mathbf{s}, \mathbf{l}) \triangleq ((s_1, l_1), ..., (s_k, l_k)) \in \mathcal{S}^k \times \mathcal{L}^k$. For each pair of source messages $(\mathbf{s}, \mathbf{l})$ drawn from the above joint distribution, we need to transmit the *common* *message* **s** over the channel $W_{YZ|UX}$ to Receivers Y and Z and transmit the *private message* **l** only to Receiver Y.

A JSC code with block length n and positive transmission rate τ (source sumbol/channel use) for transmitting Q_{SL} through $W_{YZ|UX}$ is a quadruple of mappings, $(f_n, g_n, \varphi_n, \psi_n)$, where

$$f_n : \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n} \longrightarrow \mathcal{X}^n$$

and

$$g_n : \mathcal{S}^{\tau n} \longrightarrow \mathcal{U}^n$$

are called encoders, and

$$\varphi_n : \mathcal{Y}^n \longrightarrow \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n}$$

and

$$\psi_n : \mathcal{Z}^n \rightarrow \mathcal{S}^{\tau n}$$

are referred to as Y-decoder and Z-decoder, respectively; see Fig. 9.1.

The probabilities of Y- and Z-error are given by

$$\begin{aligned} P_{Ye}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) &\triangleq \Pr(\{\varphi_n(Y^n) \neq (S^{\tau n}, L^{\tau n})\}) \\ &= \sum_{\mathbf{s},\mathbf{l}} Q_{SL}^{(\tau n)}(\mathbf{s}, \mathbf{l}) \sum_{\mathbf{y}:\varphi_n(\mathbf{y}) \neq (\mathbf{s},\mathbf{l})} W_{Y|UX}^{(n)}(\mathbf{y}|\mathbf{u}, \mathbf{x}) \end{aligned} \tag{9.1}$$

and

$$P_{Ze}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \triangleq \Pr(\{\psi_n(Z^n) \neq S^{\tau n}\}) = \sum_{\mathbf{s},\mathbf{l}} Q_{SL}^{(\tau n)}(\mathbf{s}, \mathbf{l}) \sum_{\mathbf{z}:\psi_n(\mathbf{z}) \neq \mathbf{s}} W_{Z|UX}^{(n)}(\mathbf{z}|\mathbf{u}, \mathbf{x}) \tag{9.2}$$

where $\mathbf{x} \triangleq f_n(\mathbf{s}, \mathbf{l})$ and $\mathbf{u} \triangleq g_n(\mathbf{s})$ are the corresponding codewords of the source message pair $(\mathbf{s}, \mathbf{l})$ and the source message $\mathbf{s}$, and $\mathbf{y}$ and $\mathbf{z}$ are the received codewords at the Receivers Y and Z, respectively. Since we will study the exponential behavior of these probabilities using the method of types, it might be a better way to rewrite the probabilities of Y- and Z- error as a sum of probabilities of types

$$P_{ie}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) = \sum_{P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L})} Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) P_{ie}(\mathbb{T}_{SL}), \qquad i = Y, Z, \tag{9.3}$$

where $\mathbb{T}_{SL} \triangleq \mathbb{T}_{P_{SL}}$, and

$$P_{Ye}(\mathbb{T}_{SL}) = \frac{1}{|\mathbb{T}_{SL}|} \sum_{(\mathbf{s},\mathbf{l})\in\mathbb{T}_{SL}} \sum_{\mathbf{y}:\varphi_n(\mathbf{y})\neq(\mathbf{s},\mathbf{l})} W^{(n)}_{Y|UX}(\mathbf{y}|\mathbf{u},\mathbf{x}) \tag{9.4}$$

and

$$P_{Ze}(\mathbb{T}_{SL}) = \frac{1}{|\mathbb{T}_{SL}|} \sum_{(\mathbf{s},\mathbf{l})\in\mathbb{T}_{SL}} \sum_{\mathbf{z}:\psi_n(\mathbf{z})\neq\mathbf{s}} W^{(n)}_{Z|UX}(\mathbf{z}|\mathbf{u},\mathbf{x}). \tag{9.5}$$

We say that the JSCC error exponent pair (E_{AY}, E_{AZ}) is achievable with respect to $\tau > 0$ if there exists a sequence of JSC codes $(f_n, g_n, \varphi_n, \psi_n)$ with transmission rate τ such that the probabilities of Y-error and Z-error are simultaneously bounded by

$$P^{(n)}_{ie}(Q_{SL}, W_{YZ|UX}, \tau) \leq 2^{-n[E_{Ai}-\delta]}, \quad i = Y, Z \tag{9.6}$$

for n sufficiently large and any $\delta > 0$. As the point-to-point system, we denote the system (overall) probability of error by

$$P^{(n)}_{e}(Q_{SL}, W_{YZ|UX}, \tau) \triangleq \Pr\left(\{\varphi_n(Y^n) \neq (S^{\tau n}, L^{\tau n})\} \cup \{\psi_n(Z^n) \neq S^{\tau n}\}\right), \tag{9.7}$$

where $(S^{\tau n}, L^{\tau n})$ are drawn according to $Q^{(\tau n)}_{SL}$.

Definition 9.1 Given CS Q_{SL}, 2-user discrete memoryless channel $W_{YZ|UX}$ and transmission rate $\tau > 0$, the system JSCC error exponent $E_J(Q_{SL}, W_{YZ|UX}, \tau)$ is defined as supremum of the set of all numbers E for which there exists a sequence of JSC codes $(f_n, g_n, \varphi_n, \psi_n)$ with blocklength n and transmission rate τ such that

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P^{(n)}_{e}(Q_{SL}, W_{YZ|UX}, \tau). \tag{9.8}$$

Since the system probability of error must be larger than $P^{(n)}_{Ye}$ and $P^{(n)}_{Ze}$ defined by (9.1) and (9.2), and is also upper bounded by the sum of the two, it follows that for any sequence of JSC codes $(f_n, g_n, \varphi_n, \psi_n)$

$$\liminf_{n\to\infty} -\frac{1}{n} \log_2 P^{(n)}_{e}(Q_{SL}, W_{YZ|UX}, \tau) = \liminf_{n\to\infty} -\frac{1}{n} \log_2 \max\left(P^{(n)}_{Ye}, P^{(n)}_{Ze}\right). \tag{9.9}$$

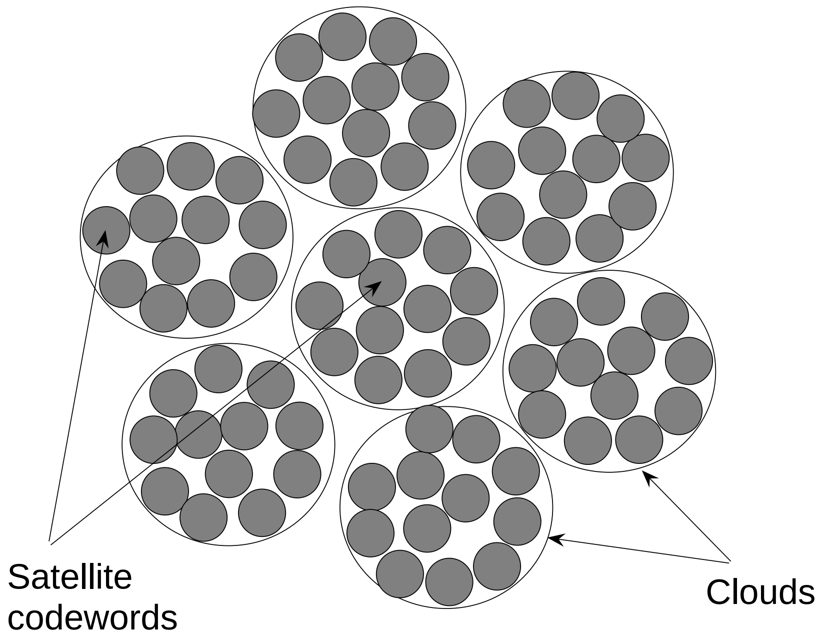

Figure 9.2: Relation between clouds and satellite codewords in superposition coding.

9.2 Superposition Encoding for Asymmetric 2-User Channels

Given an asymmetric 2-user channel $W_{YZ|UX}$, at the encoder side, we can artificially augment the channel input alphabet by introducing an auxiliary (arbitrary and finite) alphabet $\mathcal{T}$, and then look at the channel as a discrete memoryless channel $W_{YZ|TUX} = W_{YZ|UX}$ with marginal distributions $W_{Y|TUX}$ and $W_{Z|TUX}$ such that

$$W_{YZ|TUX}(y,z|t,u,x) = W_{YZ|UX}(y,z|u,x)$$

for any $t \in \mathcal{T}$, $u \in \mathcal{U}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$. In other words, we introduce a dummy RV $T \in \mathcal{T}$ such that T, (U,X), and (Y,Z) form a Markov chain in this order, i.e., $T \to (U,X) \to (Y,Z)$.

The idea of superposition coding is described as follows. The encoder g_n first maps the source message $\mathbf{s}$ to a pair of n-length sequences $(\mathbf{t},\mathbf{u}) \in \mathcal{T}^n \times \mathcal{U}^n$ with a fixed type, say P_{TU}, and then sends the codeword $\mathbf{u}$ over the channel. The encoder f_n first maps each pair

$(\mathbf{s}, \mathbf{l})$ to a triple of sequences $(\mathbf{t}, \mathbf{u}, \mathbf{x}) \in \mathcal{T}^n \times \mathcal{U}^n \times \mathcal{X}^n$ such that $\mathbf{x} \in \mathbb{T}_{P_{X|TU}}(\mathbf{t}, \mathbf{u})$, then f_n sends the codeword $\mathbf{x}$ over the channel. In other words, f_n maps $(\mathbf{s}, \mathbf{l})$ to the same $(\mathbf{t}, \mathbf{u})$ as g_n such that $(\mathbf{t}, \mathbf{u}, \mathbf{x})$ has a joint type $P_{TU}P_{X|TU}$.

Since $W^{(n)}_{YZ|TUX}(\mathbf{y}, \mathbf{z}|\mathbf{t}, \mathbf{u}, \mathbf{x})$ is equal to $W^{(n)}_{YZ|UX}(\mathbf{y}, \mathbf{z}|\mathbf{u}, \mathbf{x})$ and is independent of $\mathbf{t}$, transmitting the codewords $(\mathbf{u}, \mathbf{x})$ through the channel $W_{YZ|UX}$ can be viewed as transmitting the codewords $(\mathbf{t}, \mathbf{u}, \mathbf{x})$ over the augmented channel $W_{YZ|TUX}$. Here, the common outputs of g_n and f_n, $(\mathbf{t}, \mathbf{u})$'s, are called auxiliary *cloud centers* according to the traditional superposition coding notion [16], which convey the information of the common message $\mathbf{s}$, and the codewords $\mathbf{x}$'s corresponding to the same $(\mathbf{t}, \mathbf{u})$ are called satellite codewords of $(\mathbf{t}, \mathbf{u})$, which contain both the common and private information. At the decoding stage, Receiver Z only needs to figure out which cloud $(\mathbf{t}, \mathbf{u})$ was transmitted, and Receiver Y needs to estimate not only the cloud but also the satellite codeword $\mathbf{x}$; see Fig. 9.2. We next employ superposition encoding to derive the achievable error exponent pair and the lower bound of system JSCC error exponent.

9.3 Universal Achievable Exponent Pair and a Lower Bound for E_J

Given arbitrary and finite alphabet $\mathcal{T}$, for any joint distribution $P_{TUX} \in \mathcal{P}(\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ and every $R_1 > 0$, $R_2 > 0$, define

$$\begin{aligned} E_Y(R_1, R_2, W_{Y|TUX}, P_{TUX}) &\triangleq \min_{V_{Y|TUX}} \Big[D(V_{Y|TUX} \parallel W_{Y|TUX} | P_{TUX}) \\ + \quad \min &\left(\left| I_{P_{TUX}V_{Y|TUX}}(T, U, X; Y) - (R_1 + R_2) \right|^+, \left| I_{P_{TUX}V_{Y|TUX}}(X; Y|T, U) - R_2 \right|^+ \right) \Big], \end{aligned} \tag{9.10}$$

and

$$\begin{aligned} &E_Z(R_1, R_2, W_{Z|TUX}, P_{TUX}) \\ &\quad \triangleq \min_{V_{Z|TUX}} \left[D(V_{Z|TUX} \parallel W_{Z|TUX} | P_{TUX}) + \left| I_{P_{TUX}V_{Z|TUX}}(T, U; Z) - R_1 \right|^+ \right], \end{aligned} \tag{9.11}$$

where $|x|^+ = \max(0, x)$, and the outer minimum in (9.10) (respectively (9.11)) is taken over all conditional distributions on $\mathcal{P}(\mathcal{Y}|\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ (respectively $\mathcal{P}(\mathcal{Z}|\mathcal{T} \times \mathcal{U} \times \mathcal{X})$). Note that E_Z only depends on R_1, but as will be seen, it is convenient to write E_Z as a function of R_1 and R_2. It immediately follows by definition that $E_Y(R_1, R_2, W_{Y|TUX}, P_{TUX})$ is zero if and only if at least one of the following is satisfied

$$R_1 + R_2 \geq I_{P_{TUX}W_{Y|TUX}}(T, U, X; Y), \tag{9.12}$$

$$R_2 \geq I_{P_{TUX}W_{Y|TUX}}(X; Y|T, U), \tag{9.13}$$

and $E_Z(R_1, R_2, W_{Z|TUX}, P_{TUX})$ is zero if and only if

$$R_1 \geq I_{P_{TUX}W_{Z|TUX}}(T, U; Z). \tag{9.14}$$

Using Lemma 3.2 and employing generalized maximum mutual information decoders at the two receivers, we can prove the following auxiliary bounds.

Proposition 9.1 *Given finite sets $\mathcal{T}$, $\mathcal{U}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, a sequence of positive integers $\{m_n\}$, and a sequence of positive integers $\{m'_{in}\}$ associated with every $i = 1, 2, ..., m_n$ with*

$$\frac{1}{n}\log_2 m_n \to 0 \qquad \text{and} \qquad \frac{1}{n}\log_2 \max_i m'_{in} \to 0,$$

for any $\delta > 0$, n sufficiently large, arbitrary (not necessarily distinct) types $P_{(TU)_i} \in \mathcal{P}_n(\mathcal{T} \times \mathcal{U})$ and conditional types $P_{X_j|(TU)_i} \in \mathcal{P}_n(\mathcal{X}|P_{(TU)_i})$, and positive integers N_i and M_{ij}, $i = 1, 2, ..., m_n$ and $j = j(i) = 1, 2, ..., m'_{in}$ with $R_i < H_{P_{(TU)_i}}(T, U) - \delta$ and $R_{ij} < H_{P_{(TU)_i}P_{X_j|(TU)_i}}(X|T, U) - \delta$, where $R_i \triangleq \frac{1}{n}\log_2 N_i$ and $R_{ij} \triangleq \frac{1}{n}\log_2 M_{ij}$, there exist m_n disjoint subsets $\Omega_i = \left\{(\boldsymbol{t}, \boldsymbol{u})_p^{(i)}\right\}_{p=1}^{N_i} \subseteq \mathbb{T}_{(TU)_i}$, m'_{in} disjoint subsets

$$\Omega_{ij}((\boldsymbol{t}, \boldsymbol{u})_p^{(i)}) = \left\{\left((\boldsymbol{t}, \boldsymbol{u})_p^{(i)}, \boldsymbol{x}_{p,q}^{(j)}\right)\right\}_{q=1}^{M_{ij}}$$

with $\boldsymbol{x}_{p,q}^{(j)} \in \mathbb{T}_{X_j|(TU)_i}((\boldsymbol{t}, \boldsymbol{u})_p^{(i)})$ for every $(\boldsymbol{t}, \boldsymbol{u})_p^{(i)} \in \Omega_i$ and every i, and a pair of mappings (decoding functions) $\varphi_n^{(0)} : \mathcal{Y}^n \to \Omega$ and $\psi_n^{(0)} : \mathcal{Z}^n \to \Omega$, where $\Omega \triangleq \bigcup_{ij} \Omega_{ij}$, where $\Omega_{ij} = \bigcup_{p=1}^{N_i} \Omega_{ij}((\boldsymbol{t}, \boldsymbol{u})_p^{(i)})$, such that the probabilities of erroneous transmission of a triplet

$(\boldsymbol{t},\boldsymbol{u},\boldsymbol{x}) \in \Omega$ over the augmented channel $W_{YZ|TUX}$ using decoders $(\varphi_n^{(0)}, \psi_n^{(0)})$ are simultaneously bounded by

$$P_{Ye}^{(n)}(\boldsymbol{t},\boldsymbol{u},\boldsymbol{x}) \triangleq \sum_{\boldsymbol{y}:\varphi_n^{(0)}(\boldsymbol{y})\neq((\boldsymbol{t},\boldsymbol{u}),\boldsymbol{x})} W_{Y|TUX}^{(n)}(\boldsymbol{y}|\boldsymbol{t},\boldsymbol{u},\boldsymbol{x}) \leq 2^{-n\left[E_Y\left(R_i,R_{ij},W_{Y|TUX},P_{(TU)_i}P_{X_j|(TU)_i}\right)-\delta\right]} \quad (9.15)$$

and

$$P_{Ze}^{(n)}(\boldsymbol{t},\boldsymbol{u},\boldsymbol{x}) \triangleq \sum_{\boldsymbol{z}:\psi_n^{(0)}(\boldsymbol{z})=((\boldsymbol{t},\boldsymbol{u})',\boldsymbol{x}')\text{ such that }(\boldsymbol{t},\boldsymbol{u})'\neq(\boldsymbol{t},\boldsymbol{u})} W_{Z|TUX}^{(n)}(\boldsymbol{z}|\boldsymbol{t},\boldsymbol{u},\boldsymbol{x}) \leq 2^{-n\left[E_Z\left(R_i,R_{ij},W_{Z|TUX},P_{(TU)_i}P_{X_j|(TU)_i}\right)-\delta\right]} \quad (9.16)$$

if $((\boldsymbol{t},\boldsymbol{u}),\boldsymbol{x}) \in \Omega_{ij}$ for every i, j.

Proof: We apply the packing lemma (Lemma 3.2) and a generalized maximum mutual information decoding rule. In the sequel of the proof, we look at TU (respectively X) as the RV A (respectively B) in Lemma 3.2. For the $\{m_n\}$, $\{m'_{in}\}$, $P_{(TU)_i}$, $P_{X_j|(TU)_i}$ given in Proposition 9.1, according to Lemma 3.2, there exist pairwise disjoint subsets Ω_i and $\Omega_{ij}((\mathbf{t},\mathbf{u})_p^{(i)})$ satisfying (3.4), (3.5), and (3.6) for every $1 \leq i \leq m_n$, $1 \leq j \leq m'_{in}$, $1 \leq p \leq N_i$, $V_{(TU)'|TU} \in \mathcal{P}_n(\mathcal{T}\times\mathcal{U}|\mathcal{T}\times\mathcal{U})$, and $V_{(TU)'X'|TUX} \in \mathcal{P}_n(\mathcal{T}\times\mathcal{U}\times\mathcal{X}|\mathcal{T}\times\mathcal{U}\times\mathcal{X})$, with the exception of the two cases that $i = k$ and $V_{(TU)'|TU}$ is the conditional distribution such that $V_{(TU)'|TU}((t,u)'|(t,u))$ is 1 if $(t,u)' = (t,u)$ and 0 otherwise, and that $i = k$, $j = l$ and $V_{(TU)'X'|TUX}$ is the conditional distribution such that $V_{(TU)'X'|TUX}((t,u)',x'|t,u,x)$ is 1 if $(t,u)' = (t,u)$, $x' = x$ and 0 otherwise. Let

$$\Omega_{ij} = \bigcup_{p=1}^{N_i} \Omega_{ij}((\mathbf{t},\mathbf{u})_p^{(i)}) \quad \text{and} \quad \Omega = \bigcup_{ij} \Omega_{ij}.$$

We shall show that for such Ω_{ij}, there exists a pair of mappings $(\varphi_n^{(0)}, \psi_n^{(0)})$ such that (9.15) and (9.16) are satisfied.

We first show that there exists a Y-decoder $\varphi_n^{(0)}$ such that (9.15) holds. For any $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega$ and $\mathbf{y} \in \mathcal{Y}^n$, let

$$\alpha((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{y}) \triangleq I((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{y}) - (R_i + R_{ij}),$$

where $R_i = \frac{1}{n}\log_2 N_i$ and $R_{ij} = \frac{1}{n}\log_2 M_{ij}$ if $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega_{ij}$. Define Y-decoder $\varphi_n^{(0)} : \mathcal{Y}^n \to \Omega$ by

$$\varphi_n^{(0)}(\mathbf{y}) \triangleq \arg\max_{((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega} \alpha((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{y}).$$

Using the decoder $\varphi_n^{(0)}$, we can upper bound the probability of error (assuming that $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega_{ij}$ is sent through the channel) as follows

$$\begin{aligned} P_{Ye}^{(n)}((\mathbf{t},\mathbf{u}),\mathbf{x}) &= W_{Y|TUX}^{(n)}\left(\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq ((\mathbf{t},\mathbf{u}),\mathbf{x})\right\}\Big|(\mathbf{t},\mathbf{u}),\mathbf{x}\right) \\ &\leq \sum_{\widehat{V}_{Y|TUX}\in\mathcal{P}_n(\mathcal{Y}|P_{(TU)_iX_j})} W_{Y|TUX}^{(n)}\left(\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq ((\mathbf{t},\mathbf{u}),\mathbf{x})\right\}\Big|\mathbf{t},\mathbf{u},\mathbf{x}\right) \\ &. \end{aligned} \tag{9.17}$$

For any particular $\widehat{V}_{Y|TUX}$, since

$$\begin{aligned} &\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq ((\mathbf{t},\mathbf{u}),\mathbf{x})\right\} \\ &= \underbrace{\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) = ((\mathbf{t},\mathbf{u})',\mathbf{x}'), (\mathbf{t},\mathbf{u})' \neq (\mathbf{t},\mathbf{u})\right\}}_{\triangleq\mathcal{E}_1} \bigcup \underbrace{\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) = ((\mathbf{t},\mathbf{u}),\mathbf{x}'), \mathbf{x}' \neq \mathbf{x}\right\}}_{\triangleq\mathcal{E}_2}, \end{aligned}$$

we can upper bound

$$\begin{aligned} &W_{Y|TUX}^{(n)}\left(\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\left\{\mathbf{y} : \varphi_n^{(0)}(\mathbf{y}) \neq ((\mathbf{t},\mathbf{u}),\mathbf{x})\right\}\Big|\mathbf{t},\mathbf{u},\mathbf{x}\right) \\ \leq &\sum_{\mathbf{y}\in\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{E}_1} W_{Y|TUX}^{(n)}(\mathbf{y}|\mathbf{t},\mathbf{u},\mathbf{x}) + \sum_{\mathbf{y}\in\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{E}_2} W_{Y|TUX}^{(n)}(\mathbf{y}|\mathbf{t},\mathbf{u},\mathbf{x}). \end{aligned} \tag{9.18}$$

According to Lemma 3.1 when $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega_{ij} \subseteq \mathbb{T}_{(TU)_iX_j}$ and $\mathbf{y} \in \mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})$, we have the following identity

$$W_{Y|TUX}^{(n)}(\mathbf{y}|(\mathbf{t},\mathbf{u}),\mathbf{x}) = 2^{-n\left[D\left(\widehat{V}_{Y|TUX}\|W_{Y|TUX}|P_{(TU)_iX_j}\right)+H_{P_{(TU)_iX_j}\widehat{V}_{Y|TUX}}(Y|T,U,X)\right]}. \tag{9.19}$$

This means that we only need to bound $\left|\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{E}_1\right|$ and $\left|\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{E}_2\right|$.

Upper Bound on $\left|\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{E}_1\right|$.

If we fix a $k = 1, 2, ..., m_n$ and a $l = 1, 2, ..., m'_{kn}$, then $\mathcal{E}_1$ is the set of all $\mathbf{y}$ such that there exist some $((\mathbf{t},\mathbf{u})',\mathbf{x}') \in \Omega_{kl}$, $(\mathbf{t},\mathbf{u})' \neq (\mathbf{t},\mathbf{u})$, $((\mathbf{t},\mathbf{u}),\mathbf{x},(\mathbf{t},\mathbf{u})',\mathbf{x}',\mathbf{y})$ admits a joint type

$P_{(\mathbf{t},\mathbf{u})\mathbf{x}(\mathbf{t},\mathbf{u})'\mathbf{x}'\mathbf{y}} \in \mathcal{P}_n(\mathcal{T} \times \mathcal{U} \times \mathcal{X} \times \mathcal{T} \times \mathcal{U} \times \mathcal{X} \times \mathcal{Y})$ and

$$I((\mathbf{t},\mathbf{u})',\mathbf{x}';\mathbf{y}) - (R_k + R_{kl}) \geq I((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{y}) - (R_i + R_{ij}). \tag{9.20}$$

Note that (9.20) can be represented as for dummy R.V.'s $(TU) \in \mathcal{T}\times\mathcal{U}$, $X \in \mathcal{X}$, $(TU)' \in \mathcal{T}\times\mathcal{U}$, $X' \in \mathcal{X}$, and $Y \in \mathcal{Y}$, the following holds under the joint distribution $P_{(TU)X(TU)'X'Y} = P_{(\mathbf{t},\mathbf{u})\mathbf{x}(\mathbf{t},\mathbf{u})'\mathbf{x}'\mathbf{y}}$,

$$I_{P_{(TU)'X'Y}}((T,U)',X';Y) - (R_k + R_{kl}) \geq I_{P_{TUXY}}((T,U),X;Y) - (R_i + R_{ij}),$$

where $P_{(TU)'X'Y}$ and P_{TUXY} are the corresponding marginal distributions induced by $P_{(TU)X(TU)'X'Y}$. Thus, $\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_1$ can be written as a union of subsets

$$\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_1 = \bigcup_{k=1}^{m_n} \bigcup_{l=1}^{m'_{kn}} \bigcup_{P_{(TU)X(TU)'X'Y} \in \mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x})} \mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y}) \tag{9.21}$$

where

$$\mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x}) \triangleq \left\{ \begin{array}{ll} & P_{(TU)X} = P_{(\mathbf{t},\mathbf{u})\mathbf{x}} = P_{(TU)_i X_j}, \\ P_{(TU)X(TU)'X'Y} & P_{(TU)'X'} = P_{(TU)_k X_l}, \quad P_{Y|(TU)X} = \widehat{V}_{Y|(TU)X}, \\ \in \mathcal{P}_n(\mathcal{T}^2 \times \mathcal{U}^2 \times \mathcal{X}^2 \times \mathcal{Y}) : & I_{P_{(TU)'X'Y}}((T,U)',X';Y) - (R_k + R_{kl}) \\ & \quad \geq I_{P_{(TU)XY}}((T,U),X;Y) - (R_i + R_{ij}) \end{array} \right\},$$

where $P_{(TU)X}$, $P_{(TU)'X'}$ and $P_{Y|(TU)X}$, etc, are the corresponding marginal and conditional distributions induced by $P_{(TU)X(TU)'X'Y}$, and

$$\mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y}) \triangleq \left\{ \mathbf{y} : \begin{array}{lll} \exists & ((\mathbf{t},\mathbf{u})',\mathbf{x}') & ((\mathbf{t},\mathbf{u}),\mathbf{x},(\mathbf{t},\mathbf{u})',\mathbf{x}',\mathbf{y}) \in \mathbb{T}_{(TU)X(TU)'X'Y} \\ \text{such that} & & ((\mathbf{t},\mathbf{u})',\mathbf{x}') \in \Omega_{kl}, \quad (\mathbf{t},\mathbf{u})' \neq (\mathbf{t},\mathbf{u}) \end{array} \right\},$$

where $\mathbb{T}_{(TU)X(TU)'X'Y} \triangleq \mathbb{T}_{P_{(TU)X(TU)'X'Y}}$. Clearly, given any k, l, and $P_{(TU)X(TU)'X'Y}$,

$$\begin{aligned}
&\left|\mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y})\right| \\
\leq\ & \left|\left\{((\mathbf{t},\mathbf{u})',\mathbf{x}',\mathbf{y}) : \begin{array}{l} ((\mathbf{t},\mathbf{u}),\mathbf{x},(\mathbf{t},\mathbf{u})',\mathbf{x}',\mathbf{y}) \in \mathbb{T}_{(TU)X(TU)'X'Y} \\ ((\mathbf{t},\mathbf{u})',\mathbf{x}') \in \Omega_{kl}, \quad (\mathbf{t},\mathbf{u})' \neq (\mathbf{t},\mathbf{u}) \end{array}\right\}\right| \\
=\ & \left|\left\{((\mathbf{t},\mathbf{u})',\mathbf{x}') : \begin{array}{l} ((\mathbf{t},\mathbf{u}),\mathbf{x},(\mathbf{t},\mathbf{u})',\mathbf{x}') \in \mathbb{T}_{(TU)X(TU)'X'} \\ ((\mathbf{t},\mathbf{u})',\mathbf{x}') \in \Omega_{kl}, \quad (\mathbf{t},\mathbf{u})' \neq (\mathbf{t},\mathbf{u}) \end{array}\right\}\right| \\
& \quad \times \left|\mathbb{T}_{Y|(TU)X(TU)'X'}((\mathbf{t},\mathbf{u}),\mathbf{x},(\mathbf{t},\mathbf{u})',\mathbf{x}')\right| \\
\leq\ & N_k M_{kl} 2^{-n\left[I_{P_{(TU)X(TU)'X'}}((T,U),X;(T,U)',X')-\eta\right]} \times 2^{nH_{P_{(TU)X(TU)'X'Y}}(Y|(T,U),X,(T,U)',X')}
\end{aligned} \tag{9.22}$$

where the last inequality follows from Lemma 3.2. Meanwhile, when $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega_{ij}$, the following simple bound also holds

$$\begin{aligned}
\left|\mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y})\right| &\leq \left|\mathbb{T}_{Y|(TU)X}((\mathbf{t},\mathbf{u}),\mathbf{x})\right| \\
&\leq 2^{nH_{P_{(TU)XY}}(Y|(T,U),X)} \\
&= 2^{nH_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X)}
\end{aligned} \tag{9.23}$$

since for each $\mathbb{T}_{(TU)X(TU)'X'Y} \in \mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x})$, we have $P_{(TU)X} = P_{((TU))_i X_j}$, $P_{Y|(TU)X} = \widehat{V}_{Y|(TU)X}$ and hence $P_{(TU)XY} = P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}$. Now substituting the following inequality (cf. (6.31))

$$\begin{aligned}
& H_{P_{(TU)X(TU)'X'Y}}(Y|(T,U),X,(T,U)',X') - I_{P_{(TU)X(TU)'X'}}((T,U),X;(T,U)',X') \\
=\ & H_{P_{(TU)XY}}(Y|(T,U),X) - I_{P_{(TU)X(TU)'X'Y}}((T,U)',X';(T,U),X,Y) \\
\leq\ & H_{P_{(TU)XY}}(Y|(T,U),X) - I_{P_{(TU)'X'Y}}((T,U)',X';Y)
\end{aligned} \tag{9.24}$$

into (9.22), combining with (9.23) together, we obtain

$$\begin{aligned}
&\left|\mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y})\right| \\
&\quad \leq 2^{n\left[H_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X) - \left|I_{P_{(TU)'X'Y}}((T,U)',X';Y)-(R_k+R_{kl})\right|^+\right]}.
\end{aligned} \tag{9.25}$$

Again recall that for $P_{(TU)X(TU)'X'Y} \in \mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x})$, $P_{(TU)XY} = P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}$, and note that

$$I_{P_{(TU)'X'Y}}((T,U)', X'; Y) - (R_k + R_{kl}) \geq I_{P_{(TU)XY}}((T,U), X; Y) - (R_i + R_{ij}).$$

This implies when $P_{(TU)X(TU)'X'Y} \in \mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x})$

$$\begin{aligned}&\left|\mathcal{F}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)X(TU)'X'Y})\right| \\ \leq\ & 2^{n\left[H_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X) - \left|I_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}((T,U),X;Y) - (R_i + R_{ij})\right|^+\right]},\end{aligned}$$

and hence

$$\begin{aligned}&\left|\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_1\right| \leq m_n \left(\max_i m'_{in}\right)(n+1)^{|\mathcal{T}\times\mathcal{U}|^2|\mathcal{X}|^2|\mathcal{Y}|} \\ &\times 2^{n\left[H_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X) - \left|(I_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}((T,U),X;Y) - (R_i + R_{ij}))\right|^+\right]}, \end{aligned} \tag{9.26}$$

since by Lemma 3.1

$$|\mathcal{C}_{k,l}((\mathbf{t},\mathbf{u}),\mathbf{x})| \leq \left|\mathcal{P}_n(\mathcal{T}^2 \times \mathcal{U}^2 \times \mathcal{X}^2 \times \mathcal{Y})\right| \leq (n+1)^{|\mathcal{T}\times\mathcal{U}|^2|\mathcal{X}|^2|\mathcal{Y}|}.$$

Upper Bound on $\left|\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_2\right|$.

If we fix an $i = 1, 2, ..., m_n$ and an $l = 1, 2, ..., m'_{in}$, then $\mathcal{E}_2$ is the set of all $\mathbf{y}$ such that there exist some $((\mathbf{t},\mathbf{u}),\mathbf{x}') \in \Omega_{il}$, $\mathbf{x}' \neq \mathbf{x}$, $((\mathbf{t},\mathbf{u}),\mathbf{x},\mathbf{x}',\mathbf{y})$ admits a joint type $P_{(\mathbf{t},\mathbf{u})\mathbf{x}\mathbf{x}'\mathbf{y}} \in \mathcal{P}_n(\mathcal{T}\times\mathcal{U}\times\mathcal{X}\times\mathcal{X}\times\mathcal{Y})$ and

$$I((\mathbf{t},\mathbf{u}),\mathbf{x}';\mathbf{y}) - (R_i + R_{il}) \geq I((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{y}) - (R_i + R_{ij}). \tag{9.27}$$

Using the identity

$$I((T,U),X;Y) = I(T,U;Y) + I(X;Y|T,U),$$

on both sides of (9.27) we see it is equivalent to

$$I(\mathbf{x}';\mathbf{y}|\mathbf{t},\mathbf{u}) - R_{il} \geq I(\mathbf{x};\mathbf{y}|\mathbf{t},\mathbf{u}) - R_{ij}. \tag{9.28}$$

Note that (9.28) can be represented as for dummy R.V.'s $(TU) \in \mathcal{T} \times \mathcal{U}$, $X \in \mathcal{X}$, $X' \in \mathcal{X}$, and $Y \in \mathcal{Y}$, the following holds under the joint distribution $P_{(TU)XX'Y} = P_{(\mathbf{t},\mathbf{u})\mathbf{x}\mathbf{x}'\mathbf{y}}$,

$$I_{P_{(TU)X'Y}}(X';Y|T,U) - R_{il} \geq I_{P_{(TU)XY}}(X;Y|T,U) - R_{ij},$$

where $P_{(TU)XY}$ and $P_{(TU)X'Y}$ are the corresponding marginal distributions induced by $P_{(TU)XX'Y}$. Thus, $\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_2$ can be written as a union of subsets

$$\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_2 = \bigcup_{l=1}^{m'_{in}} \bigcup_{P_{(TU)XX'Y} \in \mathcal{C}_l((\mathbf{t},\mathbf{u}),\mathbf{x})} \mathcal{F}_l((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)XX'Y}) \tag{9.29}$$

where

$$\mathcal{C}_l((\mathbf{t},\mathbf{u}),\mathbf{x}) \triangleq \left\{ \begin{array}{ll} & P_{(TU)X} = P_{(\mathbf{t},\mathbf{u})\mathbf{x}} = P_{(TU)_i X_j}, \\ P_{(TU)XX'Y} & P_{(TU)X'} = P_{(TU)_i X_l}, \quad P_{Y|(TU)X} = \widehat{V}_{Y|TUX} \\ \in \mathcal{P}_n(\mathcal{T} \times \mathcal{U} \times \mathcal{X}^2 \times \mathcal{Y}) : & I_{P_{(TU)X'Y}}(X';Y|T,U) - R_{il} \\ & \quad \geq I_{P_{(TU)XY}}(X;Y|T,U) - R_{ij} \end{array} \right\},$$

where $P_{(TU)X}$, $P_{(TU)X'}$ and $P_{Y|(TU)X}$, etc, are the corresponding marginal and conditional distributions induced by $P_{(TU)XX'Y}$, and

$$\mathcal{F}_l((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)XX'Y}) \triangleq \left\{ \mathbf{y} : \begin{array}{lll} \exists & ((\mathbf{t},\mathbf{u}),\mathbf{x}') & ((\mathbf{t},\mathbf{u}),\mathbf{x},\mathbf{x}',\mathbf{y}) \in \mathbb{T}_{(TU)XX'Y} \\ \text{such that} & & ((\mathbf{t},\mathbf{u}),\mathbf{x}') \in \Omega_{il}, \quad \mathbf{x}' \neq \mathbf{x} \end{array} \right\},$$

where $\mathbb{T}_{(TU)XX'Y} = \mathbb{T}_{P_{(TU)XX'Y}}$. Using a similar counting argument, and applying Lemma 3.2, we can bound, for any $l = 1, 2, ..., m'_{in}$ and $P_{(TU)XX'Y} \in \mathcal{C}_l((\mathbf{t},\mathbf{u}),\mathbf{x})$,

$$\begin{aligned} &\left|\mathcal{F}_l((\mathbf{t},\mathbf{u}),\mathbf{x},P_{(TU)XX'Y})\right| \\ &\quad \leq \ 2^{n\left[H_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X) - \left|I_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(X;Y|T,U) - R_{ij}\right|^+\right]}, \end{aligned}$$

and finally, we obtain,

$$\begin{aligned} &\left|\mathbb{T}_{\widehat{V}_{Y|(TU)X}}((\mathbf{t},\mathbf{u}),\mathbf{x}) \bigcap \mathcal{E}_2\right| \leq \left(\max_i m'_{in}\right)(n+1)^{|\mathcal{T}\times\mathcal{U}||\mathcal{X}|^2|\mathcal{Y}|} \\ &\times 2^{n\left[H_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(Y|(T,U),X) - \left|I_{P_{((TU))_i X_j}\widehat{V}_{Y|(TU)X}}(X;Y|T,U) - R_{ij}\right|^+\right]} \end{aligned} \tag{9.30}$$

since $|\mathcal{C}_l((\mathbf{t},\mathbf{u}),\mathbf{x})| \le (n+1)^{|\mathcal{T}\times\mathcal{U}||\mathcal{X}|^2|\mathcal{Y}|}$.

Now using (9.19) together with (9.26) and (9.30), we obtain

$$\sum_{\mathbf{y}\in\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{A}} W^{(n)}_{Y|TUX}(\mathbf{y}|((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega_{ij}) \le m_n\left(\max_i m'_{in}\right)(n+1)^{|\mathcal{T}\times\mathcal{U}|^2|\mathcal{X}|^2|\mathcal{Y}|}$$
$$\times 2^{-n\left[D\left(\widehat{V}_{Y|TUX}\|W_{Y|TUX}|P_{(TU)_iX_j}\right)+\left|I_{P_{(TU)_iX_j}\widehat{V}_{Y|TUX}}(T,U,X;Y)-(R_i+R_{ij})\right|^+\right]}, \tag{9.31}$$

and

$$\sum_{\mathbf{y}\in\mathbb{T}_{\widehat{V}_{Y|TUX}}((\mathbf{t},\mathbf{u}),\mathbf{x})\bigcap\mathcal{B}} W^{(n)}_{Y|TUX}(\mathbf{y}|((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega_{ij}) \le \left(\max_i m'_{in}\right)(n+1)^{|\mathcal{T}\times\mathcal{U}||\mathcal{X}|^2|\mathcal{Y}|}$$
$$\times 2^{-n\left[D\left(\widehat{V}_{Y|TUX}\|W_{Y|TUX}|P_{(TU)_iX_j}\right)+\left|I_{P_{(TU)_iX_j}\widehat{V}_{Y|TUX}}(X;Y|T,U)-R_{ij}\right|^+\right]}. \tag{9.32}$$

Substituting (9.31) and (9.32) back into (9.18) and (9.17) successively, noting that

$$|\mathcal{P}_n(\mathcal{Y}|P_{(TU)_iX_j})|$$

is polynomial in n by Lemma 3.1, we obtain that, for any $\delta>0$, there exists a Y-decoder $\varphi_n^{(0)}$ such that, given $((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega_{ij}$, the probability of Y-error is bounded by

$$P^{(n)}_{Ye}((\mathbf{t},\mathbf{u}),\mathbf{x}) \le 2^{-n\left[E_Y\left(R_i,R_{ij},W_{Y|TUX},P_{(TU)_i}P_{X_j|(TU)_i}\right)-\delta\right]} \tag{9.33}$$

for sufficiently large n.

Similarly, we can design a decoder for Receiver Z as follows. For any $((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega$ and $\mathbf{z}\in\mathcal{Z}^n$, let

$$\beta((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{z}) = \beta((\mathbf{t},\mathbf{u});\mathbf{z}) \triangleq I((\mathbf{t},\mathbf{u});\mathbf{z})-R_i,$$

where $R_i=\frac{1}{n}\log_2 N_i$ if $(\mathbf{t},\mathbf{u})\in\Omega_i$. Note that $\beta((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{z})$ is independent of $\mathbf{x}$. Let $\widetilde{\Omega}=\sum_{i=1}^{m_n}\Omega_i$. The Z-decoder $\psi_n^{(0)}:\mathcal{Z}^n\to\Omega$ is defined by

$$\begin{aligned}\varphi_n^{(0)}(\mathbf{z}) &= \arg\max_{((\mathbf{t},\mathbf{u}),\mathbf{x})\in\Omega}\beta((\mathbf{t},\mathbf{u}),\mathbf{x};\mathbf{z})\\ &= ((\mathbf{t},\mathbf{u})',\mathbf{x}') \quad\text{such that}\begin{cases}(\mathbf{t},\mathbf{u})'=\arg\max_{(\mathbf{t},\mathbf{u})\in\widetilde{\Omega}}\beta((\mathbf{t},\mathbf{u});\mathbf{z}),\\ \mathbf{x}' \text{ is arbitrary.}\end{cases}\end{aligned}$$

It can be shown in a similar manner by using (3.4) in Lemma 3.2 that, under the decoder $\psi_n^{(0)}$, the probability of the Z-error is bounded by

$$P_{Ze}^{(n)}((\mathbf{t},\mathbf{u}),\mathbf{x}) \leq 2^{-n\left[E_Z\left(R_i,R_{ij},W_{Z|TUX},P_{(TU)_i}P_{X_j|(TU)_i}\right)-\delta\right]} \tag{9.34}$$

for sufficiently large n. Finally, we remark that Lemma 3.2 ensures that there exist mappings $(\varphi_n^{(0)},\psi_n^{(0)})$ such that (9.34) holds simultaneously with (9.33). ■

Proposition 9.1 is an auxiliary result for the channel coding problem for the 2-user asymmetric channel. To apply it to our 2-user source-channel system, we need to design encoders which can map a pair of correlated source messages to a particular $(\mathbf{t},\mathbf{u},\mathbf{x})$ with a joint type, so that the total probabilities of error still vanish exponentially. We hence can establish the following bounds.

Theorem 9.1 *Given an arbitrary and finite alphabet $\mathcal{T}$, for any $\widetilde{P}_{TUX} \in \mathcal{P}(\mathcal{T}\times\mathcal{U}\times\mathcal{X})$, the following exponent pair is universally achievable,*

$$\begin{aligned}&E_{JY}(Q_{SL},W_{YZ|TUX},\widetilde{P}_{TUX},\tau)\\ &\triangleq \min_{P_{SL}}\left[\tau D(P_{SL}\parallel Q_{SL})+E_Y(\tau H_P(S),\tau H_P(L|S),W_{Y|TUX},\widetilde{P}_{TUX})\right],\end{aligned} \tag{9.35}$$

and

$$\begin{aligned}&E_{JZ}(Q_{SL},W_{YZ|TUX},\widetilde{P}_{TUX},\tau)\\ &\triangleq \min_{P_{SL}}\left[\tau D(P_{SL}\parallel Q_{SL})+E_Z(\tau H_P(S),\tau H_P(L|S),W_{Z|TUX},\widetilde{P}_{TUX})\right],\end{aligned} \tag{9.36}$$

where $W_{Y|TUX}$ and $W_{Z|TUX}$ are marginal distributions of $W_{YZ|TUX}$, which is the augmented conditional distribution from $W_{YZ|UX}$. Furthermore, given Q_{SL}, $W_{YZ|UX}$, and τ, the system JSCC error exponent satisfies

$$E_J(Q_{SL},W_{YZ|UX},\tau) \geq \min_{P_{SL}}\left[\tau D(P_{SL}\parallel Q_{SL})+E_r(\tau H_P(S),\tau H_P(L|S),W_{YZ|UX})\right] \tag{9.37}$$

where

$$E_r(R_1,R_2,W_{YZ|UX}) \triangleq \sup_{\mathcal{T}}\max_{P_{TUX}} E_r(R_1,R_2,W_{YZ|TUX},P_{TUX}), \tag{9.38}$$

where the supremum is taken over all finite alphabets $\mathcal{T}$, and the maximum is taken over all the joint distributions on $\mathcal{P}(\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ and $E_r(R_1, R_2, W_{YZ|TUX}, P_{TUX})$ is given by

$$\min\left\{E_Y(R_1, R_2, W_{Y|TUX}, P_{TUX}), E_Z(R_1, R_2, W_{Z|TUX}, P_{TUX})\right\},$$

where E_Y and E_Z are given by (9.10) and (9.11), respectively.

We remark that (9.35) and (9.36) can be achieved by a sequence of codes without the knowledge of Q_{SL} and $W_{YZ|UX}$, but the lower bound (9.37) is achieved by a sequence of codes that needs to know the statistics of the channel.

Proof of Theorem 9.1: We first prove the achievable error exponent pair (9.35) and (9.36). We need to show that, for any given $\widetilde{P}_{TUX} \in \mathcal{P}(\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ and $\delta > 0$, there exists a sequence of JSC codes such that both the probabilities of decoding error are upper bounded by

$$P_{ke}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \leq 2^{-n[E_{Jk}(Q_{SL}, W_{YZ|TUX}, \widetilde{P}_{TUX}, \tau) - \delta]}, \qquad k = Y, Z,$$

where E_{JY} and E_{JZ} are given by (9.35) and (9.36).

To apply Proposition 9.1, set $m_n \triangleq |\mathcal{P}_{\tau n}(\mathcal{S})|$. For each type $P_{S_i} \in \mathcal{P}_{\tau n}(\mathcal{S})$, $i = 1, 2, ..., m_n$, denote N_i be the cardinalities of these type classes, $N_i \triangleq |\mathbb{T}_{S_i}|$, and set $m'_{in} \triangleq |\mathcal{P}_{\tau n}(\mathcal{L}|P_{S_i})|$. For each conditional type $P_{L_j|S_i} \in \mathcal{P}_{\tau n}(\mathcal{L}|P_{S_i})$, $j = 1, 2, ..., m'_{in}$, denote M_{ij} be the cardinalities of these type classes, $M_{ij} \triangleq |\mathbb{T}_{L_j|S_i}(\mathbf{s})|$ where $\mathbf{s}$ is an arbitrary sequence in $\mathbb{T}_{S_i}$. Note that $|\mathbb{T}_{L_j|S_i}(\mathbf{s})|$ is constant for all $\mathbf{s} \in \mathbb{T}_{S_i}$. R_i and R_{ij} are respectively given by $\frac{1}{n}\log_2 N_i$ and $\frac{1}{n}\log_2 M_{ij}$.

Now no matter whether the given $\widetilde{P}_{TUX}$ belongs to $\mathcal{P}_n(\mathcal{T} \times \mathcal{U} \times \mathcal{X})$ or not, we always can find a sequence of joint types $\{P_{TUX} \in \mathcal{P}_n(\mathcal{T} \times \mathcal{U} \times \mathcal{X})\}_{n=1}^{\infty}$ such that $P_{TUX} \to \widetilde{P}_{TUX}$ uniformly[1] as $n \to \infty$. Thus, we can choose, by the continuity of $E_k(R_i, R_{ij}, W_{k|TUX}, \widetilde{P}_{TUX})$ with respect to $\widetilde{P}_{TUX}$, for each $i = 1, 2, ..., m_n$, and $j = j(i) = 1, 2, ..., m'_{in}$, the joint type $P_{(TU)_i X_j} = P_{TUX}$ such that the following are satisfied

$$\left|E_k(R_i, R_{ij}, W_{k|TUX}, P_{TUX}) - E_k(R_i, R_{ij}, W_{k|TUX}, \widetilde{P}_{TUX})\right| < \frac{\delta}{4}, \qquad k = Y, Z$$

[1]We say that a sequence of distributions $\{P_{X_i} \in \mathcal{P}(\mathcal{X})\}_{i=1}^{\infty}$ uniformly converges to $P_X^* \in \mathcal{P}(\mathcal{X})$ if the variational distance between P_{X_i} and P_X^* converges to zero as $n \to \infty$.

for n sufficiently large. Since the type P_{TUX} can also be regarded as a joint distribution, let $P_{(TU)_i} = P_{TU} \in \mathcal{P}_n(\mathcal{T} \times \mathcal{U})$ be the marginal distribution on $\mathcal{T} \times \mathcal{U}$ induced by P_{TUX} and let $P_{X_j|(TU)_i} = P_{X|TU} \in \mathcal{P}_n(\mathcal{X}|P_{TU})$ be the corresponding conditional distribution, i.e., $P_{X|TU}(\mathbf{x}|\mathbf{t},\mathbf{u}) = P_{TUX}(\mathbf{t},\mathbf{u},\mathbf{x})/P_{TU}(\mathbf{t},\mathbf{u})$ for any $(\mathbf{t},\mathbf{u},\mathbf{x}) \in \mathbb{T}_{TUX}$.

Without loss of generality, we assume, for the choice of N_i, M_{ij}, $P_{(TU)_i}$, and $P_{X_j|(TU)_i}$, the following conditions are satisfied for $i = 1, 2, ..., \widehat{m}_n$, $j = 1, 2, ..., \widehat{m}'_{in}$,

$$R_i < H_{P_{(TU)_i}}(T, U) - \frac{\delta}{4}, \qquad i = 1, 2, ..., \widehat{m}_n \tag{9.39}$$

and

$$R_{ij} < H_{P_{(TU)_i X_j}}(X|T, U) - \frac{\delta}{4}, \qquad i = 1, 2, ..., \widehat{m}_n, \quad j = j(i) = 1, 2, ..., \widehat{m}'_{in}, \tag{9.40}$$

where $\widehat{m}_n \le m_n$ and $\widehat{m}'_{in} \le m'_n$. Then according to Proposition 9.1, there exist pairwise disjoint subsets $\Omega_{ij} \subseteq \mathbb{T}_{(TU)_i X_j}$ with $|\Omega_{ij}| = N_i M_{ij}$, $i = 1, 2, ..., \widehat{m}_n$, $j = 1, 2, ..., \widehat{m}'_{in}$, and a pair of mappings $\left(\varphi_n^{(0)}, \psi_n^{(0)}\right)$, such that the probabilities of erroneous transmission of a $((\mathbf{t},\mathbf{u}),\mathbf{x}) \in \Omega_{ij}$ are *simultaneously* bounded for the channel $W_{YZ|TUX}$ as

$$\begin{aligned} P_{Ye}^{(n)}(\mathbf{t},\mathbf{u},\mathbf{x}) &\le 2^{-n\left[E_Y\left(R_i, R_{ij}, W_{Y|TUX}, P_{(TU)_i X_j}\right) - \delta/4\right]} \\ &\le 2^{-n[E_Y(R_i, R_{ij}, W_{Y|TUX}, \widetilde{P}_{TUX}) - \delta/2]} \end{aligned} \tag{9.41}$$

and

$$\begin{aligned} P_{Ze}^{(n)}(\mathbf{t},\mathbf{u},\mathbf{x}) &\le 2^{-n\left[E_Z\left(R_i, R_{ij}, W_{Z|TUX}, P_{(TU)_i X_j}\right) - \delta/4\right]} \\ &\le 2^{-n[E_Z(R_i, R_{ij}, W_{Z|TUX}, \widetilde{P}_{TUX}) - \delta/2]}. \end{aligned} \tag{9.42}$$

For the N_i, M_{ij}, $P_{(TU)_i}$, and $P_{X_j|(TU)_i}$ violating (9.39) or (9.40) (i.e., for $i > \widehat{m}_n$ or $j > \widehat{m}'_{in}$), (9.41) and (9.42) trivially hold for arbitrary choice of disjoint subsets Ω_{ij} since $E_Y\left(R_i, R_{ij}, W_{Y|TUX}, P_{(TU)_i X_j}\right)$ or $E_Z\left(R_i, R_{ij}, W_{Z|TUX}, P_{(TU)_i X_j}\right)$ would be less than $\delta/4$. In fact, the functions E_Y and E_Z are trivially bounded by the following linear functions of R_i and R_{ij} with slope -1,

$$\begin{aligned} E_Y\left(R_i, R_{ij}, W_{Y|TUX}, P_{(TU)_i X_j}\right) \le \min\Big\{ & I_{P_{(TU)_i X_j} W_{Y|TUX}}(T, U, X; Y) - R_i - R_{ij}, \\ & I_{P_{(TU)_i X_j} W_{Y|TUX}}(X; Y|T, U) - R_{ij} \Big\} \end{aligned} \tag{9.43}$$

and

$$E_Z\left(R_i, R_{ij}, W_{Z|TUX}, P_{(TU)_iX_j}\right) \le I_{P_{(TU)_iX_j}W_{Z|TUX}}(T,U;Z) - R_i. \tag{9.44}$$

If

$$R_i \ge H_{P_{(TU)_i}}(T,U) - \frac{\delta}{4} \ge I_{P_{(TU)_iX_j}W_{Z|TUX}}(T,U;Z) - \frac{\delta}{4},$$

then by (9.44) $E_Z\left(R_i, R_{ij}, W_{Z|TUX}, P_{(TU)_iX_j}\right) \le \frac{\delta}{4}$. Similarly, if

$$R_{ij} \ge H_{P_{(TU)_iX_j}}(X|T,U) - \frac{\delta}{4},$$

then by (9.43) $E_Y\left(R_i, R_{ij}, W_{Y|TUX}, P_{(TU)_iX_j}\right) \le \frac{\delta}{4}$.

Therefore, we may construct the JSC code $(f_n, g_n, \varphi_n, \psi_n)$ for CS Q_{SL} and the 2-user channel $W_{YZ|UX}$ as follows.

Encoder g_n: For the message $\mathbf{s} \in \mathbb{T}_{S_i}$ such that $i > \widehat{m}_n$, let $g_n(\mathbf{s}) = \mathbf{0} \in \mathcal{U}^n$. Denote $\widetilde{\Omega} = \bigcup_i \Omega_i$. For the $\mathbf{s} \in \mathbb{T}_{S_i}$ such that $i \le \widehat{m}_n$, let $g_n^{(1)} : \mathcal{S}^{\tau n} \to \widetilde{\Omega}$ be a bijection that maps each $\mathbf{s} \in \mathbb{T}_{S_i}$ to the corresponding $(\mathbf{t}, \mathbf{u}) \in \Omega_i$, by noting that $|\Omega_i| = |\mathbb{T}_{S_i}| = N_i$. Finally, let $g_n(\mathbf{s})$ be the second component $\mathbf{u}$ of $g_n^{(1)}(\mathbf{s})$.

Encoder f_n: For the message pair $(\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{S_iL_j}$ such that $i > \widehat{m}_n$ or $j > \widehat{m}'_{in}$, let $f_n(\mathbf{s}, \mathbf{l}) = \mathbf{0} \in \mathcal{X}^n$. For the $(\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{S_iL_j}$ such that $i \le \widehat{m}_n$ and $j \le \widehat{m}'_{in}$, noting that $|\mathbb{T}_{L_j|S_i}(\mathbf{s})| = |\Omega_{ij}(\varphi_n(\mathbf{s}))| = M_{ij}$ if $\mathbf{s} \in \mathbb{T}_{S_i}$, let $f_n^{(1)}(\mathbf{s}, \cdot) : \mathbb{T}_{L_j|S_i}(\mathbf{s}) \to \Omega_{ij}(g_n(\mathbf{s}))$ be a bijection such that $f_n^{(1)}(\mathbf{s}, \mathbf{l}) = (g_n^{(1)}(\mathbf{s}), \mathbf{x}) \in \Omega_{ij}$. Let $f_n(\mathbf{s}, \mathbf{l})$ be the third component $\mathbf{x}$ of $f_n^{(1)}(\mathbf{s}, \mathbf{l})$.

Clearly, the JSC encoders (f_n, g_n), although working independently, they map each $(\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{S_iL_j}$ to a unique pair $(\mathbf{u}, \mathbf{x})$ when $i \le \widehat{m}_n$ and $j \le \widehat{m}'_{in}$, and to $(\cdot, \mathbf{0})$ otherwise (in this case an error is declared).

Y-Decoder φ_n: The $Y-$decoder is defined by

$$\varphi_n(\mathbf{y}) \triangleq \begin{cases} (\mathbf{s}', \mathbf{l}') & \text{if } \exists \quad (\mathbf{s}', \mathbf{l}') \in \mathcal{S}^n \times \mathcal{L}^n \text{ such that } f_n^{(1)}(\mathbf{s}', \mathbf{l}') = \varphi_n^{(0)}(\mathbf{y}), \\ (\mathbf{0}, \mathbf{0}) & \text{Otherwise.} \end{cases}$$

Z-Decoder ψ_n: The $Z-$decoder is defined by

$$\psi_n(\mathbf{z}) \triangleq \begin{cases} \mathbf{s}' & \text{if } \exists \quad \mathbf{s}' \in \mathcal{S}^n \text{ such that } g_n^{(1)}(\mathbf{s}') \text{ are equal to} \\ & \text{the first two components of } \psi_n^{(0)}(\mathbf{z}), \\ \mathbf{0} & \text{Otherwise.} \end{cases}$$

For such JSC code $(f_n,\, g_n,\, \varphi_n,\, \psi_n)$, the probabilities of Y-error and Z-error are bounded by

$$P_{Ye}^{(n)}(\mathbf{s},\mathbf{l}) \leq 2^{-n[E_Y\left(R_i,R_{ij},W_{Y|TUX},\widetilde{P}_{TUX}\right)-\delta/2]} \qquad \text{if } (\mathbf{s},\mathbf{l}) \in \mathbb{T}_{S_iL_j} \tag{9.45}$$

and

$$P_{Ze}^{(n)}(\mathbf{s},\mathbf{l}) \leq 2^{-nE_Z\left(R_i,R_{ij},W_{Z|TUX},\widetilde{P}_{TUX}\right)-\delta/2]} \qquad \text{if } (\mathbf{s},\mathbf{l}) \in \mathbb{T}_{S_iL_j}. \tag{9.46}$$

Substituting (9.45) and (9.46) into (9.3) and using the fact (Lemma 3.1) $Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) \leq 2^{-n\tau D(P_{SL}\|Q_{SL})}$, we obtain, for n sufficiently large,

$$\begin{aligned} & P_{Ye}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \\ \leq \quad & \sum_{i,j} 2^{-n[\tau D(P_{S_iL_j}\|Q_{SL})+E_Y(R_i,R_{ij},W_{Y|TUX},\widetilde{P}_{TUX})-\delta/2]} \\ \leq \quad & \sum_{P_{SL}} 2^{-n\left[\tau D(P_{SL}\|Q_{SL})+E_Y\left(\tau H_P(S)-o_1(n)/n,\tau H_P(L|S)-o_2(n)/n,W_{Y|TUX},\widetilde{P}_{TUX}\right)-\delta/2\right]} \\ \leq \quad & \sum_{P_{SL}} 2^{-n\left[\tau D(P_{SL}\|Q_{SL})+E_Y\left(\tau H_P(S),\tau H_P(L|S),W_{Y|TUX},\widetilde{P}_{TUX}\right)-\delta\right]} \end{aligned} \tag{9.47}$$

and

$$\begin{aligned} & P_{Ze}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \\ \leq \quad & \sum_{i,j} 2^{-n[\tau D(P_{S_iL_j}\|Q_{SL})+E_Z(R_i,R_{ij},W_{Z|TUX},\widetilde{P}_{TUX})-\delta/2]} \\ \leq \quad & \sum_{P_{SL}} 2^{-n\left[\tau D(P_{SL}\|Q_{SL})+E_Z\left(\tau H_P(S)-o_1(n)/n,\tau H_P(L|S)-o_2(n)/n,W_{Z|TUX},\widetilde{P}_{TUX}\right)-\delta/2\right]} \\ \leq \quad & \sum_{P_{SL}} 2^{-n\left[\tau D(P_{SL}\|Q_{SL})+E_Z\left(\tau H_P(S),\tau H_P(L|S),W_{Z|TUX},\widetilde{P}_{TUX}\right)-\delta\right]}, \end{aligned} \tag{9.48}$$

where $o_1(n) = |\mathcal{S}|\log_2(\tau n+1)$ and $o_2(n) = |\mathcal{S}||\mathcal{L}|\log_2(\tau n+1)$. Finally, the bounds (9.35) and (9.36) follow from (9.47) and (9.48), and the fact that the cardinality of set of joint types $\mathcal{P}_{\tau n}(\mathcal{S}\times\mathcal{L})$ is upper bounded by $(\tau n+1)^{|\mathcal{S}||\mathcal{L}|}$.

To prove the lower bound (9.37), we slightly modify the above approach by choosing $P_{(TU)_i X_j} = \widetilde{P}^*_{(TU)_i X_j}$ which achieves the maximum and the supremum of $E_r(R_i, R_{ij}, W_{YZ|UX})$ in (9.38) for every R_i and R_{ij}, $i = 1, 2, ..., m_n$, $j = 1, 2, ..., m'_{in}$. Then the probabilities of Y-error and Z-error in (9.45) and (9.46) are bounded by

$$\begin{aligned} P^{(n)}_{Ye}(\mathbf{s}, \mathbf{l}) &\le 2^{-n[E_Y\left(R_i, R_{ij}, W_{Y|TUX}, \widetilde{P}^*_{(TU)_i X_j}\right) - \delta/2]} \\ &\le 2^{-n[E_r(R_i, R_{ij}, W_{YZ|UX}) - \delta/2]} \qquad \text{if } (\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{S_i L_j} \end{aligned} \tag{9.49}$$

and

$$\begin{aligned} P^{(n)}_{Ze}(\mathbf{s}, \mathbf{l}) &\le 2^{-nE_Z\left(R_i, R_{ij}, W_{Z|TUX}, \widetilde{P}^*_{TU_i X_j}\right) - \delta/2]} \\ &\le 2^{-n[E_r(R_i, R_{ij}, W_{YZ|UX}) - \delta/2]} \qquad \text{if } (\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{S_i L_j} \end{aligned} \tag{9.50}$$

for n sufficiently large. The rest of the proof is similar to the proofs of (9.35) and (9.36). ■

9.4 JSCC Theorem for the Asymmetric 2-User System

By examining the positivity of the lower bound to E_J, we obtain a sufficient condition for reliable transmissibility for the asymmetric 2-user system. For the sake of completeness, we also prove a converse by using Fano's inequality, and hence establish the JSCC theorem for this system. Given $W_{YZ|UX}$, define

$$\mathcal{R}(W_{YZ|UX}) \triangleq \bigcup_{\mathcal{T}: |\mathcal{T}| \le |\mathcal{U}||\mathcal{X}| + 2} \; \bigcup_{P_{TUX} \in \mathcal{P}(\mathcal{T} \times \mathcal{U} \times \mathcal{X})} \mathcal{R}(W_{YZ|TUX}, P_{TUX}) \tag{9.51}$$

where

$$\mathcal{R}(W_{YZ|TUX}, P_{TUX}) \triangleq \left\{ (R_1, R_2) : \begin{array}{l} R_1 + R_2 < I(T, U, X; Y) = I(U, X; Y) \\ R_1 < I(T, U; Z) \\ R_2 < I(X; Y|T, U) \end{array} \right\},$$

where the mutual informations are taken under the joint distribution $P_{TUXYZ} = P_{TUX} W_{YZ|UX}$. It can be shown that $\mathcal{R}(W_{YZ|UX})$ is convex and denote $\overline{\mathcal{R}}(W_{YZ|UX})$ be the closure of $\mathcal{R}(W_{YZ|UX})$.

Theorem 9.2 (JSCC Theorem) *Given Q_{SL}, $W_{YZ|UX}$ and $\tau > 0$, the following statements hold.*

(1) The sources Q_{SL} can be transmitted over the channel $W_{YZ|UX}$ with probability of error $P_e^{(n)} \to 0$ as $n \to \infty$ if $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX})$;

(2) Conversely, if the sources Q_{SL} can be transmitted over the channel $W_{YZ|UX}$ with an arbitrarily small probability of error $P_e^{(n)}$ as $n \to \infty$, then $(\tau H_Q(S), \tau H_Q(L|S)) \in \overline{\mathcal{R}}(W_{YZ|UX})$.

Proof:

Forward Part (1): It follows from (9.12)-(9.14) that $E_r(R_1, R_2, W_{YZ|TUX}, P_{TUX}) > 0$ if and only if $(R_1, R_2) \in \mathcal{R}(W_{YZ|TUX}, P_{TUX})$. It then follows that $E_r(R_1, R_2, W_{YZ|UX}) > 0$ if $(R_1, R_2) \in \mathcal{R}(W_{YZ|UX})$. According to Theorem 9.1 and the definition of the system JSCC error exponent, $P_e^{(n)} \to 0$ if the lower bound (9.37) is positive, which needs $E_r(\tau H_P(S), \tau H_P(L|S), W_{YZ|UX}) > 0$. This means $P_e^{(n)} \to 0$ if the pair

$$(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX}).$$

Converse Part (2): The proof follows from a similar manner as the converse part of [38, Theorem 1] for a broadcast channel. For the sake of completeness, we also provide a full proof here since we deal with a 2-user channel. First, we remark that (as shown in [38, Theorem 2]) the region $\mathcal{R}(W_{YZ|TUX}, P_{TUX})$ can be equivalently rewritten by

$$\mathcal{R}(W_{YZ|TUX}, P_{TUX}) = \left\{ (R_1, R_2) : \begin{array}{l} R_1 + R_2 < I(U, X; Y) \\ R_1 < I(T, U; Z) \\ R_1 + R_2 < I(X; Y|T, U) + I(T, U; Z) \end{array} \right\}.$$

It suffices to show that, for any $\epsilon > 0$, if

$$\max \left\{ P_{Ye}^{(n)}(Q_{SL}, W_{YZ|XU}, \tau), P_{Ze}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \right\} \leq \epsilon_n \to 0$$

as n goes to infinity, then there exists a RV T satisfying $T \to (U, X) \to (Y, Z)$, i.e., the joint distribution P_{TUXYZ} can be factorized as $P_T P_{UX|T} W_{YZ|UX}$, such that

$$(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX}, P_{TUX})$$

with < replaced by $\leq$, i.e.,

$$\begin{aligned}\tau H_Q(S,L) &\leq \min\{I(U,X;Y), I(X;Y|T,U)+I(T,U;Z)\},\\ \tau H_Q(S) &\leq I(T,U;Z).\end{aligned}$$

Fix $k=\tau n$. Fano's inequality gives

$$H(S^k,L^k|Y^n) \leq P_{Ye}^{(n)} \log_2 |\mathcal{S}^k \times \mathcal{L}^k| + H\left(P_{Ye}^{(n)}\right) \triangleq n\epsilon_{1n} \tag{9.52}$$
$$H(S^k|Z^n) \leq P_{Ze}^{(n)} \log_2 |\mathcal{S}^k| + H\left(P_{Ze}^{(n)}\right) \triangleq n\epsilon_{2n}, \tag{9.53}$$

where $S^k \triangleq (S_1, S_2, \cdots, S_k)$; similar definitions apply for the other tuples. It follows from (9.52)-(9.53) that

$$\begin{aligned}kH(S,L) &= H(L^k|S^k)+H(S^k)\\ &= I(L^k;Y^n|S^k)+H(L^k|S^k,Y^n)+I(S^k;Z^n)+H(S^k|Z^n)\\ &\leq \sum_{i=1}^{n}[I(L^k;Y_i|S^k,Y^{i-1})+I(S^k;Z_i|\mathbf{Z}^{i+1})]+H(S^k,L^k|Y^n)+n\epsilon_{2n}\\ &\leq \sum_{i=1}^{n}\Big[I(L^k;Y_i|S^k,Y^{i-1},\mathbf{Z}^{i+1})+I(\mathbf{Z}^{i+1};Y_i|S^k,Y^{i-1})\\ &\qquad +I(S^k,\mathbf{Z}^{i+1},Y^{i-1};Z_i)-I(Y^{i-1};Z_i|S^k,\mathbf{Z}^{i+1})\Big]+n(\epsilon_{1n}+\epsilon_{2n}),\end{aligned}$$

where $Y^{i-1}=(Y_1,Y_2,...,Y_{i-1})$ and $\mathbf{Z}^{i+1} \triangleq (Z_{i+1},Z_{i+2},...,Z_n)$. Substituting the identity [32, Lemma 7]

$$\sum_{i=1}^{n} I(\mathbf{Z}^{i+1};Y_i|S^k,Y^{i-1}) = \sum_{i=1}^{n} I(Y^{i-1};Z_i|S^k,\mathbf{Z}^{i+1})$$

into the above, and setting $T_i=(S^k,Y^{i-1},\mathbf{Z}^{i+1})$ for $1\leq i\leq n$ yields

$$\begin{aligned}kH(S,L) &\leq \sum_{i=1}^{n}\Big[I(L^k;Y_i|T_i)+I(T_i;Z_i)\Big]+n(\epsilon_{1n}+\epsilon_{2n})\\ &\overset{(a)}{=} \sum_{i=1}^{n}\Big[I(L^k;Y_i|T_i,U_i)+I(T_i,U_i;Z_i)\Big]+n(\epsilon_{1n}+\epsilon_{2n})\\ &\overset{(b)}{\leq} \sum_{i=1}^{n}[I(X^n;Y_i|T_i,U_i)+I(T_i,U_i;Z_i)]+n(\epsilon_{1n}+\epsilon_{2n})\\ &\overset{(c)}{=} \sum_{i=1}^{n}[I(X_i;Y_i|T_i,U_i)+I(T_i,U_i;Z_i)]+n(\epsilon_{1n}+\epsilon_{2n}),\end{aligned} \tag{9.54}$$

where (a) holds since U_i is a deterministic function of S^k and hence of T_i, (b) follows from the data processing inequality, and (c) holds since Y_i is only determined by U_i and X_i due to the memoryless property of the channel. On the other hand, $kH(S,L)$ can also be bounded by

$$\begin{aligned} kH(S,L) &= H(S^k, L^k) \\ &= I(S^k, L^k; Y^n) + H(S^k, L^k | Y^n) \\ &\leq I(X^n, U^n; Y^n) + n\epsilon_{1n} \\ &= \sum_{i=1}^{n} I(U_i, X_i; Y_i) + n\epsilon_{1n}. \end{aligned} \tag{9.55}$$

Likewise, it follows from (9.53) that

$$\begin{aligned} kH(S) &= H(S^k) \\ &= I(S^k; Z^n) + H(S^k | Z^n) \\ &= \sum_{i=1}^{n} I(S^k; Z_i | \mathbf{Z}^{i+1}) + H(S^k | Z^n) \\ &\leq \sum_{i=1}^{n} I(S^k, \mathbf{Z}^{i+1}; Z_i) + n\epsilon_{2n} \\ &\leq \sum_{i=1}^{n} I(S^k, Y^{i-1}, \mathbf{Z}^{i+1}, U_i; Z_i) + n\epsilon_{2n} \\ &= \sum_{i=1}^{n} I(T_i, U_i; Z_i) + n\epsilon_{2n}. \end{aligned} \tag{9.56}$$

Note also that $T_i \longrightarrow (U_i, X_i) \longrightarrow (Y_i, Z_i)$ for all $1 \leq i \leq n$. According to (9.54), (9.55), and (9.56), and recalling that $k = \tau n$, it is easy to show (e.g., see [32]) that there exists an auxiliary RV T with $P_{TUXYZ} = P_T P_{UX|T} W_{YZ|UX}$ such that

$$\begin{aligned} \tau H(S,L) &\leq \min\{I_{P_{UXYZ}}(U,X;Y), I_{P_{TUXYZ}}(X;Y|T,U) + I_{P_{TUXYZ}}(T,U;Z)\} \\ \tau H(S) &\leq I_{P_{TUXYZ}}(T,U;Z), \end{aligned}$$

which is equivalent to

$$\begin{aligned} \tau H(S,L) &\leq I_{P_{UXYZ}}(U,X;Y), \\ \tau H(S) &\leq I_{P_{TUXYZ}}(T,U;Z), \\ \tau H(L|S) &\leq I_{P_{TUXYZ}}(X;Y|T,U). \end{aligned}$$

Finally, by using the Carathéodory theorem (cf. [32, p. 311]) we can show that there exists a RV $\widehat{T}$ with $|\widehat{\mathcal{T}}| \leq |\mathcal{U}||\mathcal{X}| + 1$ such that $P_{\widehat{T}UXYZ} = P_{\widehat{T}} P_{UX|\widehat{T}} W_{YZ|UX}$ and

$$\begin{aligned}(I_{P_{UXYZ}}(U, X; Y), I_{P_{TUXYZ}}(T, U; Z), I_{P_{TUXYZ}}(X; Y|T, U)) \\ = (I_{P_{UXYZ}}(U, X; Y), I_{P_{\widehat{T}UXYZ}}(\widehat{T}, U; Z), I_{P_{\widehat{T}UXYZ}}(X; Y|\widehat{T}, U)).\end{aligned}$$

This completes the proof of the converse part. ■

9.5 Separation Principle for the Asymmetric 2-User System

It can be verified that the condition $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX})$ of Theorem 9.2 can be achieved by separate source and channel coding. The separate coding system of rate τ (source symbol/channel symbol) (we refer to it as the *tandem* coding system) is depicted in Figs. 9.3 and 9.4

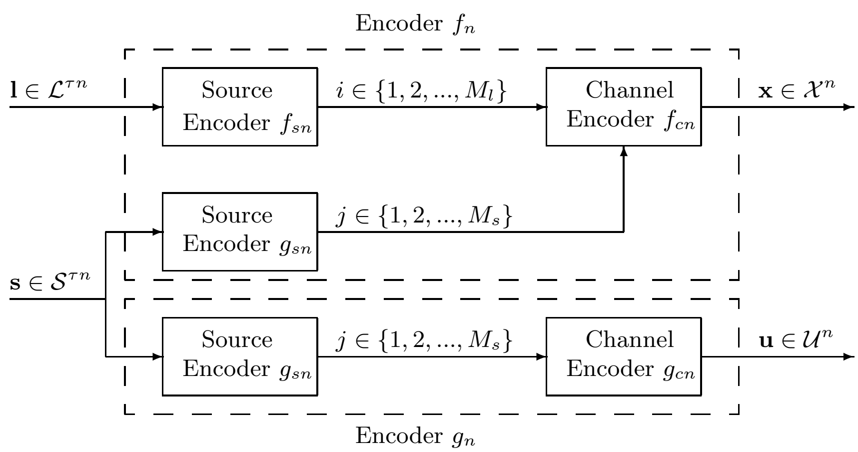

Figure 9.3: Tandem source-channel coding system - encoders.

The encoder f_n is composed of two source encoders $f_{sn} : \mathcal{L}^{\tau n} \to \{1, 2, ..., M_l\}$ and $g_{sn} : \mathcal{S}^{\tau n} \to \{1, 2, ..., M_s\}$ with private coding rate $\widehat{R}_l \triangleq \frac{1}{\tau n} \log_2 M_l$ and common coding rate $\widehat{R}_s \triangleq \frac{1}{\tau n} \log_2 M_s$ and a channel encoder $\{1, 2, ..., M_l\} \times \{1, 2, ..., M_s\} \to \mathcal{X}^n$. Similarly, the encoder g_n is composed of a source encoder $g_{sn} : \mathcal{S}^{\tau n} \to \{1, 2, ..., M_s\}$ with common coding rate $\widehat{R}_s$ and a channel encoder $g_{cn} : \{1, 2, ..., M_s\} \to \mathcal{U}^n$.

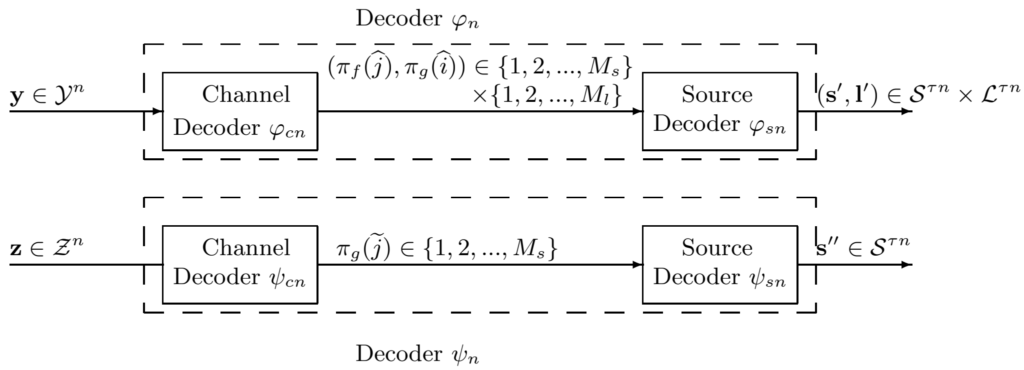

Figure 9.4: Tandem source-channel coding system - decoders.

At the receiver side, the decoder φ_n is composed of a channel decoder $\varphi_{cn} : \mathcal{Y}^n \to \{1, 2, ..., M_l\} \times \{1, 2, ..., M_s\}$, and a source decoder $\varphi_{sn} : \{1, 2, ..., M_l\} \times \{1, 2, ..., M_s\} \to \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n}$ which outputs the approximation of the source messages $\mathbf{s}'$ and $\mathbf{l}'$. Similarly, the decoder ψ_n is composed of a channel decoder $\psi_{cn} : \mathcal{Z}^n \to \{1, 2, ..., M_s\}$, and a source decoder $\psi_{sn} : \{1, 2, ..., M_s\} \to \mathcal{S}^{\tau n}$.

To show that the condition $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX})$ can be achieved by the above tandem system, we need to apply the following 2-user source and channel coding theorems (we only state the forward parts of the theorems). Note that both of these theorems are special case of Theorem 9.2.

Let $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ be a sequence of source codes for CS Q_{SL} with common source rate $\widehat{R}_s$ and private source rate $\widehat{R}_l$ as defined above. The probability of the overall 2-user source coding error is given by

$$P_{es}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) \triangleq \Pr\left(\{\varphi_{sn}(g_{sn}(S^{\tau n}), f_{sn}(L^{\tau n})) \neq (S^{\tau n}, L^{\tau n})\} \bigcup \{\psi_{sn}(g_{sn}(S^{\tau n})) \neq S^{\tau n}\}\right). \tag{9.57}$$

Then by the 2-user source coding theorem, there exists a sequence of source codes $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ with rates $\widehat{R}_s$ and $\widehat{R}_l$ such that $P_{es}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) \to 0$ as $n \to \infty$ if the rates satisfy $\widehat{R}_s > H_Q(S)$ and $\widehat{R}_l > H_Q(L|S)$, i.e., $(\widehat{R}_s, \widehat{R}_l)$ lies in the upper-right infinite rectangle with vertex given by the point $(H_Q(S), H_Q(L|S))$.

We next state the forward part of channel coding theorem for the asymmetric 2-user channel. Let the (common and private) message pair (j, i) be uniformly drawn from the finite set $\mathcal{M}_s \times \mathcal{M}_l$, where $\mathcal{M}_s \triangleq \{1, 2, ..., M_s\}$ and $\mathcal{M}_l \triangleq \{1, 2, ..., M_l\}$, and let $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$ be an asymmetric 2-user channel code with block length n and common and private message sets $\mathcal{M}_s$ and $\mathcal{M}_l$. Let $R_s \triangleq \frac{1}{n}\log_2 M_s$ and $R_l \triangleq \frac{1}{n}\log_2 M_l$ be the common and private rates of the channel code, respectively. The average probability of error for asymmetric 2-user channel coding is given by

$$P_{ec}^{(n)}(R_s, R_l, W_{YZ|UX}) \triangleq \Pr\left(\{\varphi_{cn}(Y^n) \neq (J, I)\}\bigcup\{\psi_{cn}(Z^n) \neq J\}\right), \tag{9.58}$$

where (J, I) are uniformly drawn from $\mathcal{M}_s \times \mathcal{M}_l$. The maximum probability for error of asymmetric 2-user channel coding is given by

$$\begin{aligned} &P_{ec,max}^{(n)}(R_s, R_l, W_{YZ|UX}) \\ &\triangleq \max_{(j,i)\in\mathcal{M}_s\times\mathcal{M}_l} \Pr\left(\{\varphi_{cn}(Y^n) \neq (J, I)\}\bigcup\{\psi_{cn}(Z^n) \neq J\}\,\middle|\, J = j, I = i\right), \end{aligned} \tag{9.59}$$

Then there exists a sequence of channel codes $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$ such that $P_{ec}^{(n)}(R_s, R_l, W_{YZ|UX})$ goes to 0 as $n \to \infty$ if $(R_s, R_l) \in \mathcal{R}(W_{YZ|UX})$. Furthermore, it can be readily shown by a standard expurgation argument [29, p. 204] that $P_{ec,max}^{(n)}(R_s, R_l, W_{YZ|UX}) \to 0$ as $n \to \infty$ if $(R_s, R_l) \in \mathcal{R}(W_{YZ|UX})$.

Now by (9.7), the overall probability of error for the tandem system is given by

$$P_e^{(n)} \triangleq \Pr\left(\{\varphi_{sn}\left[\varphi_{cn}(Y^n)\right] \neq (S^{\tau n}, L^{\tau n})\}\bigcup\{\psi_{sn}\left[\psi_{cn}(Z^n)\right] \neq S^{\tau n}\}\right).$$

By the union bound, it is easy to see that $P_e^{(n)}$ is upper bounded by

$$\begin{aligned} P_e^{(n)} &\leq \Pr\left(\{\varphi_{sn}(g_{sn}(S^{\tau n}), f_{sn}(L^{\tau n})) \neq (S^{\tau n}, L^{\tau n})\}\bigcup\{\psi_{sn}(g_{sn}(S^{\tau n})) \neq S^{\tau n}\}\right) \\ &+ \Pr\left(\{\varphi_{cn}(Y^n) \neq (g_{sn}(S^{\tau n}), f_{sn}(L^{\tau n}))\}\bigcup\{\psi_{cn}(Z^n) \neq g_{sn}(S^{\tau n})\}\right) \\ &= P_{es}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) + \sum_{(j,i)\in\mathcal{M}_s\times\mathcal{M}_l} \Pr\left(g_{sn}(S^{\tau n}) = j, f_{sn}(L^{\tau n}) = i\right) \\ &\qquad \Pr\left(\{\varphi_{cn}(Y^n) \neq (J, I)\}\bigcup\{\psi_{cn}(Z^n) \neq J\}\,\middle|\, J = j, I = i\right) \\ &\leq P_{es}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) + P_{ec,max}^{(n)}(\tau\widehat{R}_s, \tau\widehat{R}_l, W_{YZ|UX}) \end{aligned}$$

where $P_{ec,max}^{(n)}(\tau\widehat{R}_s, \tau\widehat{R}_l, W_{YZ|UX})$ is the maximum channel coding probability of error with common rate $\tau\widehat{R}_s$ and private rate $\tau\widehat{R}_l$. Clearly, by combining the 2-user source coding theorem and the asymmetric 2-user channel coding theorem, if $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX})$, then there exist a sequence of source codes $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ and a sequence of channel codes $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$ such that the overall tandem system probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. Therefore, separation of source and channel coding is optimal from the point of view of reliable transmissibility.

9.6 An Upper Bound for E_J

We know that Csiszár also established an upper bound for the JSCC error exponent for the point-to-point discrete memoryless source-channel system in terms of the source and channel error exponents by a simple type counting argument. He shows that the JSCC error exponent is always less than the infimum of the sum of the source and channel error exponent, even though the channel error exponent is only partially known for high rates. This conceptual bound cannot currently be computed as the channel error exponent is not yet fully known for all achievable coding rates, but it directly implies that any upper bound for the channel error exponent yields a corresponding upper bound for the JSCC error exponent. For the asymmetric 2-user channel, a similar bound can be shown.

As a special case of the JSCC system, let the (common and private) message pair (j, i) be uniformly drawn from the finite set $\mathcal{M}_s \times \mathcal{M}_l$, where $\mathcal{M}_s \triangleq \{1, 2, ..., M_s\}$ and $\mathcal{M}_l \triangleq \{1, 2, ..., M_l\}$. An asymmetric 2-user channel code with block length n for transmitting the uniform message set is a quadruple of mappings, $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$, where $f_{cn} : \mathcal{M}_s \times \mathcal{M}_l \to \mathcal{X}^n$ and $g_{cn} : \mathcal{M}_s \to \mathcal{U}^n$ are the channel encoders, and $\varphi_{cn} : \mathcal{Y}^n \to \mathcal{M}_s \times \mathcal{M}_l$ and $\psi_{cn} : \mathcal{Z}^n \to \mathcal{M}_s$ are respectively the Y-decoder and Z-decoder. Let $R_s \triangleq \frac{1}{n} \log_2 M_s$ and $R_l \triangleq \frac{1}{n} \log_2 M_l$ be the common and private rates of the code, respectively. The probabilities of Y- and Z-error of the channel coding are respectively given by

$$P_{Yec}^{(n)}(R_s, R_l, W_{YY|UX}) \triangleq \Pr(\{\varphi_{cn}(Y^n) \neq (J, I)\}) = \frac{1}{2^{R_1+R_2}} \sum_{\mathcal{M}_s \times \mathcal{M}_l} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y}) \neq (j,i)} W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{u}, \mathbf{x}) \tag{9.60}$$

and

$$P_{Zec}^{(n)}(R_s, R_l, W_{YZ|UX}) \triangleq \Pr(\{\psi_{cn}(Z^n) \neq J\}) = \frac{1}{2^{R_1+R_2}} \sum_{\mathcal{M}_s \times \mathcal{M}_l} \sum_{\mathbf{z}:\psi_{cn}(\mathbf{z})\neq j} W_{Z|X}^{(n)}(\mathbf{z}|\mathbf{u}, \mathbf{x}) \tag{9.61}$$

where $\mathbf{x} \triangleq f_{cn}(j, i)$ and $\mathbf{u} \triangleq g_{cn}(j)$. Similarly, the probability of the overall asymmetric 2-user channel coding error is given by

$$P_{ec}^{(n)}(R_s, R_l, W_{YZ|UX}) \triangleq \Pr\left(\{\varphi_{cn}(Y^n) \neq (J, I)\} \cup \{\psi_{cn}(Z^n) \neq J\}\right), \tag{9.62}$$

where (J, I) are uniformly drawn from $\mathcal{M}_s \times \mathcal{M}_l$.

Definition 9.2 The asymmetric 2-user channel coding error exponent $E(R_1, R_2, W_{YZ|UX})$, for any $R_1 > 0$ and $R_2 > 0$, is defined by the supremum of the set of all numbers E_c for which there exists a sequence of asymmetric channel codes $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$ with blocklength n, the common rate no less than R_1, and the private rate no less than R_2, such that

$$E_c \leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(R_1, R_2, W_{YZ|UX}). \tag{9.63}$$

Clearly, for any sequence of channel codes $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$, $P_e^{(n)}(R_1, R_2, W_{YZ|UX})$ must be larger than $P_{Ye}^{(n)}(R_1, R_2, W_{Y|UX})$ and $P_{Ze}^{(n)}(R_1, R_2, W_{Z|UX}))$ but less than the sum of the two, so we have

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(R_1, R_2, W_{YZ|UX}) \\ &= \liminf_{n\to\infty} -\frac{1}{n}\log_2 \max\left(P_{Ye}^{(n)}(R_1, R_2, W_{Y|UX}), P_{Ze}^{(n)}(R_1, R_2, W_{Z|UX})\right). \end{aligned} \tag{9.64}$$

Our upper bound for the system JSCC error exponent is stated as follows.

Theorem 9.3 *Given Q_{SL}, $W_{YZ|UX}$, and τ, the system JSCC error exponent satisfies*

$$E_J(Q_{SL}, W_{YZ|UX}, \tau) \leq \inf_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E(\tau H_P(S), \tau H_P(L|S), W_{YZ|UX})\right], \tag{9.65}$$

where $E(\cdot, \cdot, W_{YZ|UX})$ is the corresponding channel coding error exponent for the asymmetric 2-user channel.

Proof: First, we write from (9.3) that

$$P_{ie}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \geq \max_{P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S}\times\mathcal{L})} Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) P_{ie}(\mathbb{T}_{SL}) \quad i = Y, Z, \tag{9.66}$$

where $P_{Ye}(\mathbb{T}_{SL})$ and $P_{Ze}(\mathbb{T}_{SL})$ are given by (9.4) and (9.5). Comparing (9.4) with (9.60), and comparing (9.5) with (9.61), we note that $P_{Ye}(\mathbb{T}_{SL})$ and $P_{Ze}(\mathbb{T}_{SL})$ can be interpreted as the probabilities of Y-error and Z-error of the asymmetric 2-user channel coding with (common and private) message sets $\mathbb{T}_{SL}$, since $(\mathbf{s}, \mathbf{l})$ are uniformly distributed on $\mathbb{T}_{SL}$. For any $P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L})$, let P_S and $P_{L|S}$ be the marginal and conditional distributions induced by P_{SL}. Recall that for each $\mathbf{s} \in \mathbb{T}_S = \mathbb{T}_{P_S}$,

$$\mathbb{T}_{L|S}(\mathbf{s}) \triangleq \mathbb{T}_{P_{L|S}}(\mathbf{s}) = \{\mathbf{l} : (\mathbf{s}, \mathbf{l}) \in \mathbb{T}_{SL}\}$$

and that $\mathbb{T}_{L|S}(\mathbf{s})$ is the same set for all $\mathbf{s} \in \mathbb{T}_S$. Hence, we can write $\mathbb{T}_{SL}$ by the product of two sets $\mathbb{T}_{SL} = \mathbb{T}_S \times \mathbb{T}_{L|S}(\mathbf{s})$. Setting $\widetilde{R}_1 = \frac{1}{n}\log_2 |\mathbb{T}_S|$ and $\widetilde{R}_2 = \frac{1}{n}\log_2 |\mathbb{T}_{L|S}(\mathbf{s})|$, it follows that, by the definition of asymmetric 2-user channel coding error exponent and (10.51),

$$\begin{aligned} \liminf_{n\to\infty} -\frac{1}{n}\log_2 \max_{i=Y,Z} P_{ie}(\mathbb{T}_{SL}) &\leq E(\liminf_{n\to\infty} \widetilde{R}_1, \liminf_{n\to\infty} \widetilde{R}_2, W_{YZ|UX}) \\ &= E(\tau H_P(S), \tau H_P(L|S), W_{YZ|UX}) \end{aligned} \tag{9.67}$$

for any sequence of JSC codes (f_n, φ_n, ψ_n), recalling Lemma 3.1 that

$$(\tau n + 1)^{-|\mathcal{S}|} 2^{n\tau H_P(S)} \leq |\mathbb{T}_S| \leq 2^{n\tau H_P(S)}$$

and

$$(\tau n + 1)^{-|\mathcal{S}||\mathcal{L}|} 2^{n\tau H_P(L|S)} \leq |\mathbb{T}_{L|S}(\mathbf{s})| \leq 2^{n\tau H_P(L|S)}.$$

According to (9.9), we write

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_e^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \\ &= \liminf_{n\to\infty} -\frac{1}{n}\log_2 \max\left(P_{Ye}^{(n)}(Q_{SL}, W_{Y|X}, \tau), P_{Ze}^{(n)}(Q_{SL}, W_{Z|X}, \tau)\right) \\ &\leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 \max_{i=Y,Z} \max_{P_{SL}\in\mathcal{P}_{\tau n}(\mathcal{S}\times\mathcal{L})} Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) P_{ie}(\mathbb{T}_{SL}) \\ &= \liminf_{n\to\infty} \min_{P_{SL}\in\mathcal{P}_{\tau n}(\mathcal{S}\times\mathcal{L})} -\frac{1}{n}\log_2 Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) \max_{i=Y,Z} P_{ie}(\mathbb{T}_{SL}) \\ &= \liminf_{n\to\infty} \min_{P_{SL}\in\mathcal{P}_{\tau n}(\mathcal{S}\times\mathcal{L})} \left[-\frac{1}{n}\log_2 Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) - \frac{1}{n}\log_2 \max_{i=Y,Z} P_{ie}(\mathbb{T}_{SL})\right]. \end{aligned} \tag{9.68}$$

By Lemma 3.1, for any $P_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L})$,

$$-\frac{1}{\tau n} \log_2 Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) \leq D(P_{SL} \parallel Q_{SL}) + |\mathcal{S}||\mathcal{L}| \frac{1}{\tau n} \log_2(1 + \tau n)$$

which implies

$$\limsup_{n\to\infty} -\frac{1}{n} \log_2 Q_{SL}^{(\tau n)}(\mathbb{T}_{SL}) \leq \tau D(P_{SL} \parallel Q_{SL}). \tag{9.69}$$

Now assume that

$$\inf_{P_{SL}\in\mathcal{P}(\mathcal{S}\times\mathcal{L})} \left[\tau D(P_{SL} \parallel Q_{SL}) + E(\tau H_P(S), \tau H_P(L|S), W_{YZ|UX})\right]$$

is finite (the upper bound is trivial if it is infinity) and the infimum actually becomes a minimum. Let the minimum be achieved by distribution $P_{SL}^* \in \mathcal{P}(\mathcal{S} \times \mathcal{L})$, then there must exists a sequence of types $\left\{\widehat{P}_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L})\right\}_{n=n_o}^{\infty}$ such that $\widehat{P}_{SL} \to P_{SL}^*$ uniformly[2]. It then follows from (9.68), (9.67) and (9.69) that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n} \log_2 P_e^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \\
&\leq \liminf_{n\to\infty} \left[-\frac{1}{n} \log_2 Q_{SL}^{(\tau n)}(\mathbb{T}_{\widehat{P}_{SL}}) - \frac{1}{n} \log_2 \max_{i=Y,Z} P_{ie}(\mathbb{T}_{\widehat{P}_{SL}})\right] \\
&\leq \tau D(P_{SL}^* \parallel Q_{SL}) + E(\tau H_{P^*}(S), \tau H_{P^*}(L|S), W_{YZ|UX}).
\end{aligned} \tag{9.70}$$

Since the above bound holds for any sequence of JSC codes, we complete the proof of Theorem 9.3. ■

9.7 Applications to CS-AMAC and CS-ABC Systems

As pointed out in the introduction, our results obtained in the previous section can be directly applied to the CS-AMAC and CS-ABC source-channel systems.

9.7.1 CS-AMAC System

Setting $|\mathcal{Z}| = 1$ and removing the decoder ψ_n, the 2-user asymmetric channel $W_{YZ|UX}$ reduces to an AMAC $W_{Y|UX}$. Since the CS-AMAC system is a special case of the 2-user

---

[2] The sequence $\left\{\widehat{P}_{SL} \in \mathcal{P}_{\tau n}(\mathcal{S} \times \mathcal{L})\right\}_{n=n_o}^{\infty}$ here denotes a sequence for $n = n_o, 2n_o, 3n_o, ...,$ where n_o is the smallest integer such that τn_o is also an integer.

system, the quantities defined before, including the system (overall) probability of error, the system JSCC error exponent, and also the channel error exponent still hold for the CS-AMAC system. Note that there is only one decoder, so we do not have the Z-error probability (nor exponent) here. The first union in (9.51) can be removed since the largest region is given by $|\mathcal{T}| = 1$. In fact, for any $T \to (U, X) \to Y$, $I(T, U, X; Y) = I(U, X; Y)$ and $I(X; Y|T, U) \le I(X; Y|U)$. Thus Theorem 9.2 reduces to the same JSCC theorem established in [20] for the CS-AMAC system.

Given $W_{Y|UX}$, $\mathcal{R}(W_{YZ|UX})$ of (9.51) reduces to $\mathcal{R}(W_{Y|UX})$ given by

$$\mathcal{R}(W_{Y|UX}) \triangleq \bigcup_{P_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})} \mathcal{R}(W_{Y|UX}, P_{UX}) \tag{9.71}$$

where

$$\mathcal{R}(W_{Y|TX}, P_{UX}) = \left\{ (R_1, R_2) : \begin{array}{l} R_1 + R_2 < I(U, X; Y) \\ R_2 < I(X; Y|U) \end{array} \right\},$$

where the mutual informations are taken under the joint distribution $P_{UXY} = P_{UX} W_{Y|UX}$. We remark that the following JSCC theorem for the CS-AMAC system coincides with the one established in [20]. Note that $\mathcal{R}(W_{Y|TX}, P_{UX})$ is convex and denote $\overline{\mathcal{R}}(W_{Y|TX}, P_{UX})$ be the closure of $\mathcal{R}(W_{Y|TX}, P_{UX})$.

Corollary 9.1 (JSCC Theorem for CS-AMAC system [20]) *Given Q_{SL}, $W_{Y|UX}$ and $\tau > 0$, the sources can be transmitted over the channel with $P_e^{(n)} \to 0$ as $n \to \infty$ if*

$$(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{Y|UX});$$

Conversely, if the sources can be transmitted over the channel with an arbitrarily small probability of error $P_e^{(n)}$ as $n \to \infty$, then $(\tau H_Q(S), \tau H_Q(L|S)) \in \overline{\mathcal{R}}(W_{Y|UX}$.

To specialize Theorems 9.1 and 9.3 to the CS-AMAC system, we simply choose the auxiliary alphabet $|\mathcal{T}| = 1$, which yields the following corollaries.

Corollary 9.2 *Given Q_{SL}, $W_{Y|UX}$ and τ, the system JSCC error exponent satisfies*

$$E_J(Q_{SL}, W_{Y|UX}, \tau) \ge \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_r(\tau H_P(S), \tau H_P(L|S), W_{Y|UX}) \right], \tag{9.72}$$

and

$$E_J(Q_{SL}, W_{Y|UX}, \tau) \leq \inf_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E(\tau H_P(S), \tau H_P(L|S), W_{Y|UX})\right], \quad (9.73)$$

where $E(\tau H_P(S), \tau H_P(L|S), W_{Y|UX})$ *is the channel error exponent of the AMAC* $W_{Y|UX}$ *defined in (9.63) with* $|\mathcal{Z}| = 1$*, and*

$$E_r(R_1, R_2, W_{Y|UX}) = \max_{P_{UX}} E_Y(R_1, R_2, W_{Y|UX}, P_{UX}) \quad (9.74)$$

where $E_Y(R_1, R_2, W_{Y|UX}, P_{UX})$ *is defined in (9.10) with* $|\mathcal{T}| = 1$*.*

It has been shown in [10] that for any $R_1 > 0$ and $R_2 > 0$, the channel exponent for AMAC $W_{Y|UX}$ satisfies

$$E(R_1, R_2, W_{YZ|X}) \leq E_{sp}(R_1, R_2, W_{Y|UX}),$$

where

$$E_{sp}(R_1, R_2, W_{Y|UX}) \triangleq \max_{P_{UX} \in \mathcal{P}(\mathcal{U}\times\mathcal{X})} \min D(V_{Y|UX} \parallel W_{Y|UX}|P_{UX}), \quad (9.75)$$

where the minimum is taken over $V_{Y|UX} \in \mathcal{P}(\mathcal{Y}|\mathcal{U} \times \mathcal{X})$ such that $I_{P_{UX}V_{Y|UX}}(U, X; Y) \leq R_1 + R_2$ or $I_{P_{UX}V_{Y|UX}}(X; Y|U) \leq R_2$.

As a consequence, we obtain that

$$E_J(Q_{SL}, W_{Y|UX}, \tau) \leq \inf_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_{sp}(\tau H_P(S), \tau H_P(L|S), W_{Y|UX})\right]. \quad (9.76)$$

In Section 9.8, we investigate the evaluation of lower bound (9.72) and upper bound (9.76) when the AMAC has a symmetric distribution.

9.7.2 CS-ABC System

Setting $|\mathcal{U}| = 1$ and removing the encoder g_n, the 2-user asymmetric channel $W_{YZ|UX}$ reduces to an ABC $W_{YZ|X}$. The quantities defined before, including the probabilities of error at Y-decoder and Z-decoder, the achievable error exponent pair, system (overall) probability of error, the system JSCC error exponent, and the channel error exponent still hold for the CS-ABC system. Given an arbitrary and finite auxiliary alphabet $\mathcal{T}$, we augment the channel $W_{YZ|X}$ to $W_{YZ|TX}$ by introducing a RV $T \in \mathcal{T}$ such that $T \to X \to (YZ)$. Similarly, the marginal distributions of the augmented channel are denoted by $W_{Y|TX}$ and $W_{Z|TX}$. We then specialize Theorems 9.1, 9.2 and 9.3 to the following corollaries.

Given $W_{YZ|X}$, $\mathcal{R}(W_{YZ|UX})$ of (9.51) reduces to $\mathcal{R}(W_{YZ|X})$ given by

$$\mathcal{R}(W_{YZ|X}) \triangleq \bigcup_{\mathcal{T}:|\mathcal{T}|\leq|\mathcal{X}|+2} \bigcup_{P_{TX}\in\mathcal{P}(\mathcal{T}\times\mathcal{X})} \mathcal{R}(W_{YZ|TX}, P_{TX}) \tag{9.77}$$

where

$$\mathcal{R}(W_{YZ|TX}, P_{TX}) = \left\{ (R_1, R_2) : \begin{array}{l} R_1 + R_2 < I(T, X; Y) = I(X; Y) \\ R_1 < I(T; Z) \\ R_2 < I(X; Y|T) \end{array} \right\},$$

where the mutual informations are taken under the joint distribution $P_{TXYZ} = P_{TX}W_{YZ|X}$. We remark that the closure of $\mathcal{R}(W_{YZ|X})$, denoted by $\overline{\mathcal{R}}(W_{YZ|X})$, is the capacity region of the ABC $W_{YZ|X}$ [59].

Corollary 9.3 (JSCC Theorem for CS-ABC system) *Given Q_{SL}, $W_{YZ|X}$ and $\tau > 0$, the following statements hold.*

(1) The sources Q_{SL} can be transmitted over the ABC $W_{YZ|X}$ with $P_e^{(n)} \to 0$ as $n \to \infty$ if $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|X})$;

(2) Conversely, if the sources Q_{SL} can be transmitted over the ABC $W_{YZ|X}$ with an arbitrarily small probability of error $P_e^{(n)}$ as $n \to \infty$, then $(\tau H_Q(S), \tau H_Q(L|S)) \in \overline{\mathcal{R}}(W_{YZ|X})$..

Corollary 9.4 *Given an arbitrary and finite alphabet $\mathcal{T}$, for any $\widetilde{P}_{TX} \in \mathcal{P}(\mathcal{T} \times \mathcal{X})$, the following exponent pair is universally achievable,*

$$\begin{aligned} &E_{JY}(Q_{SL}, W_{YZ|TX}, \widetilde{P}_{TX}, \tau) \\ &\triangleq \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_Y(\tau H_P(S), \tau H_P(L|S), W_{Y|TX}, \widetilde{P}_{TX}) \right], \end{aligned} \tag{9.78}$$

and

$$\begin{aligned} &E_{JZ}(Q_{SL}, W_{YZ|TX}, \widetilde{P}_{TX}, \tau) \\ &\triangleq \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_Z(\tau H_P(S), \tau H_P(L|S), W_{Z|TX}, \widetilde{P}_{TX}) \right], \end{aligned} \tag{9.79}$$

where E_Y and E_Z are defined in (9.10) and (9.11) by setting $|\mathcal{U}| = 1$. Furthermore, given Q_{SL}, $W_{YZ|X}$, and τ, the system JSCC error exponent satisfies

$$E_J(Q_{SL}, W_{YZ|X}, \tau) \geq \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_r(\tau H_P(S), \tau H_P(L|S), W_{YZ|X})\right] \quad (9.80)$$

and

$$E_J(Q_{SL}, W_{YZ|X}, \tau) \leq \inf_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E(\tau H_P(S), \tau H_P(L|S), W_{YZ|X})\right] \quad (9.81)$$

where $E_r(R_1, R_2, W_{YZ|X})$ is given by $E_r(R_1, R_2, W_{YZ|UX})$ in (9.38) with $|\mathcal{U}| = 1$, and $E(R_1, R_2, W_{YZ|X})$ is the channel error exponent for the ABC $W_{YZ|X}$.

9.8 Evaluation of the Bounds for E_J

We established the lower and upper bounds for the JSCC error exponent of the asymmetric 2-user JSCC system. However, we are not able to simplify these bounds for general 2-user JSCC systems (not even for general CS-AMAC and CS-ABC systems) into computable parametric forms as we did for the point-to-point systems [107, 109]. In the following, we only address a special case of CS-AMAC systems where the channel admits a symmetric transition probability distribution. We first introduce the parametric forms of functions $E_r(R_1, R_2, W_{Y|UX})$ and $E_{sp}(R_1, R_2, W_{Y|UX})$ defined in (9.74) and (9.75), respectively. For any $R_1, R_2 > 0$, rewrite

$$E_Y(R_1, R_2, W_{Y|UX}, P_{UX}) = \min\left\{E_r^{(1)}(R_1 + R_2, W_{Y|UX}, P_{UX}), E_r^{(2)}(R_2, W_{Y|UX}, P_{UX})\right\}$$

where

$$E_r^{(1)}(R, W_{Y|UX}, P_{UX}) \triangleq \min_{V_{Y|UX}} \left[D(V_{Y|UX} \parallel W_{Y|UX}|P_{UX}) + \left|I_{P_{UX}V_{Y|UX}}(U, X; Y) - R\right|^+\right] \quad (9.82)$$

and

$$E_r^{(2)}(R, W_{Y|UX}, P_{UX}) \triangleq \min_{V_{Y|UX}} \left[D(V_{Y|UX} \parallel W_{Y|UX}|P_{UX}) + \left|I_{P_{UX}V_{Y|UX}}(X; Y|U) - R\right|^+\right]. \quad (9.83)$$

Also, rewrite

$$E_{sp}(R_1, R_2, W_{Y|UX}) = \max_{P_{UX}} E_{sp}(R_1, R_2, W_{Y|UX}, P_{UX})$$

where

$$E_{sp}(R_1, R_2, W_{Y|UX}, P_{UX}) = \min\left\{E_{sp}^{(1)}(R_1 + R_2, W_{Y|UX}, P_{UX}), E_{sp}^{(2)}(R_2, W_{Y|UX}, P_{UX})\right\}$$

where

$$E_{sp}^{(1)}(R, W_{Y|UX}, P_{UX}) \triangleq \min_{V_{Y|UX}} \Big(D(V_{Y|UX} \parallel W_{Y|UX}|P_{UX}) : \quad I_{P_{UX}V_{Y|UX}}(U, X; Y) \leq R\Big) \tag{9.84}$$

and

$$E_{sp}^{(2)}(R, W_{Y|UX}, P_{UX}) \triangleq \min_{V_{Y|UX}} \Big(D(V_{Y|UX} \parallel W_{Y|UX}|P_{UX}) : \quad I_{P_{UX}V_{Y|UX}}(X; Y|U) \leq R\Big). \tag{9.85}$$

Note that $E_r^{(1)}$ and $E_r^{(2)}$ (respectively $E_{sp}^{(1)}$ and $E_{sp}^{(2)}$) are the random-coding (respectively sphere-packing) type exponents expressed in terms of constrained Kullback-Leibler divergences and mutual informations [32]. In fact, it has been shown in [10] that

$$E_{sp}^{(i)}(R, W_{Y|UX}, P_{UX}) = \max_{\rho \geq 0}[E_i(\rho, W_{Y|UX}, P_{UX}) - \rho R], \quad i = 1, 2,$$

where

$$E_1(\rho_1, W_{Y|UX}, P_{UX}) \triangleq -\log_2 \sum_{y \in \mathcal{Y}} \left(\sum_{(u,x) \in \mathcal{U} \times \mathcal{X}} P_{UX}(u, x) W_{Y|UX}(y|u, x)^{\frac{1}{1+\rho_1}} \right)^{1+\rho_1}, \tag{9.86}$$

and

$$E_2(\rho_2, W_{Y|UX}, P_{UX}) = -\log_2 \sum_{u \in \mathcal{U}} P_U(u) \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_{X|U}(x|u) W_{Y|UX}(y|u, x)^{\frac{1}{1+\rho_2}} \right)^{1+\rho_2}. \tag{9.87}$$

Analogous to [32, Lemma 5.4, Corollary 5.4, p. 168], we can prove the following results; some of them have been proved in [10].

Lemma 9.1 *Let* $i = 1, 2$. $E_r^{(i)}(R, W_{Y|UX}, P_{UX})$ *coincides with* $E_{sp}^{(i)}(R, W_{Y|UX}, P_{UX})$ *if* $R \geq R_{cr}^{(i)}(W_{Y|UX}, P_{UX})$ *where*

$$R_{cr}^{(i)}(W_{Y|UX}, P_{UX}) = \left.\frac{\partial E_i(\rho, W_{Y|UX}, P_{UX})}{\partial \rho}\right|_{\rho=1},$$

and is a straight line tangent on $E_{sp}^{(i)}(R, W_{Y|UX}, P_{UX})$ with slope -1 if $R \le R_{cr}^{(i)}(W_{Y|UX}, P_{UX})$, i.e.

$$E_r^{(i)}(R, W_{Y|UX}, P_{UX}) = \begin{cases} E_{sp}^{(i)}(R, W_{Y|UX}, P_{UX}), \\ \qquad \text{if} \quad R \ge R_{cr}^{(i)}(W_{Y|UX}, P_{UX}), \\ E_{sp}^{(i)}\left(R_{cr}^{(i)}(W_{Y|UX}, P_{UX}), W_{Y|UX}, P_{UX}\right) \\ \quad + R_{cr}^{(i)}(W_{Y|UX}, P_{UX}) - R, \\ \qquad \text{if} \quad 0 < R \le R_{cr}^{(i)}(W_{Y|UX}, P_{UX}). \end{cases}$$

Furthermore, $E_r^{(i)}(R, W_{Y|UX}, P_{UX})$ has the parametric form

$$E_r^{(i)}(R, W_{Y|UX}, P_{UX}) = \max_{0 \le \rho \le 1} [E_i(\rho, W_{Y|UX}, P_{UX}) - \rho R]$$

where $E_1(\rho, W_{Y|UX}, P_{UX})$ and $E_2(\rho, W_{Y|UX}, P_{UX})$ are given in (9.86) and (9.87) respectively.

Therefore, we can write the functions E_r in (9.74) and E_{sp} in (9.75) as follows.

$$E_r(R_1, R_2, W_{Y|UX}) = \max_{P_{UX}} \min_{i=1,2} \max_{0 \le \rho \le 1} [E_i(\rho, W_{Y|UX}, P_{UX}) - \rho_i \widehat{R}_i] \tag{9.88}$$

and

$$E_{sp}(R_1, R_2, W_{Y|UX}) = \max_{P_{UX}} \min_{i=1,2} \max_{\rho \ge 0} [E_i(\rho_i, W_{Y|UX}, P_{UX}) - \rho \widehat{R}_i] \tag{9.89}$$

where $\widehat{R}_1 = R_1 + R_2$ and $\widehat{R}_2 = R_2$. Since it is in general hard to find the optimizing solution P_{UX} for E_r and E_{sp} above, we next confine our attention to multiple access channels with some symmetric distributions.

Definition 9.3 [10] We say that the multiple access channel $W_{Y|UX}$ is U-symmetric if for every $u \in \mathcal{U}$ the transition matrix $W_{Y|UX}(\cdot|u, \cdot)$ is symmetric in the sense that the rows (respectively columns) are permutations of each other. An X-symmetric multiple access channel is defined similarly. We then say that $W_{Y|UX}$ is symmetric if it is both U-symmetric and X-symmetric.

It follows that the multiple access channel with additive noise is symmetric (e.g., see the example below), where a multiple access channel $W_{Y|UX}$ with (modulo B) additive noise

$\{P_F : \mathcal{F}\}$ is described as

$$Y_i = U_i \oplus X_i \oplus F_i \pmod B$$

where $Y_i \in \mathcal{Y}$, $X_i \in \mathcal{X}$, $U_i \in \mathcal{U}$ and $F_i \in \mathcal{F}$ are the channel's output, two input and noise symbols at time i such that $\mathcal{Y} = \mathcal{U} = \mathcal{X} = \mathcal{F} = \{0, 1, 2, ..., B-1\}$, and F_i is independent of X_i and U_i, $i = 1, 2, ..., n$.

It is shown in [10] that if the multiple access channel $W_{Y|UX}$ is U-symmetric, then the outer maximum of (9.90) and (9.91) is achieved by a joint distribution of the form $P_{UX}(u,x) = P_U(u)/|\mathcal{X}|$ for every x and u. It then follows that for the symmetric multiple access channel, the maximum of (9.90) and (9.91) is achieved by a uniform joint distribution

$$P^*_{UX}(u,x) = \frac{1}{|\mathcal{U}||\mathcal{X}|},$$

which is independent of ρ. Substituting P^*_{UX} in (9.90) and (9.91) yields

$$E_r(R_1, R_2, W_{Y|UX}) = \min_{i=1,2} \max_{0\le\rho\le1} [\widetilde{E}_i(\rho, W_{Y|UX}) - \rho \widehat{R}_i] \tag{9.90}$$

and

$$E_{sp}(R_1, R_2, W_{Y|UX}) = \min_{i=1,2} \max_{\rho\ge0} [\widetilde{E}_i(\rho, W_{Y|UX}) - \rho \widehat{R}_i] \tag{9.91}$$

where $\widehat{R}_1 = R_1 + R_2$, $\widehat{R}_2 = R_2$,

$$\widetilde{E}_1(\rho, W_{Y|UX}) = (1+\rho)\log_2(|\mathcal{U}||\mathcal{X}|) - \log_2 \sum_{y\in\mathcal{Y}} \left(\sum_{(u,x)\in\mathcal{U}\times\mathcal{X}} W_{Y|UX}(y|u,x)^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

and

$$\widetilde{E}_2(\rho, W_{Y|UX}) = (1+\rho)\log_2|\mathcal{X}| + \log_2|\mathcal{U}| - \log_2 \sum_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} W_{Y|UX}(y|u,x)^{\frac{1}{1+\rho}} \right)^{1+\rho}.$$

We recall the following identities.

$$\min_{P_{SL}:H_{P_{SL}}(S,L)=R} D(P_{SL}\|Q_{SL}) = \max_{\rho\ge0} \left[\rho R - E_{s1}(\rho, Q_{SL})\right], \tag{9.92}$$

$$\min_{P_{SL}:H_{P_{SL}}(L|S)=R} D(P_{SL}\|Q_{SL}) = \max_{\rho\ge0} \left[\rho R - E_{s2}(\rho, Q_{SL})\right], \tag{9.93}$$

where

$$E_{s1}(\rho, Q_{SL}) = (1+\rho)\log_2 \sum_{(s,l)\in\mathcal{S}\times\mathcal{L}} Q_{SL}(s,l)^{\frac{1}{1+\rho}}$$

and

$$E_{s2}(\rho, Q_{SL}) = (1+\rho) \sum_{s \in \mathcal{S}} Q_S(s) \log_2 \sum_{l \in \mathcal{L}} Q_{L|S}(l|s)^{\frac{1}{1+\rho}}.$$

Note that $E_{s1}(\rho, Q_{SL})$ and $E_{s2}(\rho, Q_{SL})$ are both concave in ρ. Clearly, if the marginal distribution $Q_S(s)$ is uniform, then (9.92) and (9.93) are equal. Using (9.90) we now can write (9.72) as

$$\begin{aligned}
&\min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_r(\tau H_P(S), \tau H_{P_{SL}}(L|S), W_{Y|UX})\right] \\
= &\min \Bigg\{ \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + \max_{0 \le \rho_1 \le 1} [\widetilde{E}_1(\rho_1, W_{Y|UX}) - \rho_1 \tau H_{P_{SL}}(S,L)]\right], \\
&\quad \min_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + \max_{0 \le \rho_2 \le 1} [\widetilde{E}_2(\rho_2, W_{Y|UX}) - \rho_2 \tau H_{P_{SL}}(L|S)]\right] \Bigg\} \\
= &\min \Bigg\{ \min_{R} \left[\min_{P_{SL}: \tau H_{P_{SL}}(S,L) = R} \tau D(P_{SL} \parallel Q_{SL}) + \max_{0 \le \rho_1 \le 1} [\widetilde{E}_1(\rho_1, W_{Y|UX}) - \rho_1 R] \right], \\
&\quad \min_{R} \left[\min_{P_{SL}: \tau H_{P_{SL}}(L|S) = R} \tau D(P_{SL} \parallel Q_{SL}) + \max_{0 \le \rho_2 \le 1} [\widetilde{E}_2(\rho_2, W_{Y|UX}) - \rho_2 R] \right] \Bigg\} \qquad (9.94)
\end{aligned}$$

and similarly using (9.91) we can write (9.73) as

$$\begin{aligned}
&\inf_{P_{SL}} \left[\tau D(P_{SL} \parallel Q_{SL}) + E_{sp}(\tau H_P(S), \tau H_{P_{SL}}(L|S), W_{Y|UX})\right] \\
= &\min \Bigg\{ \inf_{R} \left[\min_{P_{SL}: \tau H_{P_{SL}}(S,L) = R} \tau D(P_{SL} \parallel Q_{SL}) + \max_{\rho_1 \ge 0} [\widetilde{E}_1(\rho_1, W_{Y|UX}) - \rho_1 R] \right], \\
&\quad \inf_{R} \left[\min_{P_{SL}: \tau H_{P_{SL}}(L|S) = R} \tau D(P_{SL} \parallel Q_{SL}) + \max_{\rho_2 \ge 0} [\widetilde{E}_2(\rho_2, W_{Y|UX}) - \rho_2 R] \right] \Bigg\}. \qquad (9.95)
\end{aligned}$$

Consequently, applying Fenchel duality theorem as in the precious chapters and (9.92) and (9.93), we obtain the following.

Theorem 9.4 *Given Q_{SL}, $W_{Y|UX}$, and the transmission rate τ, the lower bound of the JSCC error exponent given in (9.72) and the upper bound given in (9.76) can be equivalently expressed as*

$$\begin{aligned}
\min_{i=1,2} \max_{0 \le \rho \le 1} [\widetilde{E}_i(\rho, W_{Y|UX}) - \tau E_{si}(\rho, Q_{SL})] &\le E_J(Q_{SL}, W_{Y|UX}, \tau) \\
&\le \min_{i=1,2} \max_{\rho \ge 0} [\widetilde{E}_i(\rho, W_{Y|UX}) - \tau E_{si}(\rho, Q_{SL})].
\end{aligned} \qquad (9.96)$$

Example 9.1 (Binary CS-AMAC System) Now consider binary CS Q_{SL} with distribution

$$Q_{SL}(S=0,L=0)=\frac{2(1-q)}{3},\qquad Q_{SL}(S=1,L=0)=\frac{q}{2},$$
$$Q_{SL}(S=0,L=1)=\frac{q}{2},\qquad Q_{SL}(S=1,L=1)=\frac{1-q}{3},$$

where $0<q<1/2$. Then

$$E_{s1}(\rho,Q_{SL})=(1+\rho)\log_2\left\{\left[\left(\frac{2}{3}\right)^{\frac{1}{1+\rho}}+\left(\frac{1}{3}\right)^{\frac{1}{1+\rho}}\right](1-q)^{\frac{1}{1+\rho}}+2\left(\frac{q}{2}\right)^{\frac{1}{1+\rho}}\right\},$$

$$\begin{aligned} E_{s2}(\rho,Q_{SL}) &= (1+\rho)\left(\frac{2(1-q)}{3}+\frac{q}{2}\right)\log_2\left[\left(\frac{\frac{2(1-q)}{3}}{\frac{2(1-q)}{3}+\frac{q}{2}}\right)^{\frac{1}{1+\rho}}+\left(\frac{\frac{q}{2}}{\frac{2(1-q)}{3}+\frac{q}{2}}\right)^{\frac{1}{1+\rho}}\right] \\ &\quad +(1+\rho)\left(\frac{1-q}{3}+\frac{q}{2}\right)\log_2\left[\left(\frac{\frac{1-q}{3}}{\frac{1-q}{3}+\frac{q}{2}}\right)^{\frac{1}{1+\rho}}+\left(\frac{\frac{q}{2}}{\frac{1-q}{3}+\frac{q}{2}}\right)^{\frac{1}{1+\rho}}\right]. \end{aligned}$$

Consider a binary multiple access channel $W_{Y|UX}$ with binary additive noise $P_F(F=1)=\epsilon$ $(0<\epsilon<1/2)$. That is, the transition probabilities are given by

$$\begin{aligned} P_{Y|UX}(Y=0|U=0,X=0)&=1-\epsilon, & P_{Y|UX}(Y=1|U=0,X=0)&=\epsilon \\ P_{Y|UX}(Y=0|U=0,X=1)&=\epsilon, & P_{Y|UX}(Y=1|U=0,X=1)&=1-\epsilon \\ P_{Y|UX}(Y=0|U=1,X=0)&=\epsilon, & P_{Y|UX}(Y=1|U=1,X=0)&=1-\epsilon \\ P_{Y|UX}(Y=0|U=1,X=1)&=1-\epsilon, & P_{Y|UX}(Y=1|U=1,X=1)&=\epsilon. \end{aligned}$$

It follows that

$$\widetilde{E}_1(\rho,W_{Y|UX})=\widetilde{E}_2(\rho,W_{Y|UX})=\rho\log_2 2-(1+\rho)\log_2\left(\epsilon^{\frac{1}{1+\rho}}+(1-\epsilon)^{\frac{1}{1+\rho}}\right).$$

In Fig. 9.5, we plot the lower and upper bounds for the JSCC error exponent E_J for different (q,ϵ) pairs with transmission rate $t=0.25$ and 0.35. As illustrated, the upper and lower bounds coincide (this can also be proved by checking that the two outer minimums in (9.96) are achieved by the same i and that the inner maximum in the upper bound is achieved by $\rho\le 1$) for many (q,ϵ) pairs (e.g., when $\tau=0.25, q=0.1, \epsilon\ge 0.0205$ and when $\tau=0.35, q=0.1$, $\epsilon\ge 0.0056$), and hence exactly determine the exponent.

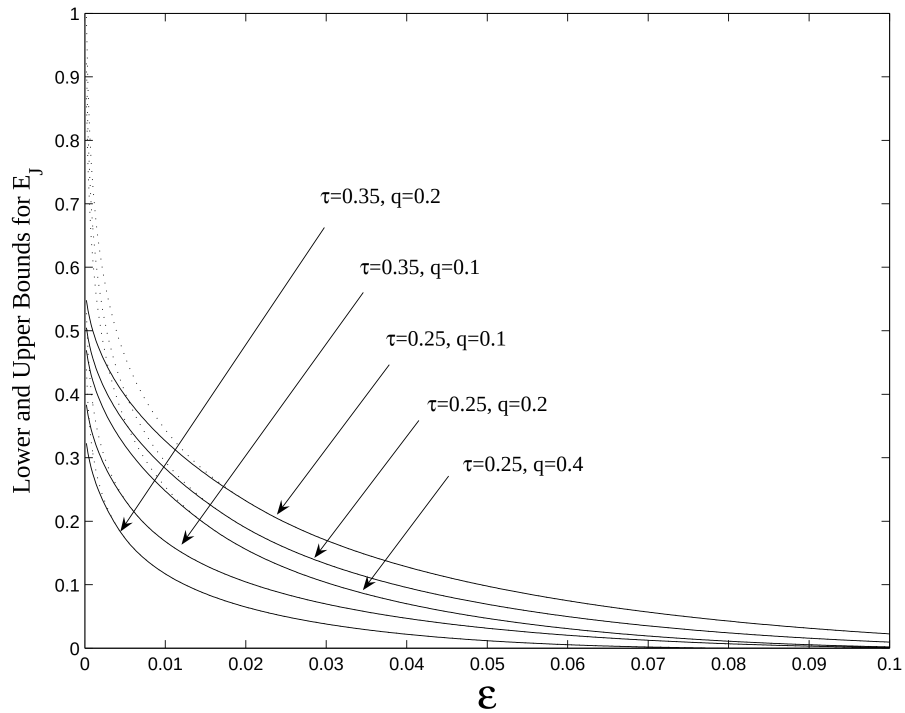

Figure 9.5: The lower bound (solid line) and the upper bound (dash line) for the system JSCC error exponent for transmitting binary CS over the binary AMAC with binary additive noise in Example 9.1.

9.9 Conclusion

In this chapter, we studied the exponential behavior of the probabilities of error for transmitting correlated sources over asymmetric 2-user channels by JSCC. We first established universally achievable error exponent pairs for the two receivers by using the joint type-packing lemma and generalized mutual information decoders. We also defined the system JSCC error exponent and derived lower and upper bounds for the exponent. By definition, when the system JSCC error exponent is positive, the sources can be reliably transmitted over the channel. Thus by examining the positivity of the lower bound for the exponent we obtained the forward part for the JSCC theorem. We also proved the converse part by Fano's inequality and hence we established the JSCC theorem for the asymmetric multi-terminal scenario. It is demonstrated that the condition can actually be achieved by a tandem coding scheme, which combines separate source and channel coding. This means that tandem coding does not lose optimality from the point of view of reliable transmissibility.

We next specialized our results to CS-AMAC systems and CS-ABC systems. We analytically computed the system JSCC error exponent for the CS-AMAC case when the channel admits a symmetric distribution. Numerical examples showed that our bounds are tight, hence exactly determining the exponent, for a large range of source-channel parameters.

Chapter 10

When is JSCC Worthwhile: JSCC vs Tandem Coding Reliability Functions

So far we have obtained computable lower and upper bounds for the JSCC reliability function (error exponent/excess distortion exponent) for different single-user communication systems and the asymmetric 2-user system. In this chapter, we employ these results to provide a systematic comparison of the JSCC reliability function and the corresponding tandem coding reliability function with the same transmission rate. As can be shown, JSCC reliability function is at least as large as the tandem coding reliability function; however, we are particularly interested in investigating the situation where a strict inequality holds. Indeed, this inequality, when it holds, provides a theoretical underpinning and justification for JSCC design as opposed to the widely used tandem approach, since the former method will yield a faster exponential rate of decay for the error probability, which may translate into substantial reductions in complexity and delay for real-world communication systems.

In Section 10.1, we derive a formula for the (lossless) tandem coding error exponent E_T for discrete systems (consisting of an arbitrary discrete source and an arbitrary discrete channel). The exponent is conceptually represented in terms of the discrete source and channel error exponents and is in general not computable, as the channel error exponent is not fully known even for DMCs. However, for discrete memoryless systems and SEM systems, we can obtain upper and lower bounds for the tandem error exponent by replacing the channel error exponent by its upper and lower bounds.

We then address the comparison of the JSCC error exponent E_J with the tandem coding error exponent E_T for discrete memoryless systems in Section 10.2. We first show that E_J can at most double E_T, and then we establish sufficient (computable) conditions for which $E_J > E_T$ for any given source-channel pair $(Q_S, W_{Y|X})$, which are satisfied for a large class of memoryless source-channel pairs. As an application, we estimate the power savings of JSCC over tandem coding for transmitting binary DMS over binary-input quantized-output additive white Gaussian noise and memoryless Rayleigh-fading channels; it turns out that the advantage of JSCC in terms of the reliability function translates into more than 2 dB power gain for those systems. The comparison of E_J and E_T for SEM systems (which consist of an SEM source and an SEM channel) is provided in Section 10.3. As in the preceding section, we prove that $E_J \leq 2E_T$ and establish sufficient (computable) conditions for which $E_J > E_T$. We observe via numerical examples that such conditions are satisfied by a wide class of SEM source-channel pairs.

It is seen in Section 10.4 that our formula for the tandem error exponent remains valid for a discrete memoryless system involving channel output feedback and source SI. For systems with feedback, we show that $E_{T,fb} \leq E_{J,fb} \leq 2E_{T,fb}$; for systems with source SI at the decoder, we also prove that $E_T^{SID} \leq E_J^{SID}$. Additionally, we provide numerical examples to show that the JSCC error exponent is superior to the corresponding tandem coding error exponent in most cases.

In Section 10.5, we compare the JSCC excess distortion exponent with the tandem coding excess distortion exponent for Gaussian systems. For an MGS and an MGC, the tandem coding excess exponent results from separately performing and concatenating optimal lossy Gaussian source coding and channel coding for MGC. We derive a formula for the tandem coding excess distortion exponent for the case when SDR ≥ 4 ($\approx$ 6dB) (i.e., when the distortion threshold is less than 1/4 of the source variance). The exponent admits a similar form as the discrete tandem error exponent in terms of the MGS exponent and the MGC exponent. We next numerically compare the lower bound of the JSCC exponent with the upper bound of the tandem exponent and observe that the JSCC exponent can be strictly superior to the tandem exponent for many SNR-SDR pairs.

We next address the tandem coding error exponent for asymmetric 2-user systems in Section 10.6. As for the point-to-point systems, we derive a conceptual formula for the tandem coding error exponent in terms of the corresponding 2-user source error exponent and the asymmetric 2-user channel error exponent. By numerically comparing the lower bound of the JSCC error exponent and the upper bound of the tandem coding error exponent, we illustrate that, JSCC can considerably outperform tandem coding in terms of error exponent for a large class of binary CS-AMAC systems with additive noise. Finally a conclusion is given in Section 10.7.

10.1 Tandem Error Exponent for Discrete Systems

Our aim in this section is to derive a formula for the tandem coding error exponent for a general discrete system (with memory) consisting of a discrete source $\mathbf{Q_S} = \{Q_{S^{\tau n}} \in \mathcal{P}(\mathcal{S}^{\tau n})\}_{\tau n=1}^{\infty}$ and a discrete channel $\mathbf{W_{Y|X}} = \{W_{Y^n|X^n} \in \mathcal{P}(\mathcal{Y}^n|\mathcal{X}^n)\}_{n=1}^{\infty}$.

Conceptually and essentially, a tandem code $(f_n^*, \varphi_n^*) \triangleq (f_{sn}, f_{cn}, \varphi_{cn}, \varphi_{sn})$ is composed of two "separately" designed codes: a $(\tau n, M_n)$ block source code (f_{sn}, φ_{sn}) with source code rate[1]

$$R_{s,n} \triangleq \frac{\log_2 M_n}{\tau n} \quad \text{source code bits/source symbol,}$$

and an (n, M_n) block channel code (f_{cn}, φ_{cn}) with channel code rate

$$R_{c,n} \triangleq \frac{\log_2 M_n}{n} \quad \text{source code bits/channel use,}$$

assuming that the limit $\lim_{n\to\infty} \log M_n/n$ exists, i.e.,

$$\limsup_{n\to\infty} \frac{\log M_n}{n} = \liminf_{n\to\infty} \frac{\log M_n}{n}.$$

[1]Since in the tandem system $k = \tau n$ and τ is a constant, to simplify our notation, we denote the source rate by $R_{s,n}$ instead of $R_{s,k}$ or $R_{s,\tau n}$.

Here "separately" means that the source code is designed without the knowledge of the channel statistics, and the channel code is designed without the knowledge of the source statistics. However, as long as the source encoder is directly concatenated by a channel encoder, the source statistics would be automatically brought into the channel coding stage. Thus common randomization is also needed to decouple source and channel coding (e.g., [51]). Specifically, we assume that the source coding index $i = f_{sn}(\mathbf{s})$ is mapped to a channel index through a permutation mapping $\pi_m : \{1, 2, ..., M_n\} \to \{1, 2, ..., M_n\}$, commonly called an index assignment (π_m is assumed to be known at both the transmitter and the receiver; see Fig. 10.1). Furthermore, the choice of π_m is assumed random and equally likely from all the $M_n!$ different possible index assignments, so that the indices fed into the channel encoder have a uniform distribution, i.e.,

$$\begin{aligned}\Pr(\pi_m[f_{sn}(S^{\tau n})] = l) &= \sum_{i=1}^{M_n} \Pr(f_{sn}(S^{\tau n}) = i)\Pr(\pi_m(i) = l | f_{sn}(S^{\tau n}) = i) \\ &= \sum_{i=1}^{M_n} \Pr(f_{sn}(S^{\tau n}) = i)\frac{(M_n-1)!}{M_n!} \\ &= \frac{1}{M_n} \end{aligned} \tag{10.1}$$

for any $l \in \{1, 2, ..., M_n\}$. Hence common randomization achieves statistical independence between the source and channel coding operations.

In what follows we need to make another assumption regarding the source code in order to analyze the probability of error. Let the source codebook be $\mathcal{C} = \{\mathbf{c}_1, ..., \mathbf{c}_{M_n}\} \subseteq \mathcal{S}^{\tau n}$. We assume that (A1) the source code satisfies the condition (for every n): $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) > 0$ and $\mathbf{c}_i \in f_{sn}^{-1}(i)$ for every $i = 1, 2, ..., M_n$, where $f_{sn}^{-1}(i) \triangleq \{\mathbf{s} \in \mathcal{S}^{\tau n} : f_{sn}(\mathbf{s}) = i\}$; see Fig. 10.2. Clearly, the assumption has practical meaning. If $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) = 0$ for some i, then the codeword $\mathbf{c}_i$ is redundant, and we can remove it from the codebook $\mathcal{C}$. If $\mathbf{c}_i \notin f_{sn}^{-1}(i)$, we can map the index i to some source message $\widehat{\mathbf{s}}$ such that $Q_{S^{\tau n}}(\widehat{\mathbf{s}}) > 0$ and $f_{sn}(\widehat{\mathbf{s}}) = i$, so that the source coding probability of error

$$P_{es}^{(\tau n)}(\mathbf{Q_S}, R_{s,n}) = \sum_{\mathbf{s}:\varphi_{sn}(f_{sn}(\mathbf{s}))\neq \mathbf{s}} Q_{S^{\tau n}}(\mathbf{s}) \tag{10.2}$$

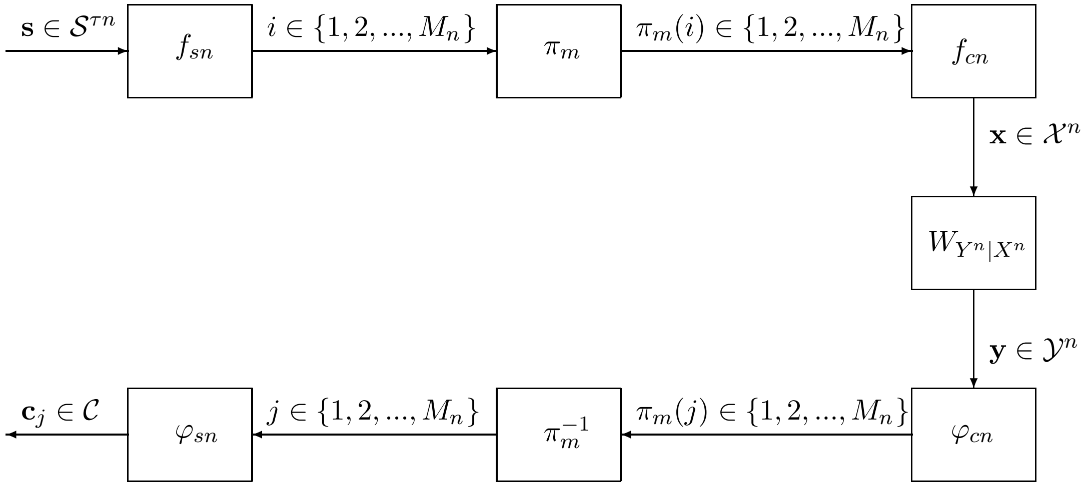

Figure 10.1: Tandem coding system for discrete sources and discrete channels.

is strictly reduced by setting $\widehat{\mathbf{s}}$ as the codeword $\mathbf{c}_i$. We remark that the source code satisfying (A1) does not lose optimality (in the sense of achieving the source error exponent).

By introducing the uniform index assignment assumption and (A1), the error probability of the tandem code (f_n^*, φ_n^*) is given by

$$\begin{aligned}
&P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)\\
&\triangleq \Pr\left(\varphi_{sn}\left[\pi_m^{-1}(\varphi_{cn}(Y^n))\right] \neq S^{\tau n}\right)\\
&= \sum_{l=1}^{M_n} \underbrace{\Pr(\pi_m[f_{sn}(S^{\tau n})] = l)}_{=1/M_n} \big[\Pr\left(\varphi_{cn}(Y^n) \neq l \middle| \pi_m[f_{sn}(S^{\tau n})] = l\right) + \\
&\qquad \Pr\left(\{\varphi_{cn}(Y^n) = l\} \cap \{\varphi_{sn}[\pi_m^{-1}(l))] \neq S^{\tau n}\} \middle| \pi_m[f_{sn}(S^{\tau n})] = l\right)\big] \quad (10.3)\\
&= \sum_{l=1}^{M_n} \frac{1}{M_n} \Pr\left(\varphi_{cn}(Y^n) \neq l | l \text{ is sent}\right)\\
&\qquad + \Pr\left(\varphi_{sn}(f_{sn}(S^{\tau n})) \neq S^{\tau n}\right) \sum_{l=1}^{M_n} \frac{1}{M_n} \Pr\left(\varphi_{cn}(Y^n) = l | l \text{ is sent}\right) \quad (10.4)\\
&= P_{ec}^{(n)}(\mathbf{W_{Y|X}}, R_{c,n}) + (1 - P_{ec}^{(n)}(\mathbf{W_{Y|X}}, R_{c,n}))P_{es}^{(\tau n)}(\mathbf{Q_S}, R_{s,n}), \quad (10.5)
\end{aligned}$$

where (10.3) follows from the assumption (A1), which implies that a channel decoding error must cause an overall system decoding error, (10.4) holds due to the independence of source

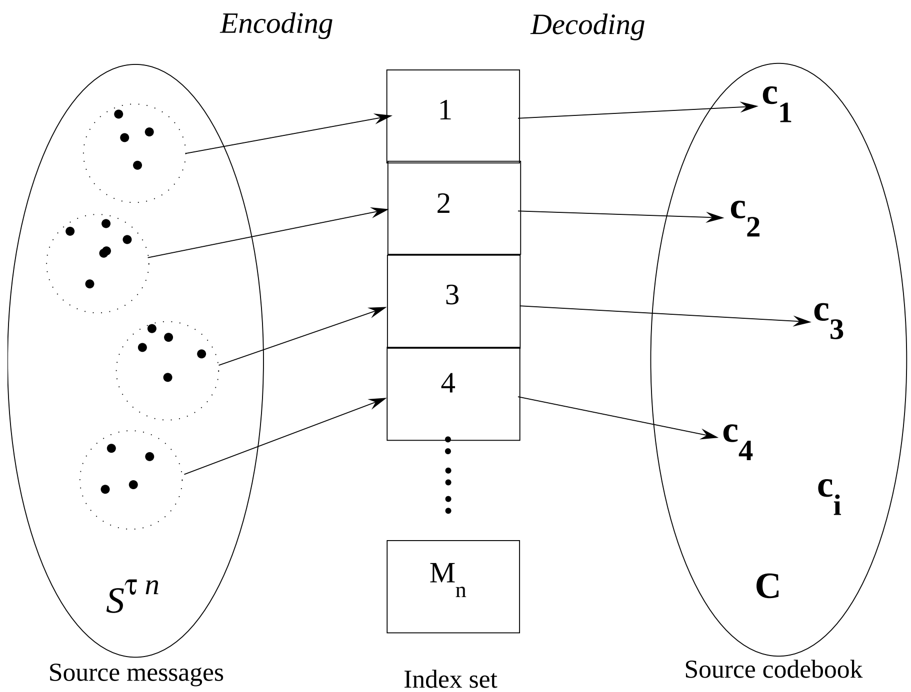

Figure 10.2: (A1): $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) > 0$ and $\mathbf{c}_i \in f_{sn}^{-1}(i)$ for all $i = 1, 2, ..., M_n$.

and channel coding, and in (10.5)

$$P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n}) = \frac{1}{M_n}\sum_{l=1}^{M_n}\sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y})\neq l} W_{Y^n|X^n}(\mathbf{y}|f_{cn}(l)) \tag{10.6}$$

is the probability of error for channel coding, and $P_{es}^{(\tau n)}(\mathbf{Q_S}, R_{s,n})$ is the probability of error for source coding given by (10.2).

Definition 10.1 The tandem coding error exponent $E_T(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau)$ for source $\mathbf{Q_S}$ and channel $\mathbf{W}_{\mathbf{Y}|\mathbf{X}}$ is defined as the supremum of the set of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes (f_n^*, φ_n^*) satisfying (A1) with transmission rate τ such that

$$\widehat{E} \leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau).$$

When there is no possibility of confusion, $E_T(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau)$ will often be written as E_T.

Proposition 10.1 $E_J(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \geq E_T(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau)$.

Proof: For any sequence of tandem codes $\{(f_n^*, \varphi_n^*)\}_{n=1}^\infty$ composed of a sequence of source codes $\{(f_{sn}, \varphi_{sn})\}_{n=1}^\infty$ and a sequence of channel codes $\{(f_{cn}, \varphi_{cn})\}_{n=1}^\infty$, we have

$$\begin{aligned} P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) &= \sum_{m=1}^{M_n!}\frac{1}{M_n!}\Pr\left(\varphi_{sn}\left[\pi_m^{-1}(\varphi_{cn}(Y^n))\right] \neq S^{\tau n}|\pi_m \text{ is fixed}\right) \\ &\geq \min_{1\leq m\leq M_n!}\Pr\left(\varphi_{sn}\left[\pi_m^{-1}(\varphi_{cn}(Y^n))\right] \neq S^{\tau n}|\pi_m \text{ is fixed}\right). \end{aligned}$$

Let the above minimum be achieved by $m^* = m^*(n)$, then there exists a sequence of JSC codes $\{(f_n, \varphi_n)\}_{n=1}^\infty$ with $f_n = f_{cn}\circ\pi_{m^*}\circ f_{sn}$ and $\varphi_n = \varphi_{sn}\circ\pi_{m^*}^{-1}\circ\varphi_{cn}$ such that

$$P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \geq P_e^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \quad \text{for any } n \geq 1,$$

where "$\circ$" denotes composition and $P_e^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau)$ is the probability of error induced by the JSC codes $\{(f_n, \varphi_n)\}$. Since this holds for any sequence of tandem codes (satisfying (A1)), it then follows from the definition of joint and tandem exponents that $E_J \geq E_T$. ■

Theorem 10.1 *For a discrete source-channel pair* $(\mathbf{Q_S}, \mathbf{W_{Y|X}})$ *and transmission rate* τ*,*

$$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = \sup_{R>0} \min \left\{ \tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E(R, \mathbf{W_{Y|X}}) \right\}$$

where $e(R, \mathbf{Q_S})$ *is the source error exponent defined in Def. 7.2 and* $E(R, \mathbf{W_{Y|X}})$ *is the channel error exponent defined in Def. 7.3.*

Remark 10.1 Note that the formula for the tandem error exponent has been simply mentioned by Csiszár [30] without stating any assumptions (i.e., common randomization, Assumption (A1)) for the tandem system.

Proof:

Forward Part: We show that there exists a sequence of tandem codes

$$(f_n^*, \varphi_n^*) = (f_{sn}, f_{cn}, \varphi_{cn}, \varphi_{sn})$$

satisfying (A1) with rate τ such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > \sup_{R>0} \min \left\{ \tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E(R, \mathbf{W_{Y|X}}) \right\} - \delta$$

for any $\delta > 0$. It follows from (10.5) that

$$P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \le 2\max\{P_{es}^{(\tau n)}(\mathbf{Q_S}, R_{s,n}), P_{ec}^{(n)}(\mathbf{W_{Y|X}}, R_{c,n})\},$$

or equivalently,

$$\begin{aligned}\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \quad \ge \quad & \min \left\{ \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{es}^{(\tau n)}(\mathbf{Q_S}, R_{s,n}), \right. \\ & \left. \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{ec}^{(n)}(\mathbf{W_{Y|X}}, R_{c,n}) \right\}.\end{aligned}$$

Now fix $R > 0$ and $\delta > 0$. According to the definition of the source error exponent, there exists a sequence of $(\tau n, \widetilde{M}_n)$ source codes $(\widetilde{f}_{sn}, \widetilde{\varphi}_{sn})$ satisfying (A1) (since (A1) does not lose optimality) such that

$$\liminf_{n\to\infty} -\frac{1}{\tau n}\log_2 \Pr\left(\widetilde{\varphi}_{sn}\left[\widetilde{f}_{sn}(S^{\tau n})\right] \ne S^{\tau n}\right) > e(R, \mathbf{Q_S}) - \delta \quad \text{and} \quad \limsup_{n\to\infty} \frac{\log_2 \widetilde{M}_n}{\tau n} \le R.$$

Since a source code with a codebook size larger than $\widetilde{M}_n$ would have a smaller probability of error, there must exist a sequence of $(\tau n, \lceil 2^{\tau n R} \rceil)$ source codes (f_{sn}, φ_{sn}) (satisfying (A1)) such that

$$\liminf_{n\to\infty} -\frac{1}{\tau n} \Pr\left(\varphi_{sn}\left[f_{sn}(S^{\tau n})\right] \neq S^{\tau n}\right) > e(R, \mathbf{Q_S}) - \delta.$$

Similarly, for given τR, the definition of channel error exponent asserts that there exists a sequence of $(n, \widehat{M}_n)$ channel codes $(\widehat{f}_{cn}, \widehat{\varphi}_{cn})$ such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log_2 \Pr\left(\widehat{\varphi}_{cn}(Y^n) \neq L\right) > E(\tau R, \mathbf{W_{Y|X}}) - \delta \quad \text{and} \quad \liminf_{n\to\infty} \frac{\log_2 \widehat{M}_n}{n} \geq \tau R.$$

Since a channel code with a codebook size smaller than $\widehat{M}_n$ would have a smaller probability of error, there must exist a sequence of $(n, \lceil 2^{\tau n R} \rceil)$ channel codes (f_{cn}, φ_{cn}) such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log_2 \Pr\left(\varphi_{cn}(Y^n) \neq L\right) > E(\tau R, \mathbf{W_{Y|X}}) - \delta.$$

Therefore, there exists a sequence of tandem codes, composed by a sequence of $(\tau n, \lceil 2^{\tau n R} \rceil)$ source codes, and a sequence of $(n, \lceil 2^{\tau n R} \rceil)$ channel codes (with the same $M_n = \lceil 2^{\tau n R} \rceil$), such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > \min\left\{\tau e(R, \mathbf{Q_S}), E(\tau R, \mathbf{W_{Y|X}})\right\} - \delta.$$

Finally, since R and δ are arbitrary, we can take the supremum over $R > 0$, completing the proof of the forward part.

Converse Part: We show that for any sequence of tandem codes (f_n^*, φ_n^*) with rate τ composed by source codes (f_{sn}, φ_{sn}) and channel codes (f_{cn}, φ_{cn}) satisfying assumption (A1),

$$\liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq \sup_{R>0} \min\left\{\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E(R, \mathbf{W_{Y|X}})\right\}. \qquad (10.7)$$

Consider the tandem code sequence (f_n^*, φ_n^*) of rate τ composed by a $(\tau n, M_n)$ block source code (f_{sn}, φ_{sn}) (with rate $R_{s,n}$) and an (n, M_n) block channel code (f_{cn}, φ_{cn}) (with rate $R_{c,n} = \tau R_{s,n}$).

We first assume that $\limsup_{n\to\infty} -\frac{1}{n}\log_2(1-P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n})) \geq \delta$ for some positive δ independent of n, which implies that there exists a sequence $n_0 \leq n_1 \leq n_2 \leq \cdots \leq \infty$ such that

$$\lim_{i\to\infty} P_{ec}^{(n_i)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n_i}) \geq 1 - \lim_{i\to\infty} 2^{-n_i\delta} = 1$$

In this trivial case,

$$\begin{aligned}
\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q}_\mathbf{S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) &\leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n}) \\
&\leq \lim_{i\to\infty} -\frac{1}{n_i}\log_2 P_{ec}^{(n_i)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n_i}) \\
&= 0
\end{aligned}$$

and (10.7) holds. Next we assume that $\limsup_{n\to\infty} -\frac{1}{n}\log_2(1-P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n})) = 0$. It then follows from (10.5) that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q}_\mathbf{S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \\
&\leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 \left[(1-P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n}))P_{es}^{(\tau n)}(\mathbf{Q}_\mathbf{S}, R_{s,n})\right] \\
&\leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{es}^{(\tau n)}(\mathbf{Q}_\mathbf{S}, R_{s,n})
\end{aligned} \tag{10.8}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q}_\mathbf{S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{ec}^{(n)}(\mathbf{W}_{\mathbf{Y}|\mathbf{X}}, R_{c,n}). \tag{10.9}$$

Let

$$R = \lim_{n\to\infty} R_{s,n} = \lim_{n\to\infty} \frac{\log_2 M_n}{\tau n}.$$

By definition, the source error exponent $e(R, \mathbf{Q}_\mathbf{S})$ is the largest (supremum) number e such that there exists a sequence of $(\tau n, \widetilde{M}_n)$ source codes $(\widetilde{f}_{sn}, \widetilde{\varphi}_{sn})$ satisfying

$$\liminf_{n\to\infty} -\frac{1}{\tau n}\log_2 \Pr\left(\widetilde{\varphi}_{sn}\left[\widetilde{f}_{sn}(S^{\tau n})\right] \neq S^{\tau n}\right) \geq e \quad \text{and} \quad \limsup_{n\to\infty} \frac{\log_2 \widetilde{M}_n}{\tau n} \leq R.$$

This means that

$$\liminf_{n\to\infty} -\frac{1}{\tau n}\log_2 \Pr\left(\widetilde{\varphi}_{sn}\left[\widetilde{f}_{sn}(S^{\tau n})\right] \neq S^{\tau n}\right) \leq e(R, \mathbf{Q}_\mathbf{S})$$

holds for all the codes $(\widetilde{f}_{sn}, \widetilde{\varphi}_{sn})$ with $\limsup_{n\to\infty} \frac{\log_2 \widetilde{M}_n}{\tau n} \leq R$, and hence holds for the sequence of $(\tau n, M_n)$ block codes satisfying (A1).

Similarly, note that

$$\tau R = \lim_{n\to\infty} R_{c,n} = \lim_{n\to\infty} \frac{\log_2 M_n}{n}.$$

By the definition of the channel error exponent,

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 \Pr\left(\varphi_{cn}(Y^n) \neq L\right) \leq E(\tau R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})$$

holds for all the $(n, \widehat{M}_n)$ block channel codes $(\widehat{f}_{cn}, \widehat{\varphi}_{cn})$ with $\liminf_{n\to\infty} \frac{\log_2 \widehat{M}_n}{n} \geq \tau R$, and of course holds for the sequence of (n, M_n) block codes.

Putting things together, (10.8) and (10.9) yield

$$\liminf_{n\to\infty} -\frac{1}{n}\log_2 P_{e^*}^{(n)}(\mathbf{Q_S}, \mathbf{W}_{\mathbf{Y}|\mathbf{X}}, \tau) \leq \min\{\tau e(R, \mathbf{Q_S}), E(\tau R, \mathbf{W}_{\mathbf{Y}|\mathbf{X}})\},$$

holds for all the $(\tau n, M_n)$ block source codes satisfying (A1) and all the (n, M_n) block channel codes with $\lim_{n\to\infty} \frac{\log_2 M_n}{\tau n} = R$. Since the above is satisfied for any sequence of $M_n > 0$ and hence for all $R > 0$, we take the supremum of $R > 0$ and obtain (10.7). ■

10.2 Discrete Memoryless Systems

In this section, we assume that the source is a DMS, i.e., $\mathbf{Q_S} = Q_S^{(\tau n)}$, and that the channel is a DMC, i.e., $\mathbf{W}_{\mathbf{Y}|\mathbf{X}} = W_{Y|X}^{(n)}$. We hence have the following corollary.

Corollary 10.1 *Let $\tau H_{Q_S}(S) < C(W_{Y|X})$ and let $\tau \log_2 |\mathcal{S}| > R_\infty(W_{Y|X})$. Then*

$$\begin{aligned} \underline{E}_{Tr}(Q_S, W_{Y|X}, \tau) &\leq E_T(Q_S, W_{Y|X}, \tau) \\ &= \sup_{\tau H_{Q_S}(S) \leq R \leq C(W_{Y|X})} \min\left\{\tau e\left(\frac{R}{\tau}, Q_S\right), E(R, W_{Y|X})\right\} \quad (10.10) \\ &\leq \overline{E}_{Tsp}(Q_S, W_{Y|X}, \tau), \quad (10.11) \end{aligned}$$

where

$$\underline{E}_{Tr}(Q_S, W_{Y|X}, \tau) \triangleq \sup_{\tau H_{Q_S}(S) \leq R \leq C(W_{Y|X})} \min\left\{\tau e\left(\frac{R}{\tau}, Q_S\right), E_r(R, W_{Y|X})\right\} \quad (10.12)$$

and

$$\overline{E}_{Tsp}(Q_S, W_{Y|X}, \tau) \triangleq \sup_{\tau H_{Q_S}(S) \leq R \leq C(W_{Y|X})} \min\left\{\tau e\left(\frac{R}{\tau}, Q_S\right), E_{sp}(R, W_{Y|X})\right\} \quad (10.13)$$

Remark 10.2 If $\tau H_{Q_S}(S) \geq C(W_{Y|X})$, $E_T(Q_S, W_{Y|X}, \tau) = 0$.

Note that

$$\sup_{R \leq \tau \log_2 |\mathcal{S}|} \tau e\left(\frac{R}{\tau}, Q_S\right) = \tau e(\log_2 |\mathcal{S}|, Q_S) = -\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}),$$

where $\overline{Q_S(s)}$ is the geometric mean of the source probabilities, i.e.

$$\overline{Q_S(s)} \triangleq \left(\prod_{s\in\mathcal{S}} Q_S(s)\right)^{1/|\mathcal{S}|} \leq 1/|\mathcal{S}|.$$

If $-\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) \geq E(\tau \log_2 |\mathcal{S}|, W_{Y|X})$, then the graphs of $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$ must have exactly one intersection R_o and by (10.10)

$$E_T(Q_S, W, \tau) = \tau e\left(\frac{R_o}{\tau}, Q_S\right) = E(R_o, W_{Y|X}), \tag{10.14}$$

since $\tau e(R/\tau, Q_S)$ is strictly increasing in $R \in [\tau H_{Q_S}(S), \tau \log_2 |\mathcal{S}|]$ and $E(R, W_{Y|X})$ is non-increasing in R. If $-\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) < E(\tau \log_2 |\mathcal{S}|, W_{Y|X})$, then there is no intersection between $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$. Recall (2.5) that $\tau e(R/\tau, Q_S)$ is infinite in the open interval $(\tau \log_2 |\mathcal{S}|, \infty)$. In this case, we have that

$$E_T(Q_S, W, \tau) = E(\tau \log_2 |\mathcal{S}|, W_{Y|X}) \tag{10.15}$$

by (10.10). Without loss of generality, we denote

$$R_o \triangleq \begin{cases} \text{the rate satisfying} \quad \tau e(\frac{R_o}{\tau}, Q_S) = E(R_o, W_{Y|X}) \\ \qquad\qquad \text{if } -\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) \geq E(\tau \log_2 |\mathcal{S}|, W_{Y|X}), \\ \tau \log_2 |\mathcal{S}| \\ \qquad\qquad \text{if } -\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) < E(\tau \log_2 |\mathcal{S}|, W_{Y|X}), \end{cases} \tag{10.16}$$

so that we can always write that $E_T(Q_S, W, \tau) = E(R_o, W_{Y|X})$.

10.2.1 E_J Can At Most Double E_T

When the DMS is uniform, the optimal source coding operation reduces to the trivial enumerating (identity) function with $M = |S|^{\tau n}$ as the source is incompressible. Hence only channel coding is performed in both JSCC and tandem coding and $E_J(Q_S, W_{Y|X}, \tau) =$

$E_T(Q_S, W_{Y|X}, \tau) = E(\tau \log_2 |\mathcal{S}|, W_{Y|X})$. Thus, our comparison of the two exponents is nontrivial only if the source is nonuniform and $\tau H_{Q_S}(S) < C(W_{Y|X})$. Even though we know that E_J is never worse than E_T, the following theorem gives a limit on how much E_J can outperform E_T.

Theorem 10.2 *JSCC exponent can at most be equal to double the tandem coding exponent, i.e.,*

$$E_J(Q_S, W_{Y|X}, \tau) \leq 2E_T(Q_S, W_{Y|X}, \tau),$$

with equality if $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ *and* $T_{sp}(\overline{\rho}^*, W) = \tau E_s(\overline{\rho}^*, Q_S) + 2\tau D(Q_S^{(\overline{\rho}^*)} \parallel Q_S)$.

Remark 10.3 Equivalently, this upper bound also implies that E_J can at most exceed E_T by $E_J/2$, i.e.,

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) \leq \frac{1}{2} E_J(Q_S, W_{Y|X}, \tau). \tag{10.17}$$

Proof of Theorem 10.2: We first refer to the upper bound of $E_J(Q_S, W_{Y|X}, \tau)$ given by (5.4)

$$E_J(Q_S, W_{Y|X}, \tau) \leq \min_{\tau H_{Q_S}(S) \leq R \leq \tau \log_2 |\mathcal{S}|} \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X}) \right], \tag{10.18}$$

where $\tau e(R/\tau, Q_S)$ is the source error exponent, which is strictly convex and increasing in $[\tau H_{Q_S}(S), \tau \log_2 |\mathcal{S}|]$, and $E(R, W_{Y|X})$ is the channel error exponent, which is a positive and non-increasing in $[0, C(W_{Y|X}))$. Unlike the source exponent, the behavior of $E(R, W_{Y|X})$ is unknown for $R < R_{cr}(W_{Y|X})$. Let C_0 be the zero-error capacity of the channel W, i.e., $E(R, W_{Y|X}) = \infty$ if and only if $R < C_0$ [42]. If $C_0 > \tau \log_2 |\mathcal{S}|$, obviously, we have

$$E_J(Q_S, W_{Y|X}, \tau) = E_T(Q_S, W_{Y|X}, \tau) = +\infty.$$

If $C_0 \leq \tau \log_2 |\mathcal{S}|$, the upper bound in (10.18) is finite and the minimum must be achieved by some rate, say R_m, in the interval $[C_0, \tau \log_2 |\mathcal{S}|]$. Then

$$\begin{aligned} E_J(Q_S, W_{Y|X}, \tau) &\stackrel{(a)}{\leq} \tau e\left(\frac{R_m}{\tau}, Q_S\right) + E(R_m, W_{Y|X}) \\ &\stackrel{(b)}{\leq} \tau e\left(\frac{R_o}{\tau}, Q_S\right) + E(R_o, W_{Y|X}) \\ &\stackrel{(c)}{\leq} 2E(R_o, W_{Y|X}) \\ &= 2E_T(Q_S, W_{Y|X}, \tau). \end{aligned}$$

Here, (a) holds with equality if our computable upper and lower bounds, $\overline{E}_{Jsp}$ and $\underline{E}_{Jr}$, are equal. To ensure this, we need the condition $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ by Theorem 5.2. (b) holds with equality if $R_m = R_o$ by definition of R_m. (c) holds with equality if and only if there is an intersection between $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$, i.e., $\tau e(R_o/\tau, Q_S) = E(R_o, W_{Y|X})$. Now taking these considerations together, and applying Theorem 5.2 again, we conclude that $E_J = 2E_T$ if $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ and $T_{sp}(\overline{\rho}^*, W) - \tau E_s(\overline{\rho}^*, Q_S) = 2\tau e(\overline{R}_m/\tau, Q_S) = 2\tau D(Q_S^{(\overline{\rho}^*)} \parallel Q_S)$. ■

Observation 10.1 The condition for the equality states that, if the minimum in the expression of $\underline{E}_{Jr}$ given in (5.5) is attained at the intersection of $\tau e(\frac{R}{\tau}, W_{Y|X})$ and $E_r(R, W_{Y|X})$ which is no less than the critical rate of the channel, then the JSCC exponent is *twice* as large as the tandem coding exponent. In that case, the rate of decay of the error probability for the JSCC system is *double* that for the tandem coding system. In other words, for the same probability of error P_e, the delay of (optimal) JSCC is approximately *half* of the delay of (optimal) tandem coding,

$$P_e \approx 2^{-nE_T(Q_S, W_{Y|X}, \tau)} = 2^{-\frac{n}{2}E_J(Q_S, W_{Y|X}, \tau)} \qquad \text{for } n \text{ sufficiently large.}$$

10.2.2 Sufficient Conditions for which $E_J > E_T$

In the following we will use our previous results to derive computable sufficient conditions for which $E_J > E_T$. We first define

$$\gamma \triangleq \begin{cases} \text{the root of } \tau H(Q_S^{(\gamma)}) = R_{cr}(W_{Y|X}) & \text{if } \tau H_{Q_S}(S) \le R_{cr}(W_{Y|X}) \le \tau \log_2 |\mathcal{S}|, \\ 0 & \text{if } \tau H_{Q_S}(S) > R_{cr}(W_{Y|X}). \end{cases} \tag{10.19}$$

such that the source error exponent $\tau e(R/\tau, Q_S)$ has a parametric expression at $R_{cr}(W_{Y|X})$

$$\tau e\left(\frac{R_{cr}(W_{Y|X})}{\tau}, Q_S\right) = \tau D(Q_S^{(\gamma)} \parallel Q_S). \tag{10.20}$$

Note that γ is well defined only if $R_{cr}(W_{Y|X}) \le \tau \log_2 |\mathcal{S}|$. Denote

$$T(\overline{\rho}^*) \triangleq T_{sp}(\overline{\rho}^*, W_{Y|X}) - \tau E_s(\overline{\rho}^*, Q_S). \tag{10.21}$$

Theorem 10.3 *Let $R_{cr}(W_{Y|X}) \le \tau \log_2 |\mathcal{S}|$. If*

$$\max\left\{\tau R_{cr}^{(s)}(Q_S), E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S)\right\} \ge R_{cr}(W_{Y|X}), \tag{10.22}$$

then

$$E_J(Q_S, W_{Y|X}, \tau) > E_T(Q_S, W_{Y|X}, \tau).$$

More precisely, we have the following bounds.

(a) If $\min\left\{\tau R_{cr}^{(s)}(Q_S), E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S)\right\} \ge R_{cr}(W_{Y|X})$, then

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) \ge \frac{1}{2}T(\overline{\rho}^*) - \left|\frac{1}{2}T(\overline{\rho}^*) - \tau D(Q_S^{(\overline{\rho}^*)} \parallel Q_S)\right| \ge 0, \tag{10.23}$$

where the two equalities in (10.23) cannot hold simultaneously.

(b) If $\tau R_{cr}^{(s)}(Q_S) \ge R_{cr}(W_{Y|X}) > E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S)$, then

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) > T(\overline{\rho}^*) - \tau D(Q_S^{(\gamma)} \parallel Q_S) \ge 0. \tag{10.24}$$

(c) If $E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S) \ge R_{cr}(W_{Y|X}) > \tau R_{cr}^{(s)}(Q_S)$, then

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) \ge R_{cr}(W_{Y|X}) - \tau E_s(1, Q_S) > 0. \tag{10.25}$$

Proof: We shall show that, in each of the three cases, (a), (b), and (c), we have $E_J > E_T$.

(a). Assume $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ and $E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S) \geq R_{cr}(W_{Y|X})$. By definition of γ, we have $\tau D(Q_S^{(\gamma)} \parallel Q_S) = \tau e(R_{cr}(W_{Y|X})/\tau, Q_S)$, see (2.5) and (10.20). Thus, the latter condition is equivalent to $E(R_{cr}(W_{Y|X}), W_{Y|X}) \geq \tau e(R_{cr}(W_{Y|X})/\tau, Q_S)$ and by (2.33) and the related discussion it guarantees that $R_o \geq R_{cr}(W_{Y|X})$, where R_o is defined in (10.16). According to Theorem 5.2, when $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$, $\overline{E}_{Jsp}(Q_S, W_{Y|X}, \tau)$ is attained by $\overline{R}_m \geq R_{cr}(W_{Y|X})$ and E_J is determined by

$$E_J(Q_S, W_{Y|X}, \tau) = \tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) + E_{sp}(\overline{R}_m, W_{Y|X}).$$

Since $R_o \geq R_{cr}(W_{Y|X})$, E_T is determined by $E_{sp}(R_o, W_{Y|X})$. If $R_o \neq \overline{R}_m$, we must have

$$E_T(Q_S, W_{Y|X}, \tau) < \max\left\{\tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right), E_{sp}(\overline{R}_m, W_{Y|X})\right\},$$

because $\tau e(R/\tau, Q_S)$ is strictly increasing and $E_{sp}(R, W_{Y|X})$ is strictly decreasing at $\overline{R}_m$. Thus,

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) > \min\left\{\tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right), E_r(\overline{R}_m, W_{Y|X})\right\} \geq 0, \quad (10.26)$$

where equality holds if $\overline{R}_m = C(W_{Y|X})$. If $R_o = \overline{R}_m$, then immediately,

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) = \tau e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) = \tau D(Q_S^{(\overline{\rho}^*)} \parallel Q_S), \qquad (10.27)$$

where the above is positive since $\overline{\rho}^* > 0$ by Lemma 5.2 (1). Note also that in this case $\tau e(\overline{R}_m/\tau, Q_S) = E_r(\overline{R}_m, W_{Y|X})$, so (10.26) and (10.27) can be summarized by (10.23).

(b). In this case, we have $\overline{R}_m \geq R_{cr}(W_{Y|X}) > R_o$. We can upper bound E_T by

$$E_T(Q_S, W_{Y|X}, \tau) = \tau e\left(\frac{R_o}{\tau}, Q_S\right) < \tau e\left(\frac{R_{cr}(W_{Y|X})}{\tau}, Q_S\right) = \tau D(Q_S^{(\gamma)} \parallel Q_S)$$

and hence

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) > T_{sp}(\overline{\rho}^*, W) - \tau E_s(\overline{\rho}^*, Q_S) - \tau D(Q_S^{(\gamma)} \parallel Q_S).$$

The above lower bound must be nonnegative since

$$\begin{aligned} & T_{sp}(\overline{\rho}^*, W) - \tau E_s(\overline{\rho}^*, Q_S) - \tau D(Q_S^{(\gamma)} \parallel Q_S) \\ = \quad & E_r(\overline{R}_m, W_{Y|X}) + \tau \left[e\left(\frac{\overline{R}_m}{\tau}, Q_S\right) - e\left(\frac{R_{cr}(W_{Y|X})}{\tau}, Q_S\right)\right] \\ \geq \quad & E_r(\overline{R}_m, W_{Y|X}) \\ \geq \quad & 0, \end{aligned}$$

and it is equal to 0 if $R_{cr}(W_{Y|X}) = \overline{R}_m = C(W_{Y|X})$.

(c). In this case, we have $R_o \geq R_{cr}(W_{Y|X}) > \underline{R}_m$ and from (5.17) E_J is bounded by

$$E_J(Q_S, W_{Y|X}, \tau) \geq E_0(1, W_{Y|X}) - \tau E_s(1, Q_S).$$

On the other hand, by the monotonicity of $E_r(R, W_{Y|X})$, we can upper bound E_T by

$$E_T(Q_S, W_{Y|X}, \tau) = E_r(R_o, W_{Y|X}) \leq E_r(R_{cr}(W_{Y|X}), W_{Y|X}) = E_0(1, W_{Y|X}) - R_{cr}(W_{Y|X}).$$

Thus we obtain

$$E_J(Q_S, W_{Y|X}, \tau) - E_T(Q_S, W_{Y|X}, \tau) \geq R_{cr}(W_{Y|X}) - \tau E_s(1, Q_S).$$

The above is positive since

$$\begin{aligned} E_0(1, W_{Y|X}) - \tau E_s(1, Q_S) &= \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) \\ &> E_r(\underline{R}_m, W_{Y|X}) \\ &> E_r(R_{cr}(W_{Y|X}), W_{Y|X}) \\ &= E_0(1, W_{Y|X}) - R_{cr}(W_{Y|X}), \end{aligned}$$

where the first inequality follows from the fact that $\underline{R}_m > \tau H_{Q_S}(S)$ by Lemma 5.2 and Corollary 5.1.

■

As pointed out in the proof, the condition $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ means that the JSCC exponent E_J is achieved at a rate no less than $R_{cr}(W_{Y|X})$. The second condition,

$E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S) \geq R_{cr}(W_{Y|X})$ means that the tandem coding exponent E_T is achieved at a rate no less than $R_{cr}(W_{Y|X})$. Hence (10.22) in Theorem 10.3 states that E_J would be strictly larger than E_T if either E_J or E_T is determined exactly. Conversely, if the conditions in Theorem 10.3 are not satisfied, then neither E_J nor E_T are exactly known. Nevertheless, if the lower bound of E_J is strictly larger than the upper bound of E_T, then we must have $E_J > E_T$. Hence we obtain the following sufficient conditions.

Theorem 10.4 *Let $E_{ex}(0, W_{Y|X}) < \infty$ and let $\tau \log_2 |\mathcal{S}| \geq R_{cr}(W_{Y|X})$, where $E_{ex}(R, W_{Y|X})$ is the expurgated channel error exponent. If*

$$E_0(1, W_{Y|X}) - \tau E_s(1, Q_S) \geq E_{R_l} \triangleq \frac{k_1 k_2 \tau \log_2 |\mathcal{S}| + k_2 \tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) + k_1 E_{ex}(0, W_{Y|X})}{k_1 - k_2},$$

where

$$k_1 = \frac{D\left(Q_S^{(1)} \parallel Q_S\right) + \log_2(|\mathcal{S}|\overline{Q_S(s)})}{H\left(Q_S^{(1)}\right) - \log_2 |\mathcal{S}|} \qquad \text{and} \qquad k_2 = \frac{E_0(1, W_{Y|X}) - E_{ex}(0, W_{Y|X})}{R_{cr}(W_{Y|X})} - 1,$$

then $E_J(Q_S, W_{Y|X}, \tau) > E_T(Q_S, W_{Y|X}, \tau)$.

Proof: We first recall that if $-\tau \log(|\mathcal{S}|\overline{Q_S(s)}) < E(\tau \log |\mathcal{S}|, W_{Y|X})$, then there is no intersection between $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$. Clearly, the tandem coding exponent satisfies

$$\begin{aligned}
E_T(Q_S, W, \tau) &= E(\tau \log |\mathcal{S}|, W_{Y|X}) \\
&= E_r(\tau \log |\mathcal{S}|, W_{Y|X}) \qquad (10.28) \\
&< E_r(\underline{R}_m, W_{Y|X}) \qquad (10.29) \\
&\leq E_J(Q_S, W, \tau),
\end{aligned}$$

Here, (10.28) follows by hypothesis $R_{cr}(W_{Y|X}) \leq \tau \log |\mathcal{S}|$. (10.29) holds since $\underline{R}_m$ must be a quantity smaller than $\tau \log |\mathcal{S}|$ by Corollary 5.1.

We hence assume that $-\tau \log(|\mathcal{S}|\overline{Q_S(s)} \geq E(\tau \log |\mathcal{S}|, W_{Y|X})$, i.e., we assume that $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$ intersect at rate R_o. If $R_o \geq R_{cr}(W_{Y|X})$, which means $E_o(1, W_{Y|X}) - R_{cr}(W_{Y|X}) \geq \tau e(R_{cr}(W_{Y|X})/\tau, Q_S)$, then Theorem 10.3 guarantees that

$E_J > E_T$. If $\underline{R}_m \geq R_{cr}(W_{Y|X})$, which implies $\tau R_{cr}^{(s)}(Q_S) \geq R_{cr}(W_{Y|X})$ by Corollary 5.2. This ensures $E_J > E_T$ by Theorem 10.3. Furthermore, if $R_{cr}(W_{Y|X}) > \underline{R}_m \geq R_o$, then

$$\begin{aligned} E_J(Q_S, W, \tau) &\geq \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) + E_r(\underline{R}_m, W_{Y|X}) \\ &> \tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) \\ &\geq \tau e\left(\frac{R_o}{\tau}, Q_S\right) \\ &= E_T(Q_S, W, \tau). \end{aligned}$$

In the remaining, we assume that $\tau e(R/\tau, Q_S)$ and $E(R, W_{Y|X})$ intersect at rate R_o and that $\underline{R}_m < R_o < R_{cr}$.

For a DMC with $E_{ex}(0, W_{Y|X}) < \infty$, we may use the straight-line upper bound for the channel error exponent $E_{sl}(R, W_{Y|X})$ given by (2.30) such that $E_{sl}(R, W_{Y|X})$ is a straight line passing $(0, E_{ex}(0, W_{Y|X}))$ in $[0, R_l]$ $(R_l \leq R_{cr}(W_{Y|X}))$ and is tangent to the sphere-packing bound at R_l. So $E_{sl}(R, W_{Y|X})$ is also convex in $0 \leq R \leq C(W_{Y|X})$, and

$$E_{sl}(0, W_{Y|X}) = E_{ex}(0, W_{Y|X}).$$

Now connect $(0, E_{sl}(0, W_{Y|X}))$ and $(R_{cr}(W_{Y|X}), E_{sl}(R_{cr}(W_{Y|X}), W_{Y|X}))$ with a straight line, denoted by l_1, where

$$E_{sl}(R_{cr}(W_{Y|X}), W_{Y|X}) = E_r(R_{cr}(W_{Y|X}), W_{Y|X}) = E_0(1, W_{Y|X}) - R_{cr}(W_{Y|X}).$$

Again, connect $(\underline{R}_m, \tau e(\underline{R}_m/\tau, Q_S))$ and $(\tau \log|\mathcal{S}|,\ \tau e(\log|\mathcal{S}|, Q_S))$ with a straight line, denoted by l_2, where

$$\tau e\left(\frac{\underline{R}_m}{\tau}, Q_S\right) = \tau D(Q_S^{(1)} \parallel Q_S),$$

and

$$\tau e(\log|\mathcal{S}|, Q_S) = -\tau \log(|\mathcal{S}|\overline{Q_S(s)}).$$

Suppose that the intersection of $E_{sl}(R, W_{Y|X})$ and $\tau e(R/\tau, Q_S)$ is $(R_1, \tau e(R_1/\tau, Q_S))$, and that the intersection of l_1 and l_2 is (R_l, E_{R_l}). By assumption, R_o, the intersection of $\tau e(R/\tau, W_{Y|X})$ and $E(R, W_{Y|X})$, is strictly larger than $\underline{R}_m$ and strictly less than $R_{cr}(W_{Y|X})$; hence by definition, R_1, the intersection of $\tau e(R/\tau, W_{Y|X})$ and $E_s(R, W_{Y|X})$, must be strictly larger than $\underline{R}_m$ and strictly less than $R_{cr}(W_{Y|X})$, i.e., $\underline{R}_m < R_1 \le R_o < R_{cr}(W_{Y|X})$. Likewise, it is easily seen that $\underline{R}_m < R_l < R_{cr}(W_{Y|X})$. Furthermore, because of the convexity of $\tau e(R/\tau, Q_S)$ and $E_{sl}(R, W_{Y|X})$ in the region $[\underline{R}_m, R_{cr}(W_{Y|X})]$, E_{R_l} must be strictly larger than $\tau e(R_1/\tau, Q_S)$ (as $\tau e(R/\tau, W_{Y|X})$ is strictly convex in this interval). It follows that

$$\begin{aligned} E_J(Q_S, W, \tau) &\ge E_0(1, W_{Y|X}) - \tau E_s(1, Q_S) \ge E_{R_l} \\ &> \tau e\left(\frac{R_1}{\tau}, Q_S\right) \ge \tau e\left(\frac{R_o}{\tau}, Q_S\right) = E_T(Q_S, W, \tau). \end{aligned}$$

■

Theorem 10.5 *Let $\tau \log_2 |\mathcal{S}| \ge R_{cr}(W_{Y|X})$. If $E_0(1, W_{Y|X}) - \tau E_s(1, Q_S) \ge \tau D\left(Q_S^{(\gamma)} \parallel Q_S\right)$, where γ is defined in (10.19), then $E_J(Q_S, W_{Y|X}, \tau) > E_T(Q_S, W_{Y|X}, \tau)$.*

Proof: As in the previous proof, we only consider the case

$$-\tau \log_2(|\mathcal{S}|\overline{Q_S(s)}) \ge E(\tau \log_2 |\mathcal{S}|, W_{Y|X})$$

and $\underline{R}_m < R_o < R_{cr}(W_{Y|X})$. Thus, we can upper bound E_T by

$$\begin{aligned} E_T(Q_S, W, \tau) &= \tau e\left(\frac{R_o}{\tau}, Q_S\right) \\ &< \tau e\left(\frac{R_{cr}(W_{Y|X})}{\tau}, Q_S\right) \\ &= \tau D\left(Q_S^{(\gamma)} \parallel Q_S)\right) \end{aligned}$$

by the strict monotonicity of the source error exponent. On the other hand, Theorem 5.2 gives that

$$E_J(Q_S, W, \tau) \ge E_0(1, W_{Y|X}) - \tau E_s(1, Q_S).$$

By assumption, if $E_0(1, W_{Y|X}) - \tau E_s(1, Q_S) \ge \tau D\left(Q_S^{(\gamma)} \parallel Q_S)\right)$, then $E_J > E_T$. ■

In Theorems 10.4 and 10.5, we establish the sufficient conditions by comparing the source-channel random-coding bound derived in Theorem 5.2, with the upper bound of tandem coding exponent obtained by using the geometric characteristics of $e(R, Q_S)$ and

$E(R, W_{Y|X})$. These conditions can be readily computed since it only requires the knowledge of $R_{cr}(W_{Y|X})$ and $E_{ex}(0, W_{Y|X})$. Note that the condition $E_{ex}(0, W_{Y|X}) < \infty$ in Theorem 10.4 is satisfied by the DMCs with zero-error capacity equal to 0, see [32, p. 187]. Thus, Theorem 10.4 applies to equidistant channels, in particular, to every channel with binary input alphabet.

Example 10.1 (When Does the JSCC Exponent Outperform the Tandem Coding Exponent?) We apply Theorems 10.3, 10.4 and 10.5 to the binary DMS with distribution $\{q, 1-q\}$ and BSC with crossover probability ε, and the binary DMS $\{q, 1-q\}$ and BEC with erasure probability α, under different transmission rates τ. If any one of the conditions in these theorems holds, then $E_J > E_T$. The above conditions are summarized by Region **F** in Fig. 10.3. Indeed, Region **F** shows that $E_J > E_T$ for a wide range of (ε, q) or (α, q) pairs. Region **G** consists of the pairs (ε, q) or (α, q) such that $\tau H_{Q_S}(S) \geq C(W_{Y|X})$; in this case, $E_J = E_T = 0$. Finally, when (ε, q) or (α, q) falls in Region **H**, we are not sure whether E_J is still strictly larger than E_T.

Example 10.2 (By How Much Can the JSCC Exponent Be Larger Than the Tandem Coding Exponent?) In the last example we have seen that $E_J > E_T$ holds for a wide large class of source-channel pairs. Now we evaluate the performance of E_J over E_T by looking at the ratio of the two quantities. Recall that when Theorem 10.3 (a) is satisfied, both E_J and E_T are exactly determined. In this case we can directly compute E_J (using the results of Section 5.2) and E_T (using (10.14) and (10.15)). When E_J (E_T, respectively) is not known, i.e., when $\tau R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$ ($E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S) < R_{cr}(W_{Y|X})$, respectively), we can calculate the lower bound of E_J (the upper bound of E_T, respectively) instead and thus obtain a lower bound of E_J/E_T. For general DMCs, we lower bound E_J by its random-coding lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$. For equidistant DMCs, particularly for binary DMCs, when $\tau R_{cr}^{(s)}(Q_S) < R_{ex}(W_{Y|X})$, we use the expurgated lower bound $\underline{E}_{ex}(Q_S, W_{Y|X}, \tau)$; when $\tau R_{cr}^{(s)}(Q_S) \geq R_{ex}(W_{Y|X})$, we use the random-coding lower bound $\underline{E}_{Jr}(Q_S, W_{Y|X}, \tau)$. To calculate the upper bound of E_T, when $E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel$

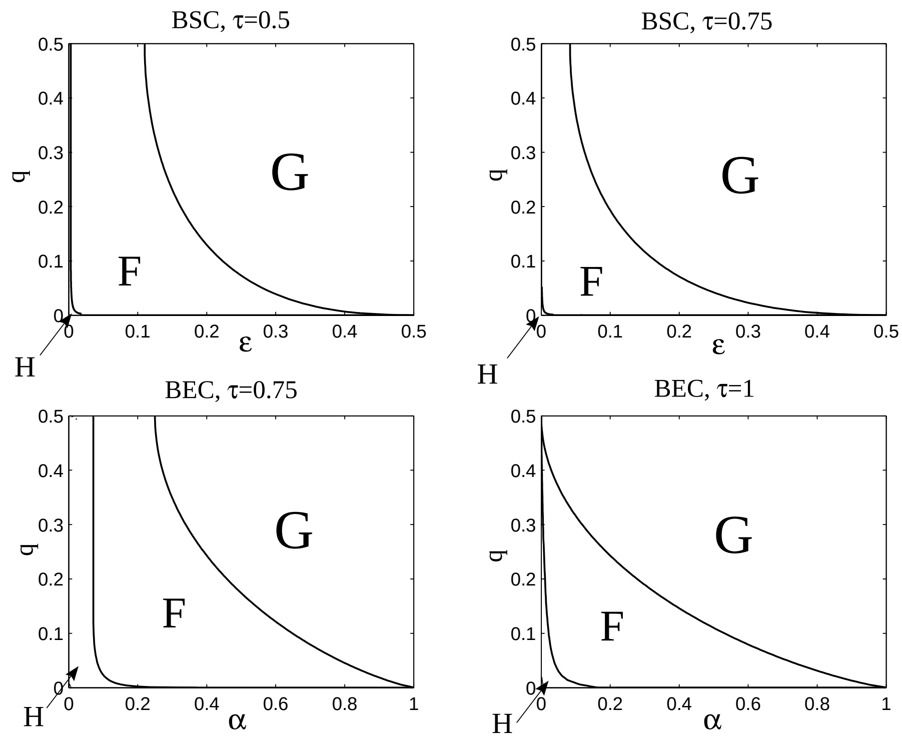

Figure 10.3: The regions for binary DMS-BSC (q, ε) pairs and binary DMS-BEC (q, α) pairs under different transmission rates τ of Example 10.1. In region **F** (including the boundary between **F** and **H**), $E_J > E_T > 0$; in region **G** (including the boundary between **G** and **F**), $E_J = E_T = 0$; and in region **H**, $E_J \geq E_T > 0$.

$Q_S) < R_{cr}(W_{Y|X}) \leq R_{cr}^{(s)}(Q_S)$, or equivalently when $R_o < R_{cr}(W_{Y|X}) \leq \overline{R}_m$, we can bound E_T by

$$E_T(Q_S, W_{Y|X}, \tau) \leq \min\left\{\tau D\left(Q_S^{(\gamma)} \parallel Q_S\right), E_{sp}(R_s, W_{Y|X})\right\},$$

where R_s is the intersection of $E_{sp}(R, W_{Y|X})$ and $\tau e(R/\tau, Q_S)$ if any; otherwise $R_s = \tau \log_2 |S|$. When $E_o(1, W_{Y|X}) - \tau D(Q_S^{(\gamma)} \parallel Q_S) < R_{cr}(W_{Y|X})$ and $R_{cr}^{(s)}(Q_S) < R_{cr}(W_{Y|X})$, we bound E_T by

$$E_T(Q_S, W_{Y|X}, \tau) \leq E_{sp}(R_s, W_{Y|X}).$$

Table 10.1 exhibits E_J/E_T (or its lower bound, which must be no less than 1) for the binary DMS $\{q, 1-q\}$ and BSC (ε) systems under transmission rates $\tau = 0.5$, 0.75 and 1. It is seen that the ratio E_J/E_T can be very close to 2 (its upper bound) for many (q,ε) pairs. For other systems, we have similar results: E_J substantially outperforms E_T. For instance, for binary DMS $\{q,1-q\}$ and BEC (α) with $\tau=1$, we can obtain $E_J/E_T \geq 1.4$ for a wide range of (q,α)'s; for ternary DMS and BSC or for DMS and ternary symmetric channel, if transmission rate τ is chosen suitably (such that $\tau H_{Q_S}(S) < C(W_{Y|X})$), we can obtain $E_J/E_T \geq 1.5$ for many source-channel pairs.

E_J/E_T	$\tau=0.5$, $q=0.1$	$\tau=0.75$, $q=0.1$	$\tau=0.75$, $q=0.15$	$\tau=1$, $q=0.05$
$\varepsilon=0.0005$	$1.0^\dagger$	$1.60^\dagger$	$1.58^\dagger$	$1.87^\dagger$
$\varepsilon=0.001$	$1.0^\dagger$	$1.70^\dagger$	$1.68^\dagger$	$1.93^\dagger$
$\varepsilon=0.005$	$1.36^\dagger$	$1.94^\dagger$	1.89	1.99
$\varepsilon=0.01$	$1.70^\dagger$	1.95	1.91	2.0
$\varepsilon=0.04$	1.85	1.97	1.95	2.0
$\varepsilon=0.08$	1.91	1.99	1.96	2.0
$\varepsilon=0.12$	1.95	1.97	2.0	2.0
$\varepsilon=0.16$	1.96	1.95	N/A	2.0
$\varepsilon=0.2$	1.86	N/A	N/A	N/A

Table 10.1: E_J/E_T for the binary DMS and BSC pairs of Example 10.2. "N/A" means that $\tau H(Q) > C$ such that $E_J = E_T = 0$. "†" means that this quantity is only a lower bound of E_J/E_T.

10.2.3 Power Gain Due to JSCC for DMS over Binary-input AWGN and Rayleigh-Fading Channels with Finite Output Quantization

It is well known that M-ary modulated additive white Gaussian noise (AWGN) and memoryless Rayleigh-fading channels can be converted to a DMC when finite quantization is applied at their output. For example, as illustrated in [5], [74], we know that the concatenation of a binary phase-shift keying (BPSK) modulated AWGN or Rayleigh-fading channel with m-bit soft-decision demodulation is equivalent to a binary-input, 2^m-output DMC (cf. Fig. 10.4). We next study the JSCC and tandem coding exponent for a system involving such channels to assess the potential benefits of JSCC over tandem coding in terms of power or channel signal-to-noise ratio (SNR) gains.

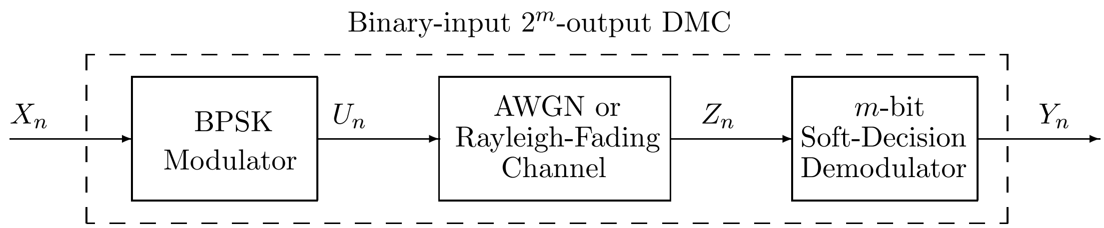

Figure 10.4: Binary-input AWGN or Rayleigh-fading channel with finite output quantization.

We assume that the BPSK signal $U_n \in \{-1, +1\}$ corresponding to the signal input X_n is of unit energy, and V_n is a zero-mean independent and identically distributed (i.i.d.) Gaussian random process with variance $N_o/2$. The channel SNR is defined by SNR $\triangleq E[U_n^2]/E[V_n^2] = 2/N_o$ and the received signal is

$$Z_n = A_n U_n + V_n, \qquad n = 1, 2, ...,$$

where A_n is 1 for the AWGN channel (no fading), and for the Rayleigh-fading channel, $\{A_n\}$ is the amplitude fading process assumed to be i.i.d. with pdf

$$f_A(a) = \begin{cases} 2ae^{-a^2}, & \text{if } a > 0, \\ 0, & \text{otherwise,} \end{cases}$$

such that $E[A_n^2] = 1$. We also assume for the Rayleigh-fading channel that A_n, U_n and V_n are independent of each other, and the values of A_n are not available at the receiver. At the receiver, as shown in Fig. 10.4, each $Z_n \in \mathbb{R}$ is demodulated via an m-bit uniform scalar quantizer with quantization step Δ to yield $Y_n \in \{0, 1\}^m$. If the channel input alphabet is $\mathcal{X} = \{0, 1\}$ and the channel output alphabet is $\mathcal{Y} = \{0, 1, 2, ..., 2^m - 1\}$, then the transition probability matrix Π is given by

$$\Pi = [\pi_{ij}], \qquad i \in \mathcal{X}, \qquad j \in \mathcal{Y},$$

where

$$\pi_{ij} \triangleq P(Y = j|X = i) = \mathcal{Q}_{\mathcal{S}}\left((T_{j-1} - (2i-1))\sqrt{\text{SNR}}\right) - \mathcal{Q}_{\mathcal{S}}\left((T_j - (2i-1))\sqrt{\text{SNR}}\right)$$

for the AWGN channel [74], and

$$\pi_{ij} \triangleq P(Y = j|X = i) = F_{Z|X}(T_j|i) - F_{Z|X}(T_{j-1}|i)$$

for the Rayleigh-fading channel [5]. Here $F_{Z|X}(z|i) = Pr\{Z \leq z|Z = i\}$ is given by [5], [92]

$$\begin{aligned} F_{Z|X}(z|1) &= 1 - F_{Z|X}(-z|0) \\ &= 1 - \mathcal{Q}_{\mathcal{S}}\left(\frac{z}{\sqrt{N_o/2}}\right) - \frac{e^{-(z^2/(N_o+1))}}{\sqrt{N_o+1}} \times \left[1 - \mathcal{Q}_{\mathcal{S}}\left(\frac{z}{\sqrt{N_o(N_o+1)/2}}\right)\right], \end{aligned}$$

where $\mathcal{Q}_{\mathcal{S}}(x)$ is the complementary error function

$$\mathcal{Q}_{\mathcal{S}}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left\{-t^2/2\right\} dt,$$

and $\{T_j\}$ are the thresholds of the receiver's soft-decision quantizer given by

$$T_j = \begin{cases} -\infty, & \text{if } j = -1, \\ (j + 1 - 2^{m-1})\Delta, & \text{if } j = 0, 1, ..., 2^m - 2, \\ +\infty, & \text{if } j = 2^m - 1 \end{cases} \tag{10.30}$$

with uniform step-size Δ. For each channel SNR, the suitable quantization step Δ is chosen as in [74], [5] to yield the maximum capacity of the binary-input 2^m-output DMC.

We compute the JSCC and tandem coding exponents for the binary source and the binary-input 2^m-output DMC converted from the AWGN (Rayleigh-fading, respectively) channel under transmission rate $\tau = 0.75$ ($\tau = 1$, respectively), and illustrate the power gain due to JSCC. In Figs. 10.5 and 10.6, we plot E_J and E_T for binary DMS $Q_S = \{0.1, 0.9\}$ and $m = 1, 2, 3$ by varying the channel SNR (in dB). We point out that in both the two figures, when SNR ≤ 6 dB for $m = 2, 3$ and when SNR ≤ 8 dB for $m = 1$, E_J and E_T are determined exactly. We observe that for the same SNR, E_J is almost twice as large as E_T ($E_J \approx 2E_T$ for 1dB$\leq$SNR$\leq$ 8dB, $m = 1$, and for 0dB$\leq$SNR$\leq$ 6dB, $m = 2, 3$). Furthermore, for the same exponent and the same (asymptotic) encoding length, JSCC would yield the same probability of error as tandem coding with a power gain of more than 2 dB. A similar behavior was noted for other values of transmission rate τ.

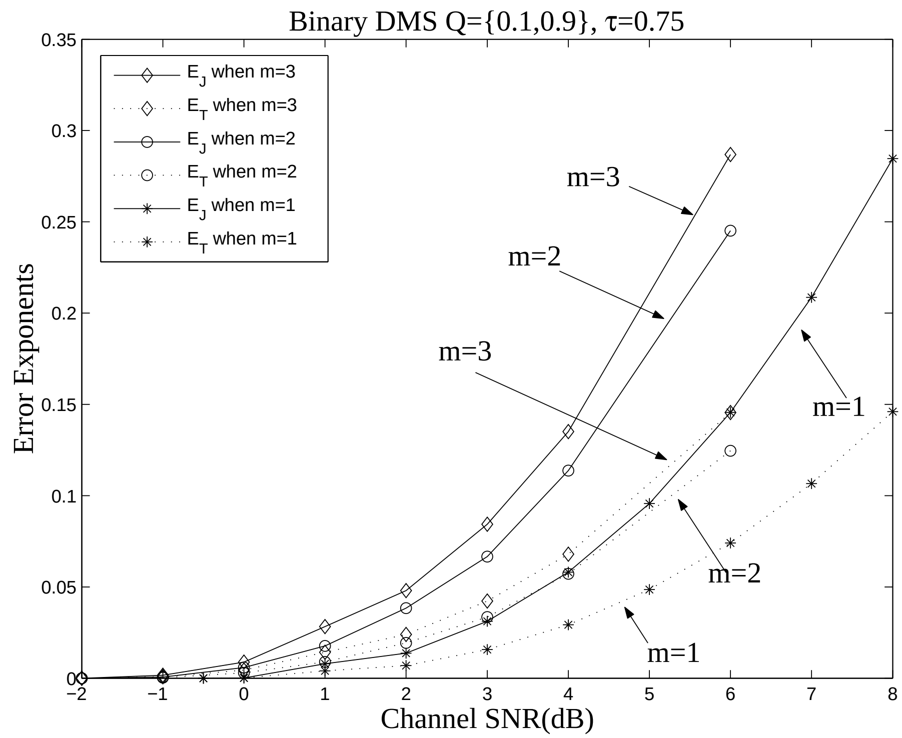

Figure 10.5: The power gain due to JSCC for binary DMS and binary-input 2^m-output DMC (AWGN channel) with $\tau = 0.75$.

10.3 Systems with Markovian Memory

In this section, we assume that the source $\mathbf{Q_S}$ is an SEM source and the channel $\mathbf{W_{Y|X}}$ is an SEM channel. We hence have the following corollary.

Corollary 10.2 *Let* $\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ *and let* $\lambda_0^\tau(\mathbf{Q_S})\lambda_0(\mathbf{P_W}) > B$*. Then*

$$\begin{aligned}\underline{E}_{Tr}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) &\leq E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \\ &= \sup_{\tau \overline{H}_{\mathbf{Q_S}}(S) \leq R \leq \overline{C}(\mathbf{W_{Y|X}})} \min\left\{\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E(R, \mathbf{W_{Y|X}})\right\} \quad (10.31) \\ &\leq \overline{E}_{Tsp}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau), \quad (10.32)\end{aligned}$$

where

$$\underline{E}_{Tr}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \triangleq \sup_{\tau \overline{H}_{\mathbf{Q_S}}(S) \leq R \leq \overline{C}(\mathbf{W_{Y|X}})} \min\left\{\tau \overline{e}\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E_r(R, \mathbf{W_{Y|X}})\right\} \quad (10.33)$$

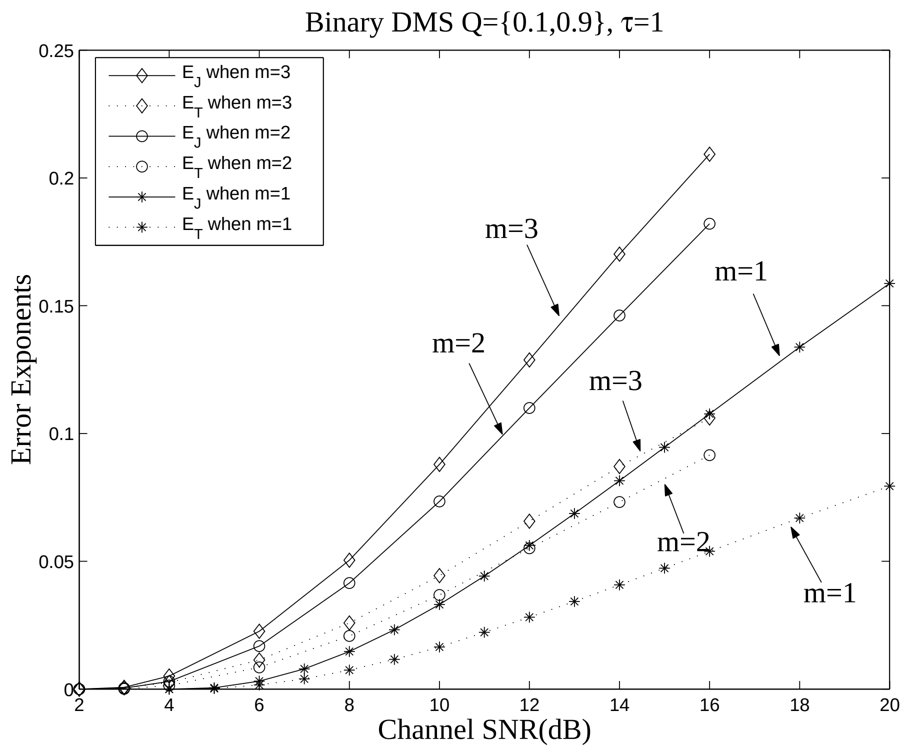

Figure 10.6: The power gain due to JSCC for binary DMS and binary-input 2^m-output DMC (Rayleigh-fading channel) with $\tau = 1$.

and

$$\overline{E}_{Tsp}(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \triangleq \sup_{\tau \overline{H}_{\mathbf{Q_S}}(S) \le R \le \overline{C}(\mathbf{W_{Y|X}})} \min \left\{ \tau \overline{e}\left(\frac{R}{\tau}, \mathbf{Q_S}\right), E_{sp}(R, \mathbf{W_{Y|X}}) \right\} \quad (10.34)$$

where $\overline{e}\,(R, \mathbf{Q_S}) = e(R, \mathbf{Q_S})$ *is the SEM source error exponent given in Corollary 7.1, and* $E_{sp}(R, \mathbf{W_{Y|X}})$ *and* $E_r(R, \mathbf{W_{Y|X}})$ *are respectively the upper and lower bounds for the SEM channel error exponent given in Corollary 7.2 and (7.28).*

Remark 10.4 If $\tau \overline{H}_{\mathbf{Q_S}}(S) \ge \overline{C}(\mathbf{W_{Y|X}})$, $E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 0$.

To evaluate E_T for an SEM source-channel pair $(\mathbf{Q_S}, \mathbf{W_{Y|X}})$, we recall that $e(R, \mathbf{Q_S})$ is 0 for $R \le H(\mathbf{Q_S})$, strictly increasing in $H(\mathbf{Q_S}) \le R \le \log_2 \lambda_0(\mathbf{Q_S})$ and infinity for $R > \log_2 \lambda_0(\mathbf{Q_S})$ ([72], [94]), while $E(R, \mathbf{W_{Y|X}})$ is non-increasing and positive in $R < C(\mathbf{W_{Y|X}})$, and vanishes at $R = C(\mathbf{W_{Y|X}})$.

Therefore, if the graphs of $\tau e\left(R/\tau, \mathbf{Q_S}\right)$ and $E(R, \mathbf{W_{Y|X}})$ have an intersection at R_o, then it immediately follows from (10.31) that

$$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = \tau e\left(\frac{R_o}{\tau}, \mathbf{Q_S}\right) = E(R_o, \mathbf{W_{Y|X}}).$$

If there is no intersection between $\tau e\left(R/\tau, \mathbf{Q_S}\right)$ and $E(R, \mathbf{W_{Y|X}})$ then by (10.31)

$$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = E(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{W_{Y|X}}).$$

10.3.1 E_J Can At Most Double E_T

Similarly as DMS-DMC pairs, for SEM source-channel pairs, the JSCC coding exponent can at most double the tandem coding exponent.

Theorem 10.6 *For an SEM source* $\mathbf{Q_S}$ *and an SEM channel* $\mathbf{W_{Y|X}}$*, the JSCC error exponent is upper bounded by twice the tandem coding exponent*

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \leq 2E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

This relation follows from Theorem 7.5 and Corollary 10.2. The proof is similar as the one of Theorem 10.2 and is omitted.

10.3.2 Sufficient Conditions for which $E_J > E_T$

When the entropy rate of the SEM source is equal to $\log_2 \lambda_0(\mathbf{Q_S})$, the error exponent would be zero for $R \leq \log_2 \lambda_0(\mathbf{Q_S})$ and infinity otherwise. In this case, the source is incompressible and only channel coding is performed in both JSCC and tandem coding; as a result,

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = E(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{W_{Y|X}})$$

by (7.29), (7.30) and (10.31). Note that $\log_2 \lambda_0(\mathbf{Q_S})$ might not be equal to $\log_2 |\mathcal{S}|$ by Lemma 7.1, as compared with the DMS. Thus, we assume in the rest of the section that $\overline{H}_{\mathbf{Q_S}}(S) < \log_2 \lambda_0(\mathbf{Q_S})$ (such that the source is compressible) and that $\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ (such that both E_J and E_T are positive).

Theorem 10.7 *Let f be defined by (7.10). If $f(1) \le 0$, i.e.,*

$$\tau \overline{H}_{\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{2})}}(S) + \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{S}}^{(\frac{1}{2})}}(Z) \ge \log_2 B,$$

then $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$.

Proof: Since we assumed that $\tau \overline{H}_{\mathbf{Q_S}}(S) < \overline{C}(\mathbf{W_{Y|X}})$ or equivalently $f(0) > 0$ (see Lemma 7.4), if now $f(1) \le 0$, then there exists some ρ $(0 < \rho \le 1)$ such that $f(\rho) = 0$ by the continuity of $f(\cdot)$. Let ρ^* be the smallest one satisfying $f(\rho^*) = 0$. According to Theorem 7.3, the JSCC error exponent is determined exactly by $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = F(\rho^*)$. On the other hand, we know from (7.29) that

$$F(\rho^*) = \min_R \left[\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right) + E_{sp}(R, \mathbf{W_{Y|X}}) \right].$$

Suppose the above minimum is achieved by some $\overline{R}_m$, i.e.,

$$F(\rho^*) = \tau e\left(\frac{\overline{R}_m}{\tau}, \mathbf{Q_S}\right) + E_{sp}(\overline{R}_m, \mathbf{W_{Y|X}}).$$

It can be shown (cf. Lemma 5.2) that $\overline{R}_m$ is related to ρ^* as follows

$$\overline{R}_m = \tau \overline{H}_{\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho^*})}}(S) = \log_2 B - \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{S}}^{(\frac{1}{1+\rho^*})}}(Z).$$

Since ρ^* is positive, from the above we know $\tau \overline{H}_{\mathbf{Q_S}}(S) \le \overline{R}_m \le \overline{C}(\mathbf{W_{Y|X}})$ by the monotonicity of $\overline{H}_{\widetilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho})}}(S)$ and $\overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{S}}^{(\frac{1}{1+\rho})}}(Z)$. In the following we first assume that $\tau e(R/\tau, \mathbf{Q_S})$ and $E(R, \mathbf{W_{Y|X}})$ intersect at R_o, i.e., there exists an $R_o \in (\tau \overline{H}_{\mathbf{Q_S}}(S), \overline{C}(\mathbf{W_{Y|X}}))$ such that

$$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = \tau e\left(\frac{R_o}{\tau}, \mathbf{Q_S}\right) = E(R_o, \mathbf{W_{Y|X}}) > 0.$$

If $\overline{R}_m > R_o$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \ge \tau e\left(\frac{\overline{R}_m}{\tau}, \mathbf{Q_S}\right) > \tau e\left(\frac{R_o}{\tau}, \mathbf{Q_S}\right) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

If $\overline{R}_m = R_o$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 2E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

If $\overline{R}_m < R_o$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \ge E_{sp}(\overline{R}_m, \mathbf{W_{Y|X}}) > E_{sp}(R_o, \mathbf{W_{Y|X}}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

We next assume that there is no intersection between $\tau e(R/\tau, \mathbf{Q_S})$ and $E(R, \mathbf{W_{Y|X}})$, i.e., $\tau e(R/\tau, \mathbf{Q_S}) < E(R, \mathbf{W_{Y|X}})$ for all $R < \tau \log_2 \lambda_0(\mathbf{Q_S})$. If $\overline{R}_m = \tau \overline{H}_{\mathbf{Q_S}}(S)$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = E_{sp}(\overline{R}_m, \mathbf{W_{Y|X}}) > E_{sp}(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{Q_S}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$$

since $\overline{H}_{\mathbf{Q_S}}(S) < \log_2 \lambda_0(\mathbf{Q_S})$ is assumed. If $\overline{R}_m > \tau H(\mathbf{Q_S})$, then

$$\begin{aligned} E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) &\geq \tau e\left(\frac{\overline{R}_m}{\tau}, \mathbf{Q_S}\right) + E_{sp}(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{Q_S}) \\ &> E_{sp}(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{Q_S}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \end{aligned}$$

since the source error exponent is positive at $\overline{R}_m > \tau \overline{H}_{\mathbf{Q_S}}(S)$. ■

Theorem 10.7 states that if E_J is determined exactly (i.e., its upper and lower bounds coincide), no matter whether E_T is known or not, then the JSC coding exponent is larger than the tandem exponent. Conversely, if E_T is determined exactly, irrespective of whether E_J is determined or not, the strict inequality between E_J and E_T also holds, as shown by the following results.

Theorem 10.8

(a) If $\tau\overline{H}_{\mathbf{Q_S}}(S) \geq R_{cr}(\mathbf{W_{Y|X}})$*, then* $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$.

(b) Otherwise, if $\tau\overline{H}_{\mathbf{Q_S}}(S) < R_{cr}(\mathbf{W_{Y|X}})$ *and* $\tau \log_2 \lambda_0(\mathbf{Q_S}) > R_{cr}(\mathbf{W_{Y|X}})$*, there must exist some* ρ *satisfying* $\tau\overline{H}_{\tilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho})}}(S) = R_{cr}$*. Let* ρ_m *be the smallest one satisfying such equation. If*

$$(1+\rho_m)\tau[\overline{H}_{\tilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho_m})}}(S) - \log_2 \lambda_{\frac{1}{1+\rho_m}}(\mathbf{Q_S})] \leq \log_2 B - 2\log_2 \lambda_{\frac{1}{2}}(\mathbf{P_Z}),$$

then $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$.

Remark 10.5 By the monotonicity of $\overline{H}_{\tilde{\mathbf{Q}}_{\mathbf{S}}^{(\frac{1}{1+\rho})}}(S)$, ρ_m can be solved numerically.

Proof: Recall that

$$R_{cr}(\mathbf{W_{Y|X}}) = \log_2 B - \overline{H}_{\tilde{\mathbf{P}}_{\mathbf{S}}^{(\frac{1}{2})}}(Z)$$

is the critical rate of the channel $\mathbf{W_{Y|X}}$ such that the channel exponent is determined for $R \geq R_{cr}(\mathbf{W_{Y|X}})$, i.e.,

$$E(R, \mathbf{W_{Y|X}}) = E_r(R, \mathbf{W_{Y|X}}) = \overline{E}(R, \mathbf{W_{Y|X}})$$

if $R \geq R_{cr}(\mathbf{W_{Y|X}})$.

We first show that $E_J > E_T$ if $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) \leq E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}})$, and then we show that $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) \leq E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}})$ if and only if (a) or (b) holds.

Now if $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) \leq E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}})$, then $E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ is determined exactly. There are two cases to consider:

(a) If $\tau e(R/\tau, \mathbf{Q_S})$ and $E(R, \mathbf{W_{Y|X}})$ intersect at R_o such that $R_{cr}(\mathbf{W_{Y|X}}) \leq R_o < C(\mathbf{W_{Y|X}})$, then

$$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = \tau e\left(\frac{R_o}{\tau}, \mathbf{Q_S}\right) = E_r(R_o, \mathbf{W_{Y|X}}) > 0.$$

On the other hand, (7.18) and (7.19) yield

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \geq \max_{0\leq\rho\leq 1} F(\rho) = F(\rho_*),$$

where $\rho_* = \min(1, \rho^*) > 0$ and recall that ρ^* is the smallest positive number satisfying $f(\rho^*) = 0$. It follows from (7.30) that

$$F(\rho_*) = \min_R \left[\tau e\left(\frac{R}{\tau}, \mathbf{Q_S}\right) + E_r(R, \mathbf{W_{Y|X}}) \right].$$

Similar to $\overline{R}_m$ in the last proof, it can be shown (cf. Lemma 5.2) that the above minimum is achieved by some $\underline{R}_m$ such that

$$\underline{R}_m = \tau \overline{H}_{\tilde{\mathbf{Q}}_\mathbf{S}^{(\frac{1}{1+\rho^*})}}(S) \geq \tau \overline{H}_{\mathbf{Q_S}}(S).$$

If $\underline{R}_m > R_o$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \geq \tau e\left(\frac{\underline{R}_m}{\tau}, \mathbf{Q_S}\right) > \tau e\left(\frac{R_o}{\tau}, \mathbf{Q_S}\right) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

If $\underline{R}_m = R_o$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) = 2E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

If $\underline{R}_m < R_o$, likewise, we have

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \geq E_r(\underline{R}_m, \mathbf{W_{Y|X}}) > E_r(R_o, \mathbf{W_{Y|X}}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

(b) If $\tau e(R/\tau, \mathbf{Q_S})$ and $E(R, \mathbf{W_{Y|X}})$ have no intersection, we still have, as in the last proof, if $\underline{R}_m = \tau \overline{H}_{\mathbf{Q_S}}(S)$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \geq E_r(\underline{R}_m, \mathbf{W_{Y|X}}) > E_r(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{Q_S}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau);$$

otherwise if $\underline{R}_m > \tau \overline{H}_{\mathbf{Q_S}}(S)$, then

$$E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) > E_r(\underline{R}_m, \mathbf{W_{Y|X}}) \geq E_r(\tau \log_2 \lambda_0(\mathbf{Q_S}), \mathbf{W_{Y|X}}) = E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau).$$

Finally, we point out that the sufficient and necessary conditions for

$$\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) \leq E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}})$$

is that

(a) $\tau \overline{H}_{\mathbf{Q_S}}(S) \geq R_{cr}(\mathbf{W_{Y|X}})$ such that $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) = 0$; or

(b) $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) > 0$ but $\tau e(R_{cr}(\mathbf{W_{Y|X}})/\tau, \mathbf{Q_S}) \leq E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}})$. Using the fact that

$$E(R_{cr}(\mathbf{W_{Y|X}}), \mathbf{W_{Y|X}}) = \overline{H}_{\widetilde{\mathbf{P}}_{\mathbf{S}}^{(\frac{1}{2})}}(Z) - 2\log_2 \lambda_{\frac{1}{2}}(\mathbf{P_Z}),$$

we obtain Condition (b) and complete the proof. ■

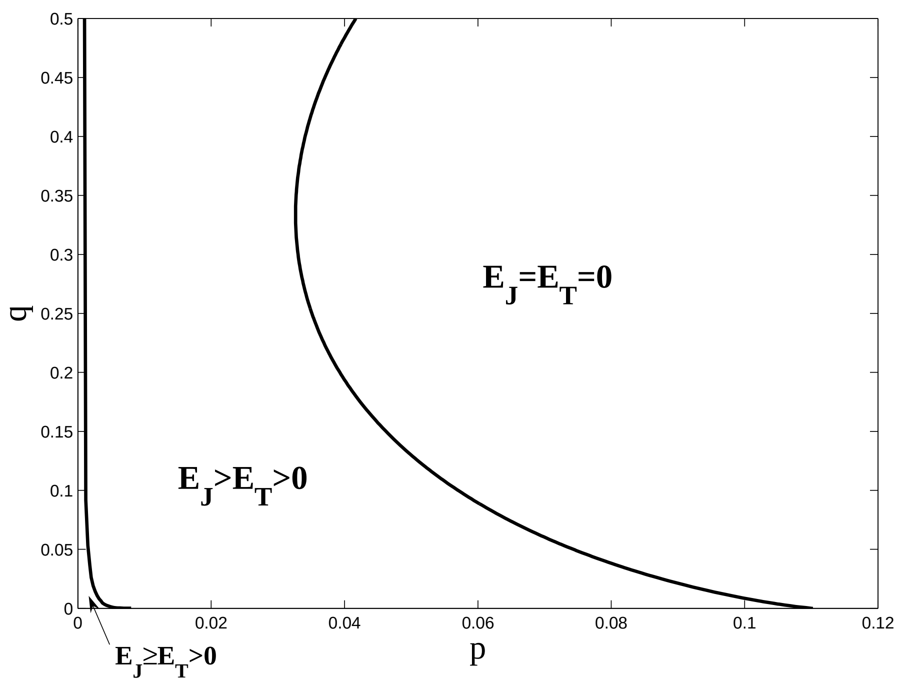

Figure 10.7: The regions for the ternary SEM source and the binary SEM channel of Example 10.3 with $\tau = 0.5$.

Example 10.3 We next examine Theorems 10.7 and 10.8 for the following simple example. Consider a ternary SEM source $\mathbf{Q_S}$ and a binary SEM channel $\mathbf{W_{Y|X}}$, both with symmetric transition matrices given by

$$Q_S = \begin{bmatrix} q & (1-q)/2 & (1-q)/2 \\ (1-q)/2 & q & (1-q)/2 \\ (1-q)/2 & (1-q)/2 & q \end{bmatrix} \quad \text{and} \quad P_Z = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

such that $0 < p, q < 0.5$. Suppose now the transmission rate $\tau = 0.5$. If (q, p) satisfies any one of the conditions of Theorems 10.7 and 10.8, then $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) >$

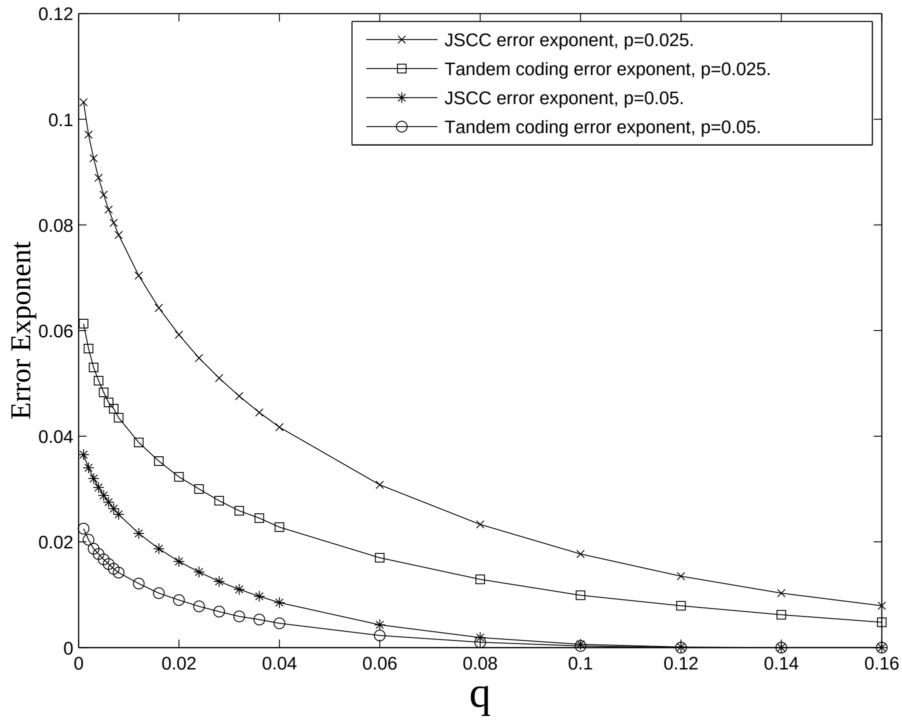

Figure 10.8: Comparison of E_J and E_T for the ternary SEM source and the binary SEM channel of Example 10.3 with $\tau = 0.5$.

$E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$. The range for which the inequality holds is summarized in Fig. 10.7. For the channel with $p = 0.025$ and $p = 0.05$, we plot the JSC coding and tandem coding error exponents against the source parameter q whenever they are exactly determined, see Fig. 10.8. We note that for these source-channel pairs, $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ substantially outperforms $E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$ (indeed $E_J(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau) \approx 2E_T(\mathbf{Q_S}, \mathbf{W_{Y|X}}, \tau)$) for a large class of (q, p) pairs. We then plot the two exponents under the transmission rate $\tau = 0.75$ whenever they are determined exactly, and obtain similar results, see Fig. 10.9. In fact, for many other SEM source-channel pairs (not necessarily binary SEM sources or ternary SEM channels) with other transmission rates, we observe similar results; this indicates that the JSC coding exponent is strictly better than the tandem coding exponent for a wide class of SEM systems.

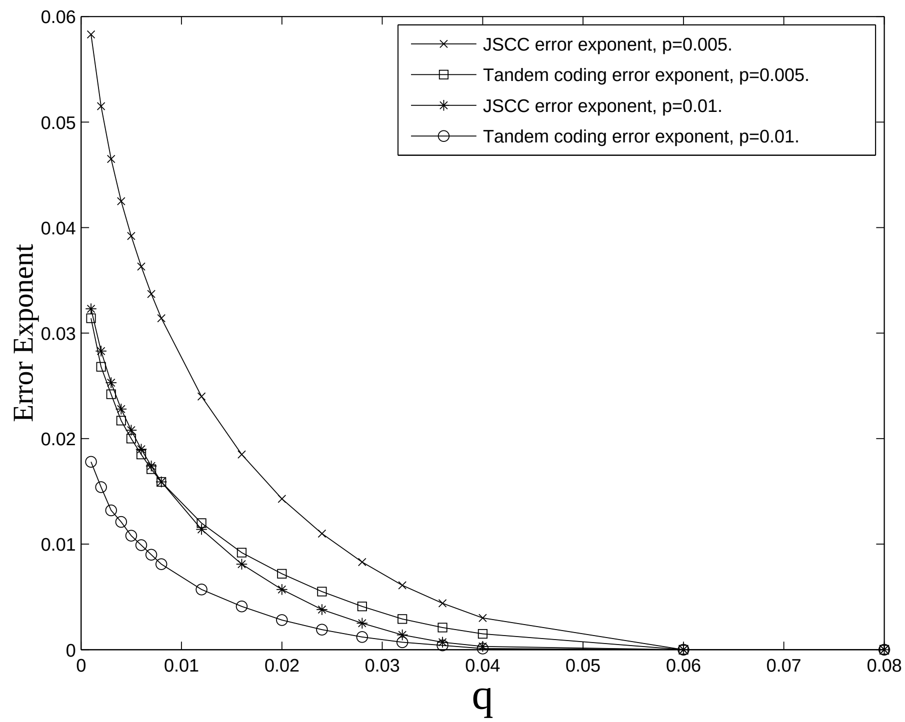

Figure 10.9: Comparison of E_J and E_T for the ternary SEM source and the binary SEM channel of Example 10.3 with $\tau = 0.75$.

In the following we present an example for the system consisting of an SEM source (of order $K_s = 1$) and the queue based channel (QBC) [100] with memory $K_c = 2$, as the QBC approximates well for a certain range of channel conditions the Gilbert-Elliott channel [100] and hard decision demodulated correlated fading channels [101].

Example 10.4 (Transmission of an SEM source over the QBC) A QBC is a binary additive channel whose noise process $\mathbf{P_Z} = \{P_{Z^n} \in \mathcal{P}(\mathcal{Z}^n)\}_{n=1}^{\infty}$ (where $\mathcal{Z} = \{0, 1\}$) is generated according to a mixture mechanism of a finite queue and a Bernoulli process [101]. At time i, the noise symbol Z_i is chosen either from the queue described by a sequence of random variables $(Q_{i,1}, ..., Q_{i,K_c})$ $(Q_{i,j} \in \{0, 1\},\ j = 1, 2, ..., K_c)$ with probability ε or from a Bernoulli process with probability $1 - \varepsilon$ such that

- If Z_i is chosen from the queue process, then

$$P_Z(Z_i = Q_{i,j}) = \begin{cases} 1/(K_c - 1 + \alpha), & j = 1, 2, ..., K_c - 1, \\ \alpha/(K_c - 1 + \alpha), & j = K_c \end{cases}$$

  if $K_c > 1$ and $\alpha \geq 0$ is arbitrary; otherwise $P_Z(Z_i = Q_{i,K_c}) = 1$ if $K_c = 1$.

- If Z_i is chosen from the Bernoulli process, then $P_Z(Z_i = 1) = p$ ($p \ll 1/2$) and $P_Z(Z_i = 0) = 1 - p$.

At time $i + 1$, we first shift the queue from left to right by the following rule

$$(Q_{i+1,1}, ..., Q_{i+1,K_c}) = (Z_i, Q_{i,1}, ..., Q_{i,K_c-1}),$$

then we generate the noise symbol Z_{i+1} according to the same mechanism. It can be shown [101] that the QBC is actually an K_c-th order SEM channel characterized only by four parameters ε, α, p and K_c.

Now we consider transmitting the first order SEM source $\mathbf{Q_S}$ with transition matrix

$$Q_S = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.05 & 0.15 & 0.8 \end{bmatrix}$$

under transmission rate $t = 1$ over the QBC with $K_c = 2$ such that the noise process $\mathbf{P_Z}$ is a second order SEM process. After 2-step blocking $\mathbf{P_Z}$, we obtain a first order SEM process $\mathbf{P_Z^{K_c}}$ with transition matrix

$$P_Z^{K_c} = \begin{bmatrix} \varepsilon + (1-\varepsilon)(1-p) & 0 & (1-\varepsilon)p & 0 \\ \frac{\varepsilon}{1+\alpha} + (1-\varepsilon)(1-p) & 0 & \frac{\varepsilon\alpha}{1+\alpha} + (1-\varepsilon)p & 0 \\ 0 & \frac{\varepsilon\alpha}{1+\alpha} + (1-\varepsilon)(1-p) & 0 & \frac{\varepsilon}{1+\alpha} + (1-\varepsilon)p \\ 0 & (1-\varepsilon)(1-p) & 0 & \varepsilon + (1-\varepsilon)p \end{bmatrix}.$$

We next compute E_J and E_T for the ternary SEM source and the QBC given above. When $p = 0.05$, $\alpha = 1$, E_J and E_T are both determined exactly if $\varepsilon \in [0.001, 0.992]$. We plot the two exponents by varying ε. We see from Fig. 10.10 that $E_J \approx 2E_T$ for all the $\varepsilon \in [0.001, 0.992]$. When we choose $p = 0.05$, $\alpha = 0.1$ for which E_J and E_T

are both determined exactly if $\varepsilon \in [0.001, 0.968]$, we have similar results, see Fig. 10.10. It is interesting to note that when ε gets smaller, E_J and E_T approach the exponents resulting from the SEM source $\mathbf{Q_S}$ and the binary symmetric channel (BSC) with crossover probability $p = 0.05$. This is indeed expected since the QBC reduces to the BSC when $\varepsilon = 0$ [101].

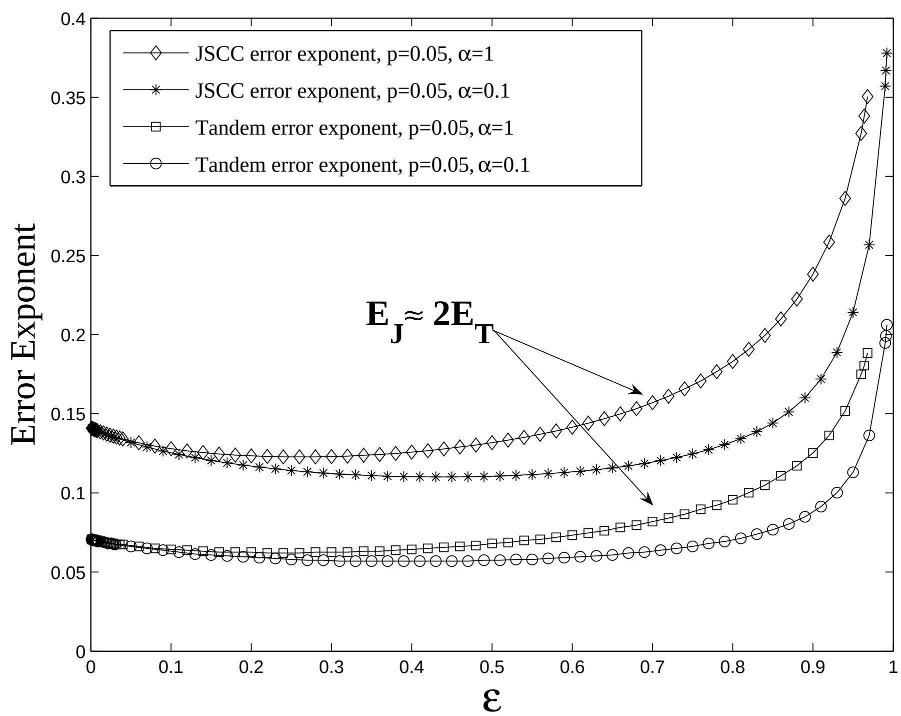

Figure 10.10: Comparison of E_J and E_T for the SEM source and the QBC of Example 10.4 with $\tau = 1$.

10.4 Tandem Error Exponent with Feedback/Source Side Information

We have obtained the formula for tandem error exponent for discrete systems, which is expressed in terms of the source and channel error exponents. In fact, it can be seen that the formula for tandem error exponent is still valid for discrete memoryless systems with channel output feedback or source SI.

10.4.1 Tandem Exponent with Perfect Feedback

A tandem code $(f^*_{n,fb}, \varphi^*_{n,fb}) \triangleq (f_{sn}, \{f_{cr}\}_{r=1}^n, \varphi_{cn}, \varphi_{sn})$ for a DMS Q_S and a DMC $W_{Y|X}$ with perfect channel output feedback is composed of two separately designed codes: a $(\tau n, M_n)$ block source code (f_{sn}, φ_{sn}) with source code rate

$$R_{s,n} \triangleq \frac{\log_2 M_n}{\tau n} \quad \text{source code bits/source symbol,}$$

and an (n, M_n) block channel code with perfect feedback $(\{f_{cr}\}_{r=1}^n, \varphi_{cn})$ with channel code rate

$$R_{c,n} \triangleq \frac{\log_2 M_n}{n} \quad \text{source code bits/channel use,}$$

assuming that the limit $\lim_{n\to\infty} \log M_n/n$ exists, i.e.,

$$\limsup_{n\to\infty} \frac{\log M_n}{n} = \liminf_{n\to\infty} \frac{\log M_n}{n}.$$

To render the source and channel coding operations independent of each other, as in Section 10.1, a random index assignment π_m is performed between the source and channel encoders/decoders, and each π_m is independently and equally likely chosen from the $M_n!$ different possible index assignments; see Fig. 10.11. Also, we assume that (A1) still holds for the source code, i.e., for every n, $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) > 0$ and $\mathbf{c}_i \in f_{sn}^{-1}(i)$ for every $i = 1, 2, ..., M_n$, where $f_{sn}^{-1}(i) \triangleq \{\mathbf{s} \in \mathcal{S}^{\tau n} : f_{sn}(\mathbf{s}) = i\}$.

Similarly, the error probability of the tandem code $(f^*_{n,fb}, \varphi^*_{n,fb})$ is given by

$$\begin{aligned} & P^{(n)}_{e^*,fb}(Q_S, W_{Y|X}, \tau) \\ &= P^{(n)}_{ecfb}(W_{Y|X}, R_{c,n}) + (1 - P^{(n)}_{ecfb}(W_{Y|X}, R_{c,n}))P^{(\tau n)}_{es}(Q_S, R_{s,n}), \end{aligned}$$

where $P^{(n)}_{ecfb}(W_{Y|X}, R_{c,n})$ is the channel coding probability of error with feedback given by (6.1), and $P^{(\tau n)}_{es}(Q_S, R_{s,n})$ is the probability of error for DMS given by (2.1).

Definition 10.2 The tandem coding error exponent $E_{T,fb}(Q_S, W_{Y|X}, \tau)$ for DMS Q_S and DMC $W_{Y|X}$ with feedback is defined as the supremum of the set of all numbers $\widehat{E}$ for which

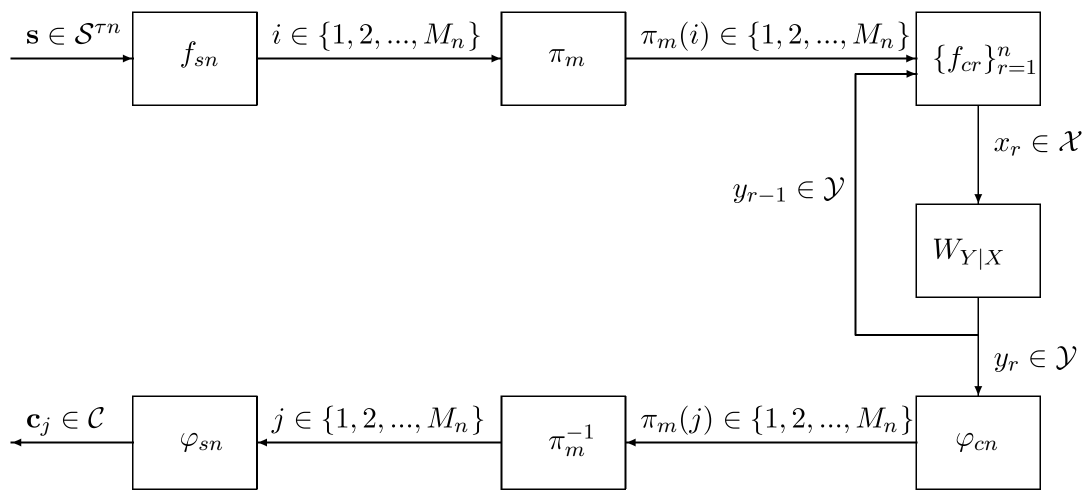

Figure 10.11: Tandem coding system for discrete memoryless source-channel systems with perfect feedback.

there exists a sequence of tandem codes $(f^*_{n,fb}, \varphi^*_{n,fb})$ satisfying (A1) with transmission rate τ such that

$$\widehat{E} \le \liminf_{n\to\infty} -\frac{1}{n}\log_2 P^{(n)}_{e^*,fb}(Q_S, W_{Y|X}, \tau).$$

We have the following similar results.

Theorem 10.9

$$E_{T,fb}(Q_S, W_{Y|X}, \tau) = \sup_{R>0} \min\left\{\tau e\left(\frac{R}{\tau}, Q_S\right), E_{fb}(R, W_{Y|X})\right\}$$

where $e(R, Q_S)$ is the source error exponent defined in (7.2) and $E_{fb}(R, W_{Y|X})$ is the channel error exponent with feedback (cf. Definition 6.1).

Remark 10.6 Recalling that the sphere-packing exponent $E_{sp}(R, W_{Y|X})$ is an upper bound for $E_{fb}(R, W_{Y|X})$ (see Section 6.1.1), we can upper bound $E_{T,fb}(Q_S, W_{Y|X}, \tau)$ by

$$E_{T,fb}(Q_S, W_{Y|X}, \tau) \le \sup_{R>0} \min\left\{\tau e\left(\frac{R}{\tau}, Q_S\right), E_{sp}(R, W_{Y|X})\right\}. \tag{10.35}$$

Clearly, if $\tau H_{Q_S}(S) \ge C(W_{Y|X})$, $E_{T,fb}(Q_S, W_{Y|X}, \tau) = 0$.

Theorem 10.10 $E_{T,fb}(Q_S, W_{Y|X}, \tau) \leq E_{J,fb}(Q_S, W_{Y|X}, \tau) \leq 2E_{T,fb}(Q_S, W_{Y|X}, \tau)$.

The proofs of Theorems 10.9 and 10.10 are similar to the proofs of Theorems 10.1 and 10.2 and are omitted.

In the following, we present an example to illustrate the gain of JSCC exponent over tandem exponent by numerically comparing the lower bound of $E_{J,fb}$ and the upper bound of $E_{T,fb}$ for binary DMS $Q = \{q, 1-q\}$ ($q < 0.5$) and BSC $W_{Y|X}$ with crossover probability ϵ ($\epsilon < 0.5$). Here we use the lower bound (6.9) for $E_{J,fb}$ and the upper bound (10.35) for $E_{T,fb}$. Since we know that the lower bound (6.9) is at least as large as Gallager's lower bound for E_J (without feedback) for binary input channels with a symmetric distribution (in the Gallager sense, cf. Corollary 6.3), it is hoped that the lower bound is larger than the upper bound (10.35) for E_T for a lot of source and channel parameters. As expected, for fixed q and τ, we see from Fig. 10.12 that the joint exponent almost doubles the tandem exponent for a wide range of ϵ. Note that similar results hold for other DMS's and binary channels with a symmetric distribution.

10.4.2 Tandem Exponent with Source Side Information

In the section we address the case of source SI at the decoder, since in Section 6.3 we have derived a valid computable lower bound for $E_J^{SID}(Q_{SL}, W_{Y|X}, \tau)$ and we could use it to evaluate the gain of JSCC error exponent over the tandem coding error exponent.

A tandem code $(f^*_{n,SID}, \varphi^*_{n,SID}) \triangleq (f_{sn}, f_{cn}, \varphi_{cn}, \varphi_{sn})$ for a DMS Q_S and a DMC $W_{Y|X}$ with source SI Q_L at the decoder (which is correlated with Q_S through $Q_{L|S}$) is composed of separately designed source and channel codes: a $(\tau n, M_n)$ block source code $(f_{sn,SID}, \varphi_{sn,SID})$ with source code rate

$$R_{s,n} \triangleq \frac{\log_2 M_n}{\tau n} \quad \text{source code bits/source symbol,}$$

where

$$f_{sn,SID} : \mathcal{S}^{tn} \to \{1, 2, ..., M_n\}$$

and

$$\varphi_{sn,SID} : \{1, 2, ..., M_n\} \times \mathcal{L}^{\tau n} \to \mathcal{S}^{tn}$$

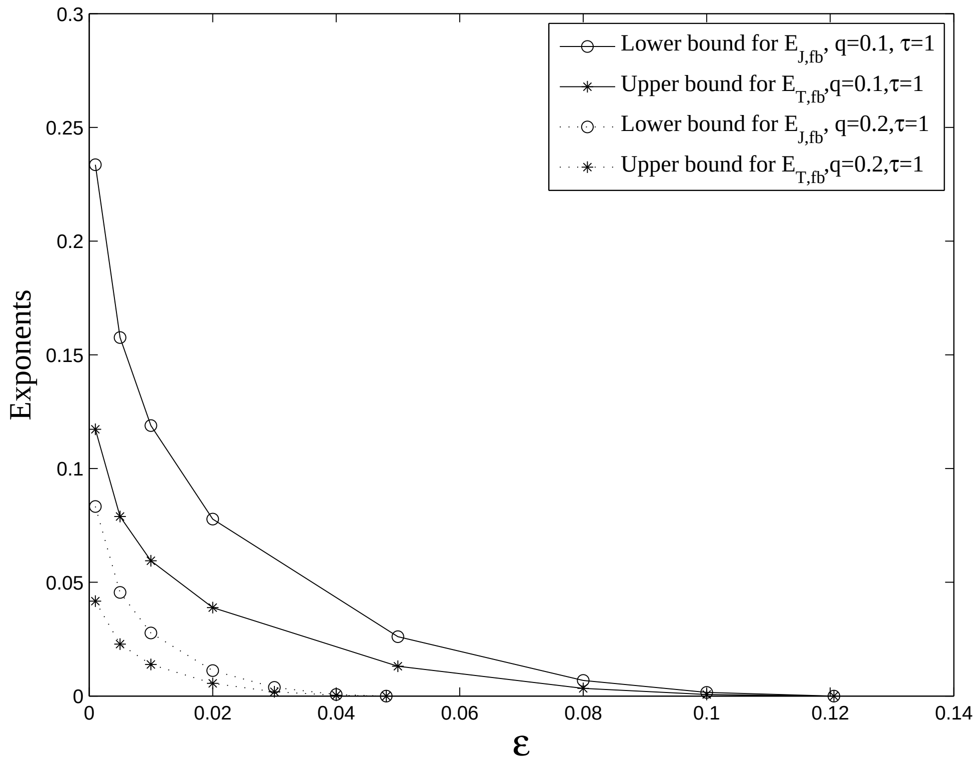

Figure 10.12: The lower bound of $E_{J,fb}$ vs the upper bound of $E_{T,fb}$.

are the source encoder and decoder with source SI, and an (n, M_n) block channel code (f_{cn}, φ_{cn}) with channel code rate

$$R_{c,n} \triangleq \frac{\log_2 M_n}{n} \quad \text{source code bits/channel use,}$$

assuming that the limit $\lim_{n\to\infty} \log M_n/n$ exists, i.e.,

$$\limsup_{n\to\infty} \frac{\log M_n}{n} = \liminf_{n\to\infty} \frac{\log M_n}{n}.$$

Again, as in Section 10.1, the source and channel operations are statistically decoupled via a random index assignment π_m between the source and channel encoders/decoders. Each π_m is independently and equally likely chosen from the $M_n!$ different possible index assignments; see Fig. 10.13. In addition, we assume (A2) that for every n, $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) > 0$ and

$$\bigcup_{\mathbf{l}\in\mathcal{L}^{\tau n}} \{\varphi_{sn,SID}(i, \mathbf{l})\} \subseteq f_{sn}^{-1}(i)$$

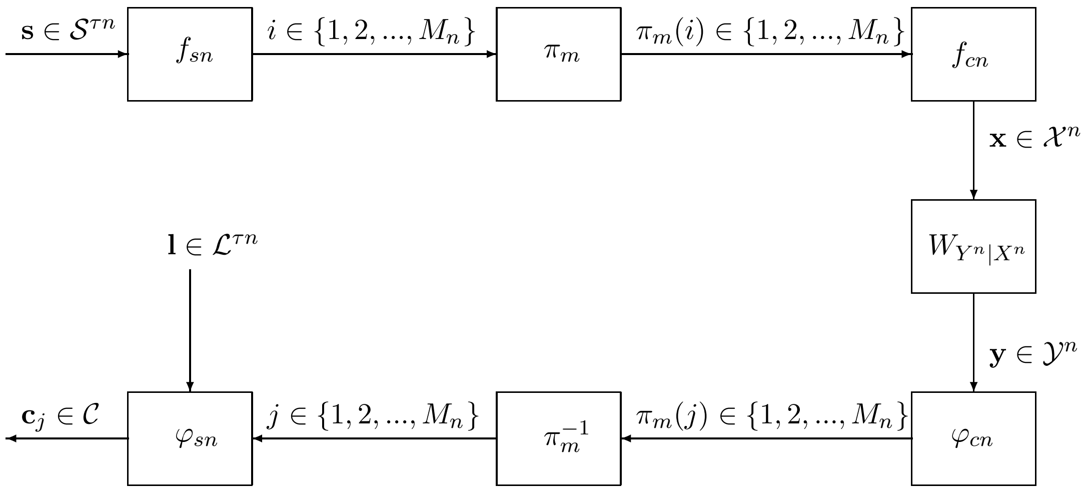

Figure 10.13: Tandem coding system for discrete memoryless source-channel systems with source SI at the decoder.

for every $i = 1, 2, ..., M_n$, where $f_{sn}^{-1}(i) \triangleq \{\mathbf{s} \in \mathcal{S}^{\tau n} : f_{sn}(\mathbf{s}) = i\}$. Like assumption (A1), the assumption (A2) has the same practical meaning. If $Q_{S^{\tau n}}(f_{sn}^{-1}(i)) = 0$ for some i, then the output codeword is redundant, and we can remove it from the codebook. If there exists one $\varphi_{sn,SID}(i, \mathbf{l}) \notin f_{sn}^{-1}(i)$, we can map the index i and the SI sequence $\mathbf{l}$ to some source message $\widehat{\mathbf{s}}$ such that $Q_{S^{\tau n}}(\widehat{\mathbf{s}}) > 0$ and $f_{sn}(\widehat{\mathbf{s}}) = i$, so that the source coding probability of error

$$P_{es,SID}^{(\tau n)}(Q_{SL}, R_{s,n}) = \sum_{(\mathbf{s},\mathbf{l}) \in \mathcal{S}^{\tau n} \times \mathcal{L}^{\tau n} : \varphi_{sn,SID}(f_{sn,SID}(\mathbf{s}),\mathbf{l}) \neq \mathbf{s}} Q_{SL}^{(\tau n)}(\mathbf{s}, \mathbf{l}). \tag{10.36}$$

is strictly reduced.

Similarly, the error probability of the tandem code $(f_{n,SID}^*, \varphi_{n,SID}^*)$ is given by

$$\begin{aligned} & P_{e^*,SID}^{(n)}(Q_{SL}, W_{Y|X}, \tau) \\ &= P_{ec}^{(n)}(W_{Y|X}, R_{c,n}) + (1 - P_{ec}^{(n)}(W_{Y|X}, R_{c,n})) P_{es,SID}^{(\tau n)}(Q_{SL}, R_{s,n}), \end{aligned}$$

where $P_{ec}^{(n)}(W_{Y|X}, R_{c,n})$ is the channel coding probability of error (10.6), and $P_{es,SID}^{(\tau n)}(Q_{SL}, R_{s,n})$ is the probability of error for DMS with source SI at the decoder given by (10.36).

Definition 10.3 For any $R > 0$, the source error exponent $e_{SID}(R, Q_{SL})$ of the DMS Q_S

with source SI Q_L at the decoder is defined as the supremum of the set of all numbers e for which there exists a sequence of (n, M_n) block codes $(f_{sn,SID}, \varphi_{sn,SID})$ with

$$e \leq \liminf_{n\to\infty} -\frac{1}{n} \log P_{se}^{(n)}(Q_S, R_n) \tag{10.37}$$

and

$$R \geq \limsup_{n\to\infty} R_n, \tag{10.38}$$

where the source code rate $R_n = \frac{1}{n}\log_2 M_n$.

Definition 10.4 The tandem coding error exponent $E_T^{SID}(Q_{SL}, W_{Y|X}, \tau)$ for DMS Q_S and DMC $W_{Y|X}$ with source SI Q_L at the decoder is defined as the supremum of the set of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes $(f^*_{n,SID}, \varphi^*_{n,SID})$ satisfying (A1) with transmission rate τ such that

$$\widehat{E} \leq \liminf_{n\to\infty} -\frac{1}{n} \log_2 P_{e^*,SID}^{(n)}(Q_{SL}, W_{Y|X}, \tau).$$

We have the following similar results.

Proposition 10.2 $E_T^{SID}(Q_S, W_{Y|X}, \tau) \leq E_J^{SID}(Q_S, W_{Y|X}, \tau)$.

Theorem 10.11

$$E_T^{SID}(Q_S, W_{Y|X}, \tau) = \sup_{R>0} \min\left\{\tau e_{SID}\left(\frac{R}{\tau}, Q_{SL}\right), E(R, W_{Y|X})\right\}$$

where $e_{SID}(R, Q_{SL})$ *is the source error exponent with source SI at the decoder and* $E(R, W_{Y|X})$ *is the channel error exponent (cf. Definition 6.1).*

Remark 10.7 If $\tau H_{Q_{SL}}(S|L) \geq C(W_{Y|X})$, $E_{T,SID}(Q_S, W_{Y|X}, \tau) = 0$.

According to [32], for source coding with source SI at the decoder, we have a somewhat loose[2] upper bound for $e_{SID}(R, Q_{SL})$,

$$e_{SID}(R, Q_{SL}) \leq \overline{e}_{SID}(R, Q_{SL}) \triangleq \min_{P_{SL}: H_{P_{SL}}(S|L)\geq R} D(P_{SL} \parallel Q_{SL}).$$

[2]In fact, this upper bound is obtained by assuming that the source SI is available at both the encoder and the decoder; thus, the upper bound is actually an upper bound for the source error exponent with source SI at both the encoder and the decoder.

It is easy to verify that the minimum of the above is achieved by a tilted distribution

$$P^*_{SL}(s,l) = Q_S(s)\frac{Q_{L|S}(l|s)^{\frac{1}{1+\rho^*}}}{\sum_{l'\in\mathcal{L}} Q_{L|S}(l'|s)^{\frac{1}{1+\rho^*}}}, \quad \forall s \in \mathcal{S}, \quad l \in \mathcal{L},$$

where ρ^* is the root of $H_{P_{SL}}(L|S) = R$. Plugging $P^*_{SL}(s,l)$ into $D(P_{SL}\|Q_{SL})$ we obtain the parametric form

$$\min_{P_{SL}:H_{P_{SL}}(S|L)=R} D(P_{SL}\|Q_{SL}) = \max_{\rho\geq 0}\left[\rho R - E_{s2}(\rho, Q_{SL})\right] \tag{10.39}$$

where

$$E_{s2}(\rho, Q_{SL}) = (1+\rho)\sum_{l\in\mathcal{L}} Q_L(l)\log_2 \sum_{s\in\mathcal{S}} Q_{S|L}(s|l)^{\frac{1}{1+\rho}}.$$

Now, replacing the source and channel error exponents by their upper bounds, we obtain a computable upper bound for $E_T^{SID}(Q_S, W_{Y|X}, \tau)$

$$E_T^{SID}(Q_S, W_{Y|X}, \tau) \leq \sup_{R>0} \min\left\{\tau\overline{e}_{SID}\left(\frac{R}{\tau}, Q_{SL}\right), E_{sp}(R, W_{Y|X})\right\}. \tag{10.40}$$

In the following, we present an example to see the gain of JSCC exponent over tandem exponent by numerically comparing the lower bound of E_J^{SID} and the upper bound of E_T^{SID} for binary DMS $Q = \{q, 1-q\}$ ($q < 0.5$) and BSC $W_{Y|X}$ with crossover probability ϵ ($\epsilon < 0.5$). The source Q_L is a noisy version of Q_S described by $L = S \oplus N \mod 2$ ($\mathcal{L} = \mathcal{N} = \{0,1\}$) with noise distribution $P_N(N=1) = 0.05$, i.e., the SI is transmitted through a dummy BSC $Q_{L|S}$ with crossover probability 0.05. Here we use the lower bound $\underline{E}_J^{SID}$ defined in (6.34) and the upper bound for E_T^{SID} given above. It is seen from Fig. 10.14 that for fixed q and transmission rate τ, $\underline{E}_J^{SID}$ substantially outperforms the upper bound for E_T^{SID}. We remark that similar results hold for other DMS and DMC pairs.

10.5 Memoryless Gaussian Source-Channel Systems

In this section we study the advantage of JSCC over tandem coding in terms of the excess distortion exponent for Gaussian systems. A tandem code

$$(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \triangleq (f_{sn}, f_{cn}, \varphi_{cn}, \varphi_{cn}, \Delta, \mathcal{E}, t, P)$$

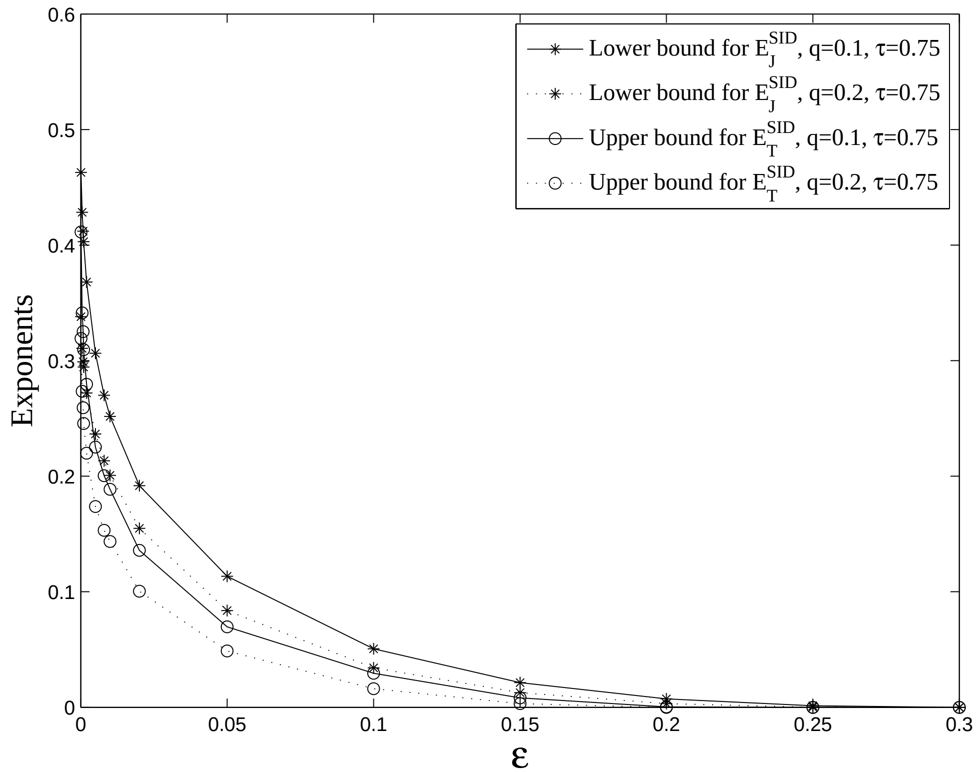

Figure 10.14: The lower bound for E_J^{SID} vs the upper bound for E_T^{SID}.

with blocklength n and transmission rate τ (source symbols/channel use) for the MGS Q_S and the MGC $W_{Y|X}$ is composed (see Fig. 10.15) of two separately designed codes: a $(\tau n, M_n)$ block source code $(f_{sn}, \varphi_{sn}, \Delta)$ with codebook $\mathcal{C} \triangleq \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{M_n}\} \subseteq \mathcal{S}^{\tau n}$ and source code rate

$$R_{s,n} = \frac{\ln M_n}{\tau n} \quad \text{source code nats/source symbol,}$$

$M_n!$ index assignments (which are permutation functions) π_m, such that each π_m is chosen with probability $1/M_n!$, $m = 1, 2, ..., M_n!$, and an (n, M_n) block channel code $(f_{cn}, \varphi_{cn}, \mathcal{E})$ with channel code rate

$$R_{c,n} = \frac{\ln M_n}{n} \quad \text{source code nats/channel use,}$$

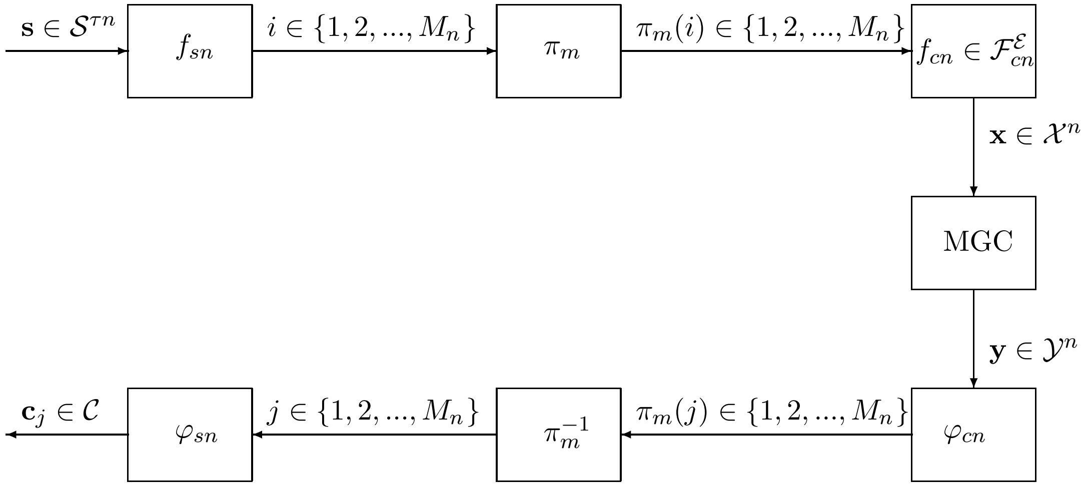

Figure 10.15: Tandem MGS-MGC system.

where $f_{cn} \in \mathcal{F}_{cn}^{\mathcal{E}}$ with $g(x) = x^2$. We assume as before that the limit $\lim_{n\to\infty} \frac{\ln M_n}{n}$ exists, i.e.,

$$\limsup_{n\to\infty} \frac{\ln M_n}{n} = \liminf_{n\to\infty} \frac{\ln M_n}{n}.$$

The (overall) excess distortion probability of the tandem code $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau)$ is hence given by

$$\begin{aligned}
& P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\
\triangleq\ & \Pr\left(d^{(\tau n)}\left(\mathbf{s}, \varphi_{sn}\left\{\pi_m^{-1}[\varphi_{cn}(\mathbf{y})]\right\}\right) > \Delta\right) \\
=\ & \sum_{m=1}^{M_n!} \frac{1}{M_n!} \Pr\left(d^{(\tau n)}\left(\mathbf{s}, \varphi_{sn}\left\{\pi_m^{-1}[\varphi_{cn}(\mathbf{y})]\right\}\right) > \Delta \middle| \pi_m\right) \\
=\ & \sum_{m=1}^{M_n!} \frac{1}{M_n!} \int_{\mathcal{S}^{\tau n}} Q_S^{(\tau n)}(\mathbf{s}) \int_{\mathbf{y}: d^{(\tau n)}\left(\mathbf{s}, \varphi_{sn}\left\{\pi_m^{-1}[\varphi_{cn}(\mathbf{y})]\right\}\right) > \Delta} W_{Y|X}^{(n)}\left(\mathbf{y} \mid f_{cn}\left\{\pi_m[f_{sn}(\mathbf{s})]\right\}\right) d\mathbf{y} d\mathbf{s}, .
\end{aligned}$$

Recall that the codebook of the source code $(f_{sn}, \varphi_{sn}, \Delta)$ is $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{M_n}\}$. To make the notations simpler, we denote (cf. Fig. 10.15)

$$\begin{aligned}
i &= f_{sn}(\mathbf{s}), \\
j &= \pi_m^{-1}(\varphi_{cn}(\mathbf{y})), \\
D_i &= \{\mathbf{s} \in \mathcal{S}^n : f_{sn}(\mathbf{s}) = i\},
\end{aligned}$$

for $i, j \in \{1, 2, ..., M_n\}$, where the (disjoint) sets $D_1, D_2, ..., D_{M_n}$ partition $\mathcal{S}^{\tau n}$. Thus, we can write

$$\begin{aligned} & P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\ &= \sum_{m=1}^{M_n!} \frac{1}{M_n!} \sum_{i=1}^{M_n} \sum_{j=1}^{M_n} P_W(\pi_m(j)|\pi_m(i)) \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \mathbb{1}\left\{ d^{(\tau n)}(\mathbf{s}, \mathbf{c}_j) > \Delta \right\} d\mathbf{s}, \quad (10.41) \end{aligned}$$

and

$$P_W(\pi_m(j)|\pi_m(i)) \triangleq \int_{\mathbf{y}:\varphi_{cn}(\mathbf{y})=\pi_m(j)} W_{Y|X}^{(n)}(\mathbf{y}|f_{cn}(\pi_m(i)))d\mathbf{y}.$$

The excess distortion probability (2.10) for the source code can also be rewritten by

$$P_{\Delta}^{(n)}(Q_S, R_{s,n}) = \sum_{i=1}^{M_n} \int_{D_i} P_{S^{\tau n}}(\mathbf{s}) \mathbb{1}\left\{ d^{(\tau n)}(\mathbf{s}, \mathbf{c}_i) > \Delta \right\} d\mathbf{s},$$

and the probability of error (2.12) for the channel code can be written by

$$P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}) = \frac{1}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} P_W(j|i).$$

Meanwhile, the maximal probability of error (2.13) for the channel code is given by

$$P_{max,ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}) = \max_{1\leq i\leq M_n} \sum_{j=1, j\neq i}^{M_n} P_W(j|i).$$

In order to facilitate the evaluation of the tandem excess distortion probability $P_{\Delta^*}^{(n)}$, we simplify the problem by making the following assumptions on the channel code and the source code.

(a) Since channel coding is performed independently and separately from source coding and the index assignment, our goal in channel coding is to make $P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})$ as small as possible. Defining

$$\Xi(W_{Y|X}, \mathcal{E}) \triangleq \left\{ (f_{cn}, \varphi_{cn}, \mathcal{E}) : \limsup_{n\to\infty} P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}) < \gamma \quad \text{for all } \gamma > 0 \right\},$$

we say that a sequence of channel codes $(f_{cn}, \varphi_{cn}, \mathcal{E})$ is a sequence of "good channel codes" if $(f_{cn}, \varphi_{cn}, \mathcal{E}) \in \Xi(W_{Y|X}, \mathcal{E})$. In the tandem system, we will restrict $(f_{cn}, \varphi_{cn}, \mathcal{E})$ to be good channel codes.

(b) In source coding, the objective is to construct a code, or equivalently, find a codebook $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{M_n}\}$ and the corresponding partition $\{D_1, D_2, ..., D_{M_n}\}$ so that $P_\Delta^{(n)}(Q_S, R_{s,n})$ is as small as possible. For the source codes $(f_{sn}, \varphi_{sn}, \Delta)$, we assume the following constraint. Letting

$$\Omega(Q_S, \Delta) \triangleq \Big\{(f_{sn}, \varphi_{sn}, \Delta) \quad : \quad \liminf_{n\to\infty} -\frac{1}{\tau n} \ln P_\Delta^{(n)}(Q_S, R_{s,n}) \geq F(R, Q_S, \Delta) > 0, \\ \text{where } R = \lim_{n\to\infty} R_{s,n}\Big\},$$

we say that a sequence of source codes $(f_{sn}, \varphi_{sn}, \Delta)$ is a sequence of "good source codes" if $(f_{sn}, \varphi_{sn}, \Delta) \in \Omega(Q_S, \Delta)$. In the tandem system, we will only consider such good source codes.

Recall that the converse JSCC theorem (Theorem 8.1) states that the MGS cannot be reliably transmitted over the MGC if $\tau R(Q_S, \Delta) > C(W_{Y|X}, \mathcal{E})$, and also note that if $\tau R(Q_S, \Delta) > C(W_{Y|X}, \mathcal{E})$ then either $\Xi(W_{Y|X}, \mathcal{E}) = \emptyset$ or $\Omega(Q_S, \Delta) = \emptyset$. Thus, we are only interested in the case $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$ as before. In order to guarantee the existence of good source and channel codes, we focus on the sequences of tandem codes with $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$, where

$$\Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau) \triangleq \left\{(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) : \tau R(Q_S, \Delta) < \lim_{n\to\infty} \frac{\ln M_n}{n} < C(W_{Y|X}, \mathcal{E})\right\}.$$

Assumptions (1) and (2) are needed for the proof of the converse part of Theorem 10.12.

Definition 10.5 The tandem coding excess distortion exponent $E_T^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$ for the MGS Q_S and the MGC $W_{Y|X}$ is defined as the supremum of the set of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau)$ composed by good source and channel codes with blocklength n provided $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$, such that

$$\widehat{E} \leq \liminf_{n\to\infty} -\frac{1}{n} \ln P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau).$$

When there is no possibility of confusion, throughout the sequel, the tandem coding excess distortion exponent $E_T^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$ will be written as $E_T^{\Delta,\mathcal{E}}$.

Theorem 10.12 *For the tandem MGS-MGC system provided $\tau R(Q_S, \Delta) < C(W_{Y|X}, \mathcal{E})$ and SDR ≥ 4,*

$$E_T^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) = \sup_{\tau R(Q_S,\Delta) < R < C(W_{Y|X},\mathcal{E})} \min \left\{ \tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right), E(R, W_{Y|X}, \mathcal{E}) \right\}$$

where $F_G(R, Q_S, \Delta)$ is the MGS excess distortion exponent given by (2.46) and (2.47) and $E(R, W_{Y|X}, \mathcal{E})$ is the MGC error exponent defined by Definition 2.4.

Remark 10.8 Since $\tau F_G(R/\tau, Q_S, \Delta)$ is a strictly increasing function of R for $R \geq 0$, and $E(R, W_{Y|X}, \mathcal{E})$ is decreasing function of R for $0 < R \leq C(W_{Y|X}, \mathcal{E})$, the supremum must be achieved at their intersection[3]

$$E_T^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) = \tau F_G\left(\frac{R_o}{\tau}, Q_S, \Delta\right) = E(R_o, W_{Y|X}, \mathcal{E}),$$

with $\tau R(Q_S, \Delta) < R_o < C(W_{Y|X}, \mathcal{E})$.

We require that the distortion threshold cannot be too large; we restrict SDR ≥ 4 ($\approx$ 6dB). As will be seen in the proof of the converse part of Theorem 10.12, this assumption ensures that the ball $B(\mathbf{0}, 4\Delta)$ is covered by $o(M_n)$ balls with size Δ; see Lemma 10.1 in the below. In fact, a large distortion threshold is useless in practice.

Lemma 10.1 *Let SDR $= \frac{\sigma_S^2}{\Delta} > 4$. Only $L_{2n} = o(M_n)$ balls of size Δ are needed to cover $B(\boldsymbol{0}, 4\Delta)$ for $R > R(Q_S, \Delta)$, i.e., every sequence in $B(\boldsymbol{0}, 4\Delta)$ is contained in the union of L_{2n} balls of size Δ.*

Proof: Let $k = tn$. For $N \in \mathbb{N}$ which will be specified later, we partition $B(\mathbf{0}, 4\Delta)$ by a sequence of sets: $\mathcal{T}_0 \triangleq \{\mathbf{s} : \|\mathbf{s}\|^2 = \mathbf{s}^T\mathbf{s} \leq k\Delta\}$ and $\mathcal{T}_i \triangleq \mathcal{T}^\epsilon(\sigma^2(i))$ by $\sigma^2(i) = \Delta + (2i-1)\epsilon$, where $\epsilon = \frac{3\Delta}{2N}$, for $i = 1, 2, \cdots, N$, i.e.,

$$\mathcal{T}_i = \left\{\mathbf{s} : k\left[\Delta + (2i-2)\epsilon\right] \leq \mathbf{s}^T\mathbf{s} \leq k\left[\Delta + 2i\epsilon\right]\right\}, \qquad i = 1, 2, \cdots, .N.$$

[3]Unlike the discrete case in [107] and Section 10.2, the intersection always exists since source exponent is continuous and increasing in $R > 0$.

Note that $\mathcal{T}_0$ is covered by one ball with size Δ, $B(\mathbf{0}, \Delta)$. It follows from the type covering lemma for Gaussian-type classes (Lemma 3.6) that each $\mathcal{T}_i$ $(1 \le i \le N)$ is covered by

$$L(i) = \exp\left\{k\left[R\left(Q_S^{(i)}, \Delta\right) + \widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right)\right] + o(k)\right\}$$

balls with size Δ for k and N sufficiently large, where $Q_S^{(i)} \sim \mathcal{N}(0, \Delta + 3(2i-1)\Delta/2N)$ and $\widetilde{\zeta}_2(\cdot)$, given by (8.35), is independent of i. Clearly, $B(\mathbf{0}, 4\Delta)$ is covered by

$$\begin{aligned}
L_{2n} &= 1 + \sum_{i=1}^{N} L(i) \\
&\le (N+1)\exp\left\{k\left[\max_{1\le i\le N} R\left(Q_S^{(i)}, \Delta\right) + \widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right)\right] + o(k)\right\} \\
&\le \exp\left\{k\left[\frac{1}{2}\ln 4 + \widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right) + \frac{\ln(N+1)}{k}\right] + o(k)\right\}
\end{aligned}$$

size Δ balls. Recall that $M_n = \exp\{kR_{s,n}\}$ and by assumption

$$\lim_{n\to\infty} R_{s,n} = R > R(Q_S, \Delta) = \frac{1}{2}\ln\frac{\sigma_S^2}{\Delta} \ge \frac{1}{2}\ln 4.$$

Set $\delta = R - \frac{1}{2}\ln 4 > 0$. Finally, if we let N be the smallest integer satisfying $\widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right) \le \frac{\delta}{2}$ (noting that $\widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right) \to 0$ as $N \to \infty$), we have

$$\lim_{n\to\infty}\frac{L_{2n}}{M_n} \le \lim_{n\to\infty}\exp\left\{-tn\left[\delta - \widetilde{\zeta}_2\left(\frac{3\Delta}{2N}\right) - \frac{\ln(N+1)}{tn}\right] + o(n)\right\} = 0.$$

■

Proof of Theorem 10.12:

Forward Part: We show that there exists a sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$ composed by good source and channel codes such that

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\
&\ge \sup_{\tau R(Q_S,\Delta) < R < C(W_{Y|X},\mathcal{E})} \min\left\{\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right), E(R, W_{Y|X}, \mathcal{E})\right\} - \delta
\end{aligned}$$

for any $\delta > 0$. First note that for any given index assignment π_m, it follows from (10.41) that

$$\begin{aligned}
&\Pr\left(d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta \middle| \pi_m\right) \\
=\ & \sum_{i=1}^{M_n} \underbrace{P_W(\pi_m(i)|\pi_m(i))}_{\leq 1} \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_i) > \Delta\right\} d\mathbf{s} \\
+\ & \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} P_W(\pi_m(j)|\pi_m(i)) \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \underbrace{\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\}}_{\leq 1} d\mathbf{s} \\
\leq\ & \sum_{i=1}^{M_n} \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_i) > \Delta\right\} d\mathbf{s} + \sum_{i=1}^{M_n} Q_S^{(\tau n)}(D_i) \sum_{j=1, j\neq i}^{M_n} P_W(\pi_m(j)|\pi_m(i)) \\
\leq\ & P_\Delta^{(n)}(Q_S, R_{s,n}) + P_{max,ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}),
\end{aligned}$$

which only depends on the source and channel codes and is independent of π_m. Thus, for any sequence of tandem codes we have

$$\begin{aligned}
P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) &\leq P_\Delta^{(n)}(Q_S, R_{s,n}) + P_{max,ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}) \\
&\leq 2\max\left\{P_\Delta^{(n)}(Q_S, R_{s,n}), P_{max,ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\right\}
\end{aligned}$$

or equivalently,

$$\begin{aligned}
\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \geq \min\Bigg\{&\liminf_{n\to\infty} -\frac{1}{n}\ln P_\Delta^{(n)}(Q_S, R_{s,n}), \\
&\liminf_{n\to\infty} -\frac{1}{n}\ln P_{max,ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\Bigg\}.
\end{aligned}$$

Now fix $R > 0$ and $\delta > 0$. According to the definition of the source error exponent (Definition 2.2), there exists a sequence of $(\tau n, \widetilde{M}_n)$ source codes $(\widetilde{f}_{sn}, \widetilde{\varphi}_{sn}, \Delta)$ (with rate $\widetilde{R}_{s,n} = \frac{\ln \widetilde{M}_n}{\tau n}$) such that

$$\liminf_{n\to\infty} -\frac{1}{\tau n}\ln P_\Delta^{(n)}(Q_S, \widetilde{R}_{s,n}) \geq F_G(\tau R, Q_S, \Delta) - \delta \quad \text{and} \quad \limsup_{n\to\infty} \widetilde{R}_{s,n} \leq R.$$

Since a source code with a larger codebook size would have a smaller probability of excess distortion, there must exist a sequence of $(\tau n, \lceil 2^{tnR} \rceil)$ source codes $(f_{sn}, \varphi_{sn}, \Delta)$ such that

$$\liminf_{n\to\infty} -\frac{1}{\tau n}\ln P_\Delta^{(n)}(Q_S, R_{s,n}) \geq F_G(\tau R, Q_S, \Delta) - \delta.$$

Similarly, for given τR, the definition of channel error exponent (Definition 2.4 and the corresponding remark) asserts that there exists a sequence of $(n, \widehat{M}_n)$ channel codes $(\widehat{f}_{sn}, \widehat{\varphi}_{cn}, \mathcal{E})$ (with rate $\widehat{R}_{c,n} = \frac{\ln \widehat{M}_n}{n}$) such that

$$\liminf_{n\to\infty} -\frac{1}{n} \ln P^{(n)}_{max,ec}(W_{Y|X}, \widehat{R}_{c,n}, \mathcal{E}) \geq E(\tau R, W_{Y|X}, \mathcal{E}) - \delta \quad \text{and} \quad \liminf_{n\to\infty} \widehat{R}_{c,n} \geq \tau R.$$

Since a channel code with a smaller codebook size would have a smaller (maximum) probability of error, there must exist a sequence of $(n, \lceil 2^{tnR} \rceil)$ channel codes $(f_{cn}, \varphi_{cn}, \mathcal{E})$ such that

$$\liminf_{n\to\infty} -\frac{1}{n} \ln P^{(n)}_{max,ec}(W_{Y|X}, R_{c,n}, \mathcal{E}) \geq E(\tau R, W_{Y|X}, \mathcal{E}) - \delta.$$

If we restrict $R \in (R(Q_S, \Delta), C(W_{Y|X}, \mathcal{E})/\tau)$, then there exists a sequence of tandem codes, composed by a sequence of $(\tau n, \lceil 2^{tnR} \rceil)$ good source codes, and a sequence of $(n, \lceil 2^{tnR} \rceil)$ good channel codes (with the same $M_n = \lceil 2^{tnR} \rceil$), such that

$$\liminf_{n\to\infty} -\frac{1}{n} \ln P^{(n)}_{\Delta^*}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \geq \min\left\{\tau F_G\left(R, Q_S, \Delta\right), E(\tau R, W_{Y|X}, \mathcal{E})\right\} - \delta.$$

Finally, since R and δ are arbitrary, we can take the supremum over $R(Q_S, \Delta) < R < C(W_{Y|X}, \mathcal{E})/\tau$, completing the proof of the forward part.

Converse Part: We next show that for any sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$ composed by good source and channel codes

$$\begin{aligned} &\liminf_{n\to\infty} -\frac{1}{n} \ln P^{(n)}_{\Delta^*}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\ &\leq \sup_{\tau R(Q_S,\Delta) < R < C(W_{Y|X},\mathcal{E})} \min\left\{\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right), E(R, W_{Y|X}, \mathcal{E})\right\}. \end{aligned} \tag{10.42}$$

As in [51], we decompose the probability of excess distortion for any given tandem codes as follow,

$$\begin{aligned} &P^{(n)}_{\Delta^*}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \\ &= \frac{1}{M_n!} \sum_{m=1}^{M_n!} \sum_{i=1}^{M_n} P_W(\pi_m(i)|\pi_m(i)) \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \mathbb{1}\left\{d^{(\tau n)}(\mathbf{s}, \mathbf{c}_i) > \Delta\right\} d\mathbf{s} \\ &+ \frac{1}{M_n!} \sum_{m=1}^{M_n!} \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} P_W(\pi_m(j)|\pi_m(i)) \int_{D_i} Q_S^{(\tau n)}(\mathbf{s}) \mathbb{1}\left\{d^{(\tau n)}(\mathbf{s}, \mathbf{c}_j) > \Delta\right\} d\mathbf{s}. \end{aligned}$$

Note that for fixed i,

$$\frac{1}{M_n!}\sum_{m=1}^{M_n!} P_W(\pi_m(i)|\pi_m(i)) = \frac{1}{M_n}\sum_{j=1}^{M_n} P_W(j|j)$$

is actually the arithmetic mean of $P_W(i|i)$ and is independent of i. Similarly, for fixed i and $j \neq i$,

$$\frac{1}{M_n!}\sum_{m=1}^{M_n!} P_W(\pi_m(j)|\pi_m(i)) = \frac{1}{M_n(M_n-1)}\sum_{k=1}^{M_n}\sum_{l=1,l\neq k}^{M_n} P_W(l|k)$$

is actually the arithmetic mean of $P_W(l|k)$ $(l \neq k)$ and is independent of l and k. Thus,

$$\begin{aligned}
&P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau)\\
&= \left[\frac{1}{M_n}\sum_{i=1}^{M_n} P_W(j|j)\right]\sum_{i=1}^{M_n}\int_{D_i} Q_S^{(\tau n)}(\mathbf{s})\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_i) > \Delta\right\} d\mathbf{s}\\
&+ \sum_{i=1}^{M_n}\sum_{j=1,j\neq i}^{M_n}\left[\underbrace{\frac{1}{M_n(M_n-1)}}_{\geq 1/M_n^2}\sum_{k=1}^{M_n}\sum_{l=1,l\neq k}^{M_n} P_W(l|k)\right]\int_{D_i} Q_S^{(\tau n)}(\mathbf{s})\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\} d\mathbf{s}\\
&\geq \left(1 - P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\right) P_{\Delta}^{(n)}(Q_S, R_{s,n}) + P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\\
&\quad \frac{1}{M_n}\sum_{i=1}^{M_n}\int_{D_i}\sum_{j=1,j\neq i}^{M_n} Q_S^{(\tau n)}(\mathbf{s})\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\} d\mathbf{s}. \qquad (10.43)
\end{aligned}$$

We then bound

$$\begin{aligned}
&\frac{1}{M_n}\sum_{i=1}^{M_n}\int_{D_i}\sum_{j=1,j\neq i}^{M_n} Q_S^{(\tau n)}(\mathbf{s})\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\} d\mathbf{s}\\
&\geq \sum_{i=1}^{M_n}\int_{D_i} Q_S^{(\tau n)}(\mathbf{s})\frac{1}{M_n}\left[\sum_{j=1}^{M_n}\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\} - 1\right] d\mathbf{s}\\
&\geq D^{(n)}(Q_S, R_{s,n}, \Delta) - \frac{1}{M_n},
\end{aligned}$$

where

$$D^{(n)}(Q_S, R_{s,n}, \Delta) \triangleq \min_{\mathbf{s}\in\mathcal{S}^{\tau n}} \frac{1}{M_n}\sum_{j=1}^{M_n}\mathbb{1}\left\{d^{(\tau n)}(\mathbf{s},\mathbf{c}_j) > \Delta\right\}.$$

Substituting the above into (10.43) gives

$$\begin{aligned}
P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \geq{}& \left(1 - P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\right) P_{\Delta}^{(n)}(Q_S, R_{s,n})\\
&+ P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\left(D^{(n)}(Q_S, R_{s,n}, \Delta) - \frac{1}{M_n}\right) \qquad (10.44)
\end{aligned}$$

By definition, for any good channel codes $(f_{cn}, \varphi_{cn}, \mathcal{E}) \in \Xi(W_{Y|X}, \mathcal{E})$, $1 - P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})$ is bounded away from zero for n sufficiently large.

Lemma 10.2 *Let* $SDR = \frac{\sigma_S^2}{\Delta} \geq 4$. *For any sequence of good source codes* $(f_{sn}, \varphi_{sn}, \Delta) \in \Omega(Q_S, \Delta)$ *with rate* $R_{s,n}$ *such that* $\lim_{n\to\infty} R_{s,n} = R$, *there exists some* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} D^{(n)}(Q_S, R_{s,n}, \Delta) > \delta.$$

Proof: It suffices to show if

$$\limsup_{n\to\infty} D^{(n)}(Q_S, R_{s,n}, \Delta) = 0,$$

then $(f_{sn}, \varphi_{sn}, \Delta) \notin \Omega(Q_S, \Delta)$, i.e.,

$$\liminf_{n\to\infty} -\frac{1}{n} \ln P_{\Delta}^{(n)}(Q_S, R_{s,n}) < F\left(\lim_{n\to\infty} R_{s,n}, Q_S, \Delta\right)$$

for such sequence of source codes $(f_{sn}, \varphi_{sn}, \Delta)$. Let $\left\{\mathbf{s}^* = \mathbf{s}^*(n) \in \mathbb{R}^{tn}\right\}$ be the sequence of source vectors achieving the minimum in $D^{(n)}(Q_S, R_{s,n}, \Delta)$ for every n. Then

$$\limsup_{n\to\infty} D^{(n)}(Q_S, R_{s,n}, \Delta) = \limsup_{n\to\infty} \frac{1}{M_n} \sum_{j=1}^{M_n} \mathbb{1}\left\{d^{(tn)}\left(\mathbf{s}^*, \mathbf{c}_j\right) > \Delta\right\} = 0 \tag{10.45}$$

implies that the source codebook $\mathcal{C}$ has only L_{1n} codewords outside the ball $B(\mathbf{s}^*, \Delta)$ such that

$$\limsup_{n\to\infty} \frac{L_{1n}}{M_n} = 0,$$

recalling that under the squared-error distortion measure

$$B(\mathbf{s}^*, \Delta) = \left\{\mathbf{s} \in \mathcal{S}^{tn} : \|\mathbf{s}^* - \mathbf{s}\|^2 \leq tn\Delta\right\},$$

where $\|\mathbf{s}^* - \mathbf{s}\| = \sqrt{\sum_{i=1}^{tn} (s_i^* - s_i)^2}$. It then follows that

$$\begin{aligned}
&\limsup_{n\to\infty} P_{\Delta}^{(n)}(Q_S, R_{s,n}) \\
&\quad = \limsup_{n\to\infty} \left\{1 - Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \in \mathcal{C}} B(\mathbf{c}_i, \Delta)\right)\right\} \\
&\quad \geq \limsup_{n\to\infty} \left\{1 - Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \in B(\mathbf{s}^*, \Delta)} B(\mathbf{c}_i, \Delta)\right) - Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \notin B(\mathbf{s}^*, \Delta)} B(\mathbf{c}_i, \Delta)\right)\right\}.
\end{aligned}$$

Clearly, the squared distance between any vector $\mathbf{s}$ in the ball $B(\mathbf{c}_i, \Delta)$ and the "center" $\mathbf{s}^*$ is bounded by

$$\|\mathbf{s}^* - \mathbf{s}\|^2 \leq (\|\mathbf{s}^* - \mathbf{c}_i\| + \|\mathbf{c}_i - \mathbf{s}\|)^2 \leq (\sqrt{tn\Delta} + \sqrt{tn\Delta})^2 = 4tn\Delta.$$

We hence can bound

$$Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \in B(\mathbf{s}^*,\Delta)} B(\mathbf{c}_i, \Delta)\right) \leq Q_S^{(tn)}\left(\mathbf{s} : \|\mathbf{s}^* - \mathbf{s}\|^2 \leq 4tn\Delta\right) \leq Q_S^{(tn)}\left(\mathbf{s} : \|\mathbf{s}\|^2 \leq 4tn\Delta\right)$$

where the last inequality holds since the zero-mean MGS has a larger density in the the neighborhood of origin $\mathbf{0}$.

Now, based on Lemma 10.1, we claim that, there exists a sequence of $(tn, L_{1n} + L_{2n})$ source codes $(\widetilde{f}_{sn}, \widetilde{\varphi}_{sn}, \Delta)$ with code rate

$$R_{L,n} = \frac{\ln(L_{1n} + L_{2n})}{n}$$

such that the probability of excess distortion is less than

$$1 - Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \in B(\mathbf{s}^*,\Delta)} B(\mathbf{c}_i, \Delta)\right) - Q_S^{(tn)}\left(\bigcup_{\mathbf{c}_i \notin B(\mathbf{s}^*,\Delta)} B(\mathbf{c}_i, \Delta)\right).$$

In other words, for any given sequence of source codes with

$$\limsup_{n\to\infty} D^{(n)}(Q_S, R_{s,n}, \Delta) = 0,$$

the corresponding probability of excess distortion can be lower bounded by another sequence of codes with rate $R_{L,n}$, i.e.,

$$\limsup_{n\to\infty} P_\Delta^{(n)}(Q_S, R_{s,n}) \geq \limsup_{n\to\infty} P_\Delta^{(n)}(Q_S, R_{L,n}).$$

It is easy to see that

$$\limsup_{n\to\infty} R_{L,n} \leq \lim_{n\to\infty} R_{s,n} - \epsilon = R - \epsilon$$

for some $\epsilon > 0$ since

$$\limsup_{n\to\infty} \frac{L_{1n} + L_{2n}}{M_n} = 0.$$

Therefore, by the definition of source excess distortion exponent (Definition 2.2),

$$\liminf_{n\to\infty} -\frac{1}{tn}\ln P_{\Delta}^{(n)}(Q_S, R_{s,n}) \le F(R-\epsilon, Q_S, \Delta) < F(R, Q_S, \Delta)$$

since $F(R, Q_S, \Delta)$ is strictly increasing and continuous at $R = \lim_{n\to\infty}\frac{M_n}{tn} > R(Q_S, \Delta)$. ■

Thus, for any sequence of tandem codes composed by good channel and source codes, there exists some $\delta > 0$ (independent of n) such that

$$\begin{aligned}\limsup_{n\to\infty} P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) &\ge \limsup_{n\to\infty} \delta\left(P_{\Delta}^{(n)}(Q_S, R_{s,n}) + P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\right)\\ &\ge \limsup_{n\to\infty} \delta \max\{P_{\Delta}^{(n)}(Q_S, R_{s,n}), P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\}\end{aligned} \tag{10.46}$$

or equivalently

$$\begin{aligned}\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \le \min\Bigg\{&\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta}^{(n)}(Q_S, R_{s,n}),\\ &\liminf_{n\to\infty} -\frac{1}{n}\ln P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E})\Bigg\}.\end{aligned}$$

Now for any sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$, let

$$R = \lim_{n\to\infty}\frac{\ln M_n}{\tau n} \in (R(Q_S, \Delta), C(W_{Y|X}, \mathcal{E})/\tau).$$

By the definition of the source excess distortion exponent (Definition 2.2)

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta}^{(n)}(Q_S, R_{s,n}) \le \tau F_G(R, Q_S, \Delta)$$

holds for any sequence of $(\tau n, M_n)$ block source codes since $\limsup_{n\to\infty}\frac{\ln M_n}{\tau n} \le R$. Similarly, by the definition of the channel error exponent (Definition 2.4)

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_{ec}^{(n)}(W_{Y|X}, R_{c,n}, \mathcal{E}) \le E(\tau R, W_{Y|X}, \mathcal{E})$$

holds for any sequence of (n, M_n) block channel codes since $\liminf_{n\to\infty}\frac{\ln M_n}{n} \ge \tau R$. Therefore,

$$\liminf_{n\to\infty} -\frac{1}{n}\ln P_{\Delta^*}^{(n)}(Q_S, W_{Y|X}, \mathcal{E}, \tau) \le \min\{\tau F_G(R, Q_S, \Delta), E(\tau R, W_{Y|X}, \mathcal{E})\}$$

holds for any sequence of tandem codes $(f_n^*, \varphi_n^*, \Delta, \mathcal{E}, \tau) \in \Lambda(Q_S, W_{Y|X}, \Delta, \mathcal{E}, \tau)$ composed by good source and channel codes. Since $R = \lim_{n\to\infty} \ln \frac{M_n}{\tau n} \in (R(Q_S,\Delta), C(W_{Y|X},\mathcal{E})/\tau)$ is arbitrary, we can take the supremum of R over this region, which yields the upper bound (10.42). ■

Since the MGC error exponent is not known for low rates, we can obtain computable lower and upper bounds to $E_T^{\Delta,\mathcal{E}}$ by replacing $E(R, W_{Y|X}, \mathcal{E})$ by its lower and upper bounds.

Corollary 10.3

$$\underline{E}_{Tr}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \le E_T^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \le \overline{E}_{Tsp}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau)$$

where

$$\underline{E}_{Tr}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \triangleq \sup_{\tau R(Q_S,\Delta)<R<C(W_{Y|X},\mathcal{E})} \min\left\{\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right), E_\dagger(R, W_{Y|X}, \mathcal{E})\right\}$$

and

$$\overline{E}_{Tsp}^{\Delta,\mathcal{E}}(Q_S, W_{Y|X}, \tau) \triangleq \sup_{\tau R(Q_S,\Delta)<R<C(W_{Y|X},\mathcal{E})} \min\left\{\tau F_G\left(\frac{R}{\tau}, Q_S, \Delta\right), E_{sp}(R, W_{Y|X}, \mathcal{E})\right\}.$$

Obviously, the tandem exponent is exactly determined if

$$\tau F_G(R/\tau, Q_S, \Delta) \text{ and } E_{sp}(R, W_{Y|X}, \mathcal{E})$$

intersects at rate $R_o' \ge R_{cr}(W_{Y|X})$ (in that case $R_o' = R_o$). Furthermore, it can be seen that the JSCC exponent strictly outperform the tandem coding exponent ($E_J^{\Delta,\mathcal{E}} > E_T^{\Delta,\mathcal{E}}$) if $E_J^{\Delta,\mathcal{E}}$ is determined exactly by its two bounds, i.e., if (8.44) is satisfied; or if the tandem coding exponent is determined by $\underline{E}_{Tr}^{\Delta,\mathcal{E}}$ and $\overline{E}_{Tsp}^{\Delta,\mathcal{E}}$, i.e. $R_o' \ge R_{cr}(W_{Y|X})$.

In contrast to the discrete systems studied in Chapters 5 and 7, the source and channel exponents for the Gaussian system have very simple analytical (computable) form, which are also continuous and differentiable functions of rate R (their expressions do not include any optimization operation). Therefore, the advantage of the JSCC exponent over the tandem exponent can be assessed by numerically comparing the lower bound of joint exponent $\underline{E}_{Jr}^{\Delta,\mathcal{E}}$ and the upper bound of tandem exponent $\overline{E}_{Tsp}^{\Delta,\mathcal{E}}$. For transmission rate $\tau = 1$, we partition in

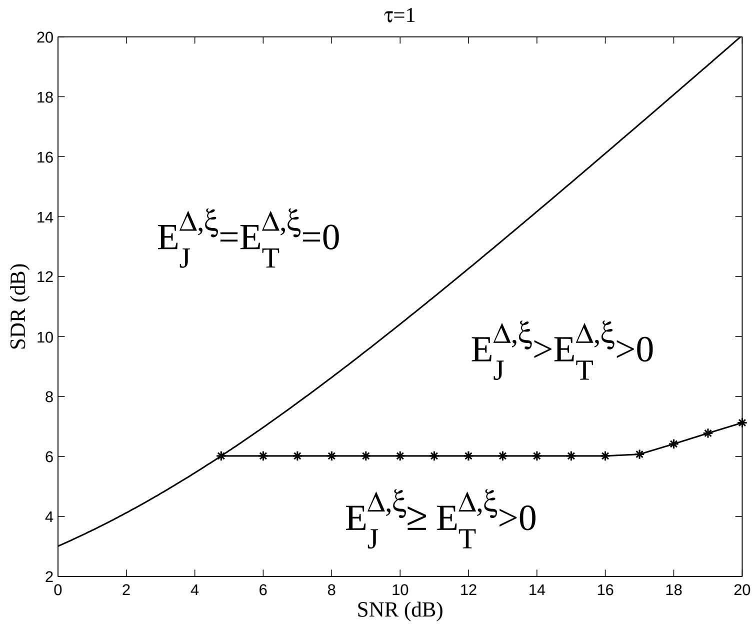

Figure 10.16: The regions for the MGS-MGC pairs with $\tau = 1$. Note that the region for $E_J^{\Delta,\mathcal{E}} > E_T^{\Delta,\mathcal{E}}$ does not include the boundary.

Fig. 10.16 the SNR-SDR plane into three regions. When SDR ≥ 4 ($\approx$ 6dB), $\underline{E}_{Jr}^{\Delta,\mathcal{E}} > \overline{E}_{Tsp}^{\Delta,\mathcal{E}}$ (which means $E_J^{\Delta,\mathcal{E}} > E_T^{\Delta,\mathcal{E}}$) for a large class of source-channel pairs. For example, when SDR $= 7$ dB, $E_J^{\Delta,\mathcal{E}} > E_T^{\Delta,\mathcal{E}}$ holds for 10 dB $\leq$ SNR $\leq$ 24 dB (approximately). We also compute the two bounds of $E_J^{\Delta,\mathcal{E}}$ and $E_T^{\Delta,\mathcal{E}}$, no matter they are determined or not, and we see from Fig. 10.17 that when SDR $= 8$ dB, $E_J^{\Delta,\mathcal{E}}$ (or its lower bound) almost double $E_T^{\Delta,\mathcal{E}}$ (or its upper bound) for 8dB $\leq$ SNR $\leq$ 15dB. It is also observed that for the same exponent (e.g. $0.2 \sim 1.1$), the gain of JSCC over tandem coding could be as large as 2dB in SNR. Similar results are observed for other parameters, see Figs. 10.18 and 10.19 for $\tau = 1.5$.

10.6 Asymmetric 2-User Systems

10.6.1 Tandem System with Common Randomization

In Section 9.5, we showed that the reliable transmissibility condition $(\tau H_Q(S), \tau H_Q(L|S)) \in \mathcal{R}(W_{YZ|UX})$ in the JSCC theorem (Theorem 9.4) for asymmetric 2-user systems can be

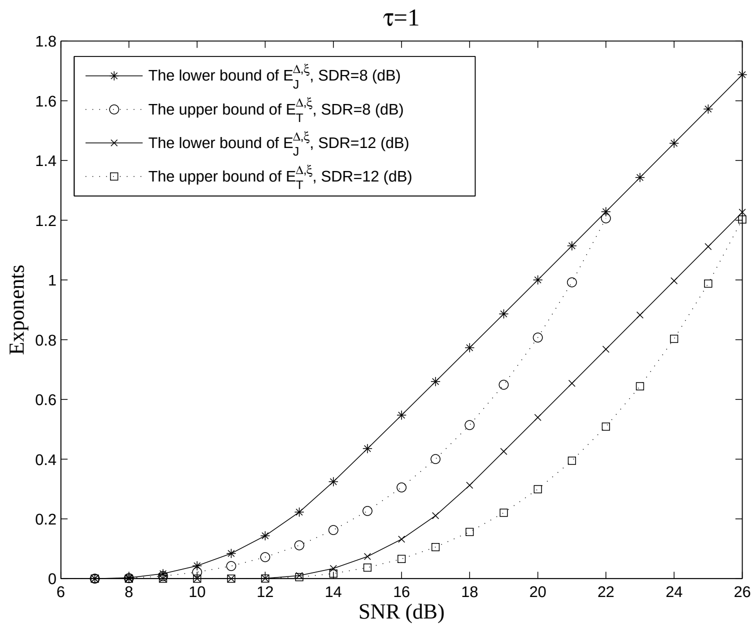

Figure 10.17: MGS-MGC source-channel pair: the lower bound of $E_J^{\Delta,\mathcal{E}}$ vs the upper bound of $E_T^{\Delta,\mathcal{E}}$ for $\tau = 1$.

achieved by a tandem coding system where separately designed source and channel coding operations are sequentially applied; see Figs. 9.3 and 9.4. Note however that, as long as the source encoder is directly concatenated by a channel encoder, the source statistics would be automatically brought into the channel coding stage. Thus, the performance of the channel code is affected by that of the source code (since the compressed messages (indices) fed into the channel encoders are not necessarily uniformly distributed). Similar to the single-user systems, to statistically decouple the source and channel coding operations, we need to employ common randomization between the source and channel coding components. This results in a "complete" tandem coding system with fully independent source and channel coding operations, and for which we can establish an expression for its error exponent in terms of the source coding and channel coding exponents. The tandem coding system is depicted in Figs. 10.20 and 10.21.

As in Section 9.5, the encoder f_n is composed of two source encoders f_{sn} and g_{sn} and

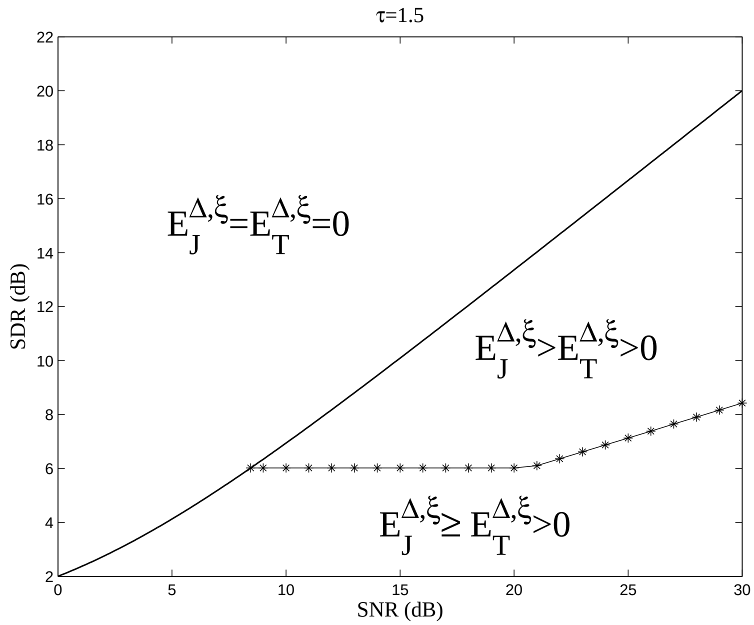

Figure 10.18: The regions for the MGS-MGC pairs with $\tau = 1.5$. Note that the region for $E_J^{\Delta,\mathcal{E}} > E_T^{\Delta,\mathcal{E}}$ does not include the boundary.

one channel encoder f_{cn}. The difference is that the indices $i = f_{sn}(\mathbf{l})$ and $j = g_{sn}(\mathbf{s})$ are separately mapped to channel indices through permutation functions $\pi_f : \{1, 2, ..., M_l\} \rightarrow \{1, 2, ..., M_l\}$ and $\pi_g : \{1, 2, ..., M_s\} \rightarrow \{1, 2, ..., M_s\}$, which are usually called index assignments (π_f and π_g are assumed to be known at both the transmitter and the receiver). Furthermore, the choice of π_f (π_g, respectively), is assumed random and equally likely from all $M_l!$ ($M_s!$, respectively) different possible index assignments, so that the indices fed into the channel encoder have a uniform distribution and are mutually independent:

$$\begin{aligned}
\Pr(\pi_f(f_{sn}(L^{\tau n})) = a) &= \sum_{i=1}^{M_l} \Pr(f_{sn}(L^{\tau n}) = i) \Pr(\pi_f(i) = a | f_{sn}(L^{\tau n}) = i) \\
&= \sum_{i=1}^{M_l} \Pr(f_{sn}(L^{\tau n}) = i) \frac{(M_l - 1)!}{M_l!} = \frac{1}{M_l}, \\
\Pr(\pi_g(g_{sn}(S^{\tau n})) = b) &= \frac{1}{M_s}, \\
\Pr(\pi_g(f_{sn}(L^{\tau n})) = a, \pi_g(g_{sn}(S^{\tau n})) = b) &= \Pr(\pi_f(f_{sn}(L^{\tau n})) = a) \Pr(\pi_g(g_{sn}(S^{\tau n})) = b),
\end{aligned}$$

for any $(a, b) \in \{1, 2, ..., M_l\} \times \{1, 2, ..., M_s\}$. Hence common randomization achieves statis-

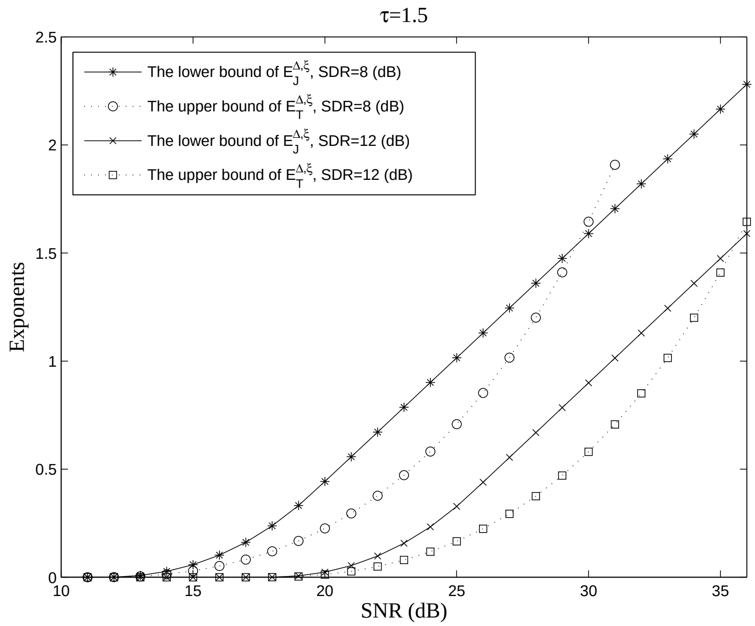

Figure 10.19: MGS-MGC source-channel pair: the lower bound of $E_J^{\Delta,\mathcal{E}}$ vs the upper bound of $E_T^{\Delta,\mathcal{E}}$ for $\tau = 1.5$.

tical independence between the source and channel coding operations.

Similarly, the encoder g_n is independently composed of a source encoder g_{sn}, an index mapping $\pi_g : \{1, 2, ..., M_s\} \to \{1, 2, ..., M_s\}$, and a channel encoder $g_{cn} : \{1, 2, ..., M_s\} \to \mathcal{U}^n$.

At the receiver side, the decoder φ_n is composed of a channel decoder φ_{cn}, a pair of index mappings (π_f^{-1}, π_g^{-1}) which maps every channel index pair $(\pi_f(\widehat{i}), \pi_g(\widehat{j}))$ back to a source index pair $(\widehat{i}, \widehat{j})$, and a source decoder φ_{sn} which outputs the approximation of the source messages $\mathbf{s}'$ and $\mathbf{l}'$. Similarly, the decoder ψ_n is composed of a channel decoder $\psi_{cn} : \mathcal{Z}^n \to \{1, 2, ..., M_s\}$, an index mapping π_g^{-1}, and a source decoder $\psi_{sn} : \{1, 2, ..., M_s\} \to \mathcal{S}^{\tau n}$.

For the above tandem system we assume that the following limits exist:[4]

$$\liminf_{n\to\infty} \frac{1}{n} \log_2 M_l = \limsup_{n\to\infty} \frac{1}{n} \log_2 M_l = \lim_{n\to\infty} \frac{1}{n} \log_2 M_l$$

[4]This assumption is used later to upper bound the tandem coding error exponent in Theorem 10.13.

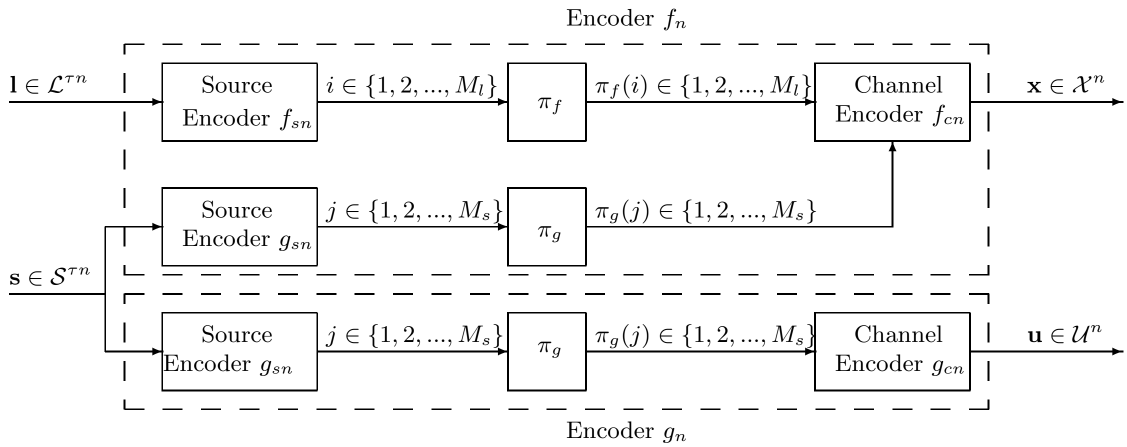

Figure 10.20: Tandem source-channel coding system - encoders.

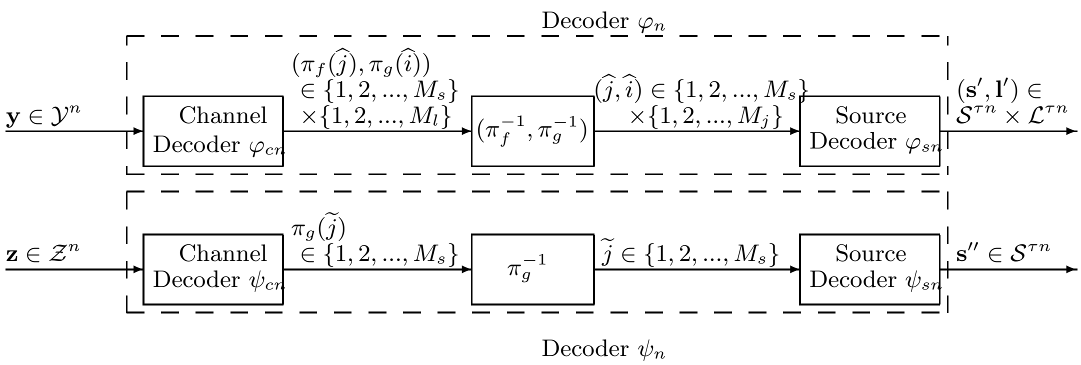

Figure 10.21: Tandem source-channel coding system - decoders.

and

$$\liminf_{n\to\infty} \frac{1}{n} \log_2 M_s = \limsup_{n\to\infty} \frac{1}{n} \log_2 M_s = \lim_{n\to\infty} \frac{1}{n} \log_2 M_s.$$

10.6.2 Tandem Coding Error Exponent

We now can study the error performance and exponent of tandem source-channel coding (with common randomization) for the asymmetric 2-user system. Since the tandem code consists of a source code $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ and a channel code $(f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$, we first define the corresponding source coding error exponent (note that the corresponding channel coding error exponent for the asymmetric 2-user channel was defined in Section 9.6).

Let $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ be a sequence of source code for CS Q_{SL} with common source rate $\widehat{R}_s$ and private source rate $\widehat{R}_l$ as defined in the last section. The probabilities of Y- and Z- error for the source coding are respectively given by

$$P^{(n)}_{Yes}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) \triangleq \Pr(\{\varphi_{sn}(i,j) \neq (S^{\tau n}, L^{\tau n})\}) = \sum_{(\mathbf{s},\mathbf{l}):\psi_{sn}(i,j)\neq(\mathbf{s},\mathbf{l})} Q^{(n)}_{SL}(\mathbf{s},\mathbf{l}) \qquad (10.47)$$

and

$$P^{(n)}_{Zes}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) = P^{(n)}_{Ze}(\widehat{R}_s, Q_S) \triangleq \Pr(\{\psi_{sn}(j) \neq S^{\tau n}\}) = \sum_{\mathbf{s}:\psi_{sn}(i)\neq\mathbf{s}} Q^{(n)}_{S}(\mathbf{s}) \qquad (10.48)$$

where $i \triangleq f_{sn}(\mathbf{l})$ and $j \triangleq g_{sn}(\mathbf{s})$. The probability of the overall 2-user source coding error is given by

$$P^{(n)}_{es}(\widehat{R}_s, \widehat{R}_l, Q_{SL}) \triangleq \Pr\left(\{\varphi_{sn}(i,j) \neq (S^{\tau n}, L^{\tau n})\} \cup \{\psi_{sn}(i) \neq S^{\tau n}\}\right). \qquad (10.49)$$

Definition 10.6 The 2-user source coding error exponent $E(R_1, R_2, Q_{SL})$, for any $R_1 > 0$ and $R_2 > 0$, is defined by the supremum of the set of all numbers E_s for which there exists a sequence of source codes $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$ with blocklength n, the common rate no larger than R_1, and the private rate no larger than R_2, such that

$$E_s \leq \liminf_{n\to\infty} -\frac{1}{n}\log_2 P^{(n)}_{es}(R_1, R_2, Q_{SL}). \qquad (10.50)$$

Clearly, for any sequence of source codes $(f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn})$, the error probability

$$P^{(n)}_{es}(R_1, R_2, Q_{SL})$$

must be larger than $P^{(n)}_{Yes}(R_1, R_2, Q_{SL})$ and $P^{(n)}_{Zes}(R_1, R_2, Q_{SL})$) but less than the sum of the two, so we have

$$\begin{aligned}
&\liminf_{n\to\infty} -\frac{1}{n}\log_2 P^{(n)}_{es}(R_1, R_2, Q_{SL}) \\
&\quad = \liminf_{n\to\infty} -\frac{1}{n}\log_2 \max\left(P^{(n)}_{Yes}(R_1, R_2, Q_{SL}), P^{(n)}_{Zes}(R_1, R_2, Q_{SL})\right). \qquad (10.51)
\end{aligned}$$

In what follows we need to make two assumptions regarding the source codes in order to analyze the probability of error of the overall tandem system. Let the source codebook for (g_{sn}, ψ_{sn}) (Receiver Z) be $\mathcal{C}^{(g)} = \{\mathbf{c}^{(g)}_1, ..., \mathbf{c}^{(g)}_{M_l}\} \subseteq \mathcal{S}^{\tau n}$, and let the source codebook for

$(f_{sn}, g_{sn}, \varphi_{sn})$ (Receiver Y) be $\mathcal{C}^{(f)} \times \mathcal{C}^{(g)}$ where $\mathcal{C}^{(f)} = \{\mathbf{c}_1^{(f)}, ..., \mathbf{c}_{M_s}^{(f)}\} \subseteq \mathcal{L}^{\tau n}$. We assume that (A1) the source encoder f_{sn} satisfies the condition (for every n): $Q_L^{\tau n}(f_{sn}^{-1}(i)) > 0$ and $\mathbf{c}_i^{(f)} \in f_{sn}^{-1}(i)$ for every $i = 1, 2, ..., M_l$, where $f_{sn}^{-1}(i) \triangleq \{\mathbf{l} \in \mathcal{L}^{\tau n} : f_{sn}(\mathbf{l}) = i\}$. If $Q_L^{\tau n}(f_{sn}^{-1}(i)) = 0$ for some i, then the codeword $\mathbf{c}_i^{(f)}$ is redundant, and we can remove it from the codebook $\mathcal{C}^{(f)}$. If $\mathbf{c}_i^{(f)} \notin f_{sn}^{-1}(i)$, we can map the index i to some source message $\widehat{\mathbf{l}}$ such that $Q_L^{\tau n}(\widehat{\mathbf{l}}) > 0$ and $f_{sn}(\widehat{\mathbf{l}}) = i$, so that the source coding probability of error $P_{Yes}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL})$ is strictly reduced by setting $\widehat{\mathbf{l}}$ as the codeword $\mathbf{c}_i^{(f)}$ (note that $P_{Zes}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL})$ is independent of f_{sn}). Similarly, we assume that (A2) the source code g_{sn} satisfies the condition (for every n): $Q_S^{\tau n}(g_{sn}^{-1}(j)) > 0$ and $\mathbf{c}_j^{(g)} \in g_{sn}^{-1}(j)$ for every $j = 1, 2, ..., M_s$, where $g_{sn}^{-1}(j) \triangleq \{\mathbf{s} \in \mathcal{S}^{\tau n} : g_{sn}(\mathbf{s}) = j\}$. If $Q_S^{\tau n}(g_{sn}^{-1}(j)) = 0$ for some j, then the codeword $\mathbf{c}_j^{(g)}$ is redundant, and we can remove it from the codebook $\mathcal{C}^{(g)}$. If $\mathbf{c}_j^{(g)} \notin g_{sn}^{-1}(j)$, we can map the index j to some source message $\widehat{\mathbf{s}}$ such that $Q_S^{\tau n}(\widehat{\mathbf{s}}) > 0$ and $g_{sn}(\widehat{\mathbf{s}}) = j$, so that the source coding error probabilities $P_{Yes}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL})$ and $P_{Zes}^{(n)}(\widehat{R}_s, \widehat{R}_l, Q_{SL})$ are strictly reduced by setting $\widehat{\mathbf{s}}$ as the codeword $\mathbf{c}_j^{(g)}$. We remark that the source code satisfying (A1) and (A2) does not lose optimality (in the sense of achieving the source error exponent).

Denote $\pi^{-1}(i,j) \triangleq (\pi_f^{-1}(i), \pi_g^{-1}(j))$. By introducing (A1) and (A2), the error probability of the tandem code $(f_n^*, \varphi_n^*) \triangleq (f_{sn}, g_{sn}, \varphi_{sn}, \psi_{sn}, f_{cn}, g_{cn}, \varphi_{cn}, \psi_{cn})$ is given by

$$
\begin{aligned}
&P_{e^*}^{(n)}(Q_{SL}, W_{YZ|UX}, \tau) \\
&\triangleq \Pr\Big(\{\varphi_{sn}\left[\pi^{-1}(\varphi_{cn}(Y^n))\right] \neq (S^{\tau n}, L^{\tau n})\} \bigcup \{\psi_{sn}\left[\pi_g^{-1}(\psi_{cn}(Z^n))\right] \neq S^{\tau n}\}\Big) \\
&= \sum_{a=1}^{M_l}\sum_{b=1}^{M_s} \underbrace{\Pr(\pi_f[f_{sn}(L^{\tau n})] = a)}_{=1/M_l} \underbrace{\Pr(\pi_g[g_{sn}(S^{\tau n})] = b)}_{=1/M_s} \\
&\quad \Big[\Pr\Big(\{\varphi_{cn}(Y^n) \neq (a,b)\} \bigcup \{\psi_{cn}(Z^n) \neq b\} \Big| \pi_f[f_{sn}(L^{\tau n})] = a, \pi_g[g_{sn}(S^{\tau n})] = b\Big) + \\
&\quad \Pr\big(\{\varphi_{cn}(Y^n) = (a,b) \text{ and } \psi_{cn}(Z^n) = b\} \\
&\quad \cap\{\varphi_{sn}[\pi^{-1}(a,b)] \neq (S^{\tau n}, L^{\tau n}) \text{ or } \psi_{sn}[\pi_g^{-1}(b)] \neq S^{\tau n}\} \big| \\
&\qquad \pi_f[f_{sn}(L^{\tau n})] = a, \pi_g[g_{sn}(S^{\tau n})] = b\big)\Big] \qquad (10.52) \\
&= \sum_{a=1}^{M_l}\sum_{b=1}^{M_s} \frac{1}{M_l M_s} \Pr\Big(\{\varphi_{cn}(Y^n) \neq (a,b)\} \bigcup \{\psi_{cn}(Z^n) \neq b\} | (a,b) \text{ is sent}\Big) \\
&\quad + \Pr\Big(\{\varphi_{sn}[S^{\tau n}, L^{\tau n}] \neq (S^{\tau n}, L^{\tau n})\} \bigcup \{\psi_{sn}[S^{\tau n}] \neq S^{\tau n}\}\Big)
\end{aligned}
$$

$$\sum_{a=1}^{M_l}\sum_{b=1}^{M_s}\frac{1}{M_lM_s}\Pr\left(\{\varphi_{cn}(Y^n)=(a,b)\}\bigcap\{\psi_{cn}(Z^n)=b\}|(a,b)\text{ is sent}\right) \quad (10.53)$$

$$= P_{ec}^{(n)}(\tau\widehat{R}_s,\tau\widehat{R}_l,W_{YZ|UX})+[1-P_{ec}^{(n)}(\tau\widehat{R}_s,\tau\widehat{R}_l,W_{YZ|UX})]P_{es}^{(\tau n)}(\widehat{R}_s,\widehat{R}_l,Q_{SL}),(10.54)$$

where (10.52) follows from assumptions (A1) and (A2), which imply that a channel decoding error must cause an overall system decoding error, (10.53) holds due to the independence of source and channel coding.

Definition 10.7 The tandem coding error exponent $E_T(Q_{SL},W_{YZ|UX},\tau)$ for source Q_{SL} and channel $W_{YZ|UX}$ is defined as the supremum of the set of all numbers $\widehat{E}$ for which there exists a sequence of tandem codes (f_n^*,φ_n^*) satisfying (A1) and (A2) with transmission rate τ such that

$$\widehat{E}\le\liminf_{n\to\infty}-\frac{1}{n}\log_2 P_{e^*}^{(n)}(Q_{SL},W_{YZ|UX},\tau).$$

When there is no possibility of confusion, $E_T(Q_{SL},W_{YZ|UX},\tau)$ will often be written as E_T. The following lemma illustrates the relation between E_T and E_J. Similar to Proposition 10.1, we have the following order.

Proposition 10.3 $E_J(Q_{SL},W_{YZ|UX},\tau)\ge E_T(Q_{SL},W_{YZ|UX},\tau)$.

We next give an explicit formula for E_T in terms of the corresponding source and channel error exponents.

Theorem 10.13

$$E_T(Q_{SL},W_{YZ|UX},\tau)=\sup_{R_1>0,R_2>0}\min\left\{\tau e\left(\frac{R_1}{\tau},\frac{R_2}{\tau},Q_{SL}\right),E(R_1,R_2,W_{YZ|UX})\right\}$$

where $e(R_1,R_2,Q_{SL})$ is the 2-user source coding error exponent defined in (10.6) and $E(R_1,R_2,W_{YZ|UX})$ is the asymmetric 2-user channel coding error exponent defined in (9.2).

Proof: The proof is basically the same as the proof for Theorem 10.1 and is omitted. ■

Although tandem source-channel coding can achieve the reliable transmission condition, it might not achieve the system JSCC error exponent. In the following we consider the tandem system consisting of CS Q_{SL} and the AMAC $W_{Y|UX}$. For the CS-AMAC tandem system, we have only one receiver, Receiver Y, and the source decoder (cf. Fig. 10.21) φ_{sn} becomes a Sepian-Wolf decoder [29]. Furthermore,

$$P_{es}^{(n)}(R_1, R_2, Q_{SL}) = P_{Yes}^{(n)}(R_1, R_2, Q_{SL}) = \sum_{(\mathbf{s},\mathbf{l}):\psi_{sn}(i,j)\neq(\mathbf{s},\mathbf{l})} Q_{SL}^{(n)}(\mathbf{s}, \mathbf{l})$$

and

$$P_{ec}^{(n)}(R_1, R_2, W_{Y|UX}) = P_{Yec}^{(n)}(R_1, R_2, W_{Y|UX}) = \frac{1}{2^{R_1+R_2}} \sum_{\mathcal{M}_s\times\mathcal{M}_l} \sum_{\mathbf{y}:\varphi_{cn}(\mathbf{y})\neq(j,i)} W_{Y|X}^{(n)}(\mathbf{y}|\mathbf{u}, \mathbf{x}).$$

In this case, we can upper bound the source error exponent by

$$e\left(\frac{R_1}{\tau}, \frac{R_2}{\tau}, Q_{SL}\right) \le \min_{P_{SL}:\tau H_P(S,L)=R_1+R_2} D(P_{SL}\|Q_{SL}) = \max_{\rho\ge 0}\left[\rho\frac{R_1+R_2}{\tau} - E_{s1}(\rho, Q_{SL})\right], \tag{10.55}$$

which is obtained by viewing the two source encoders f_{sn} and g_{sn} as a joint encoder, where $E_{s1}(\rho, Q_{SL})$ is given by (9.92). Therefore, we can upper bound the tandem coding error exponent for the CS-AMAC system by

$$\begin{aligned} &E_T(Q_{SL}, W_{YZ|UX}, \tau) \\ &\le \sup_{R_1>0,R_2>0} \min\left\{\max_{\rho\ge 0}\left[\rho(R_1+R_2) - \tau E_{s1}(\rho, Q_{SL})\right], E_{sp}(R_1, R_2, W_{Y|UX})\right\} \end{aligned} \tag{10.56}$$

where $E_{sp}(R_1, R_2, W_{Y|UX})$ is an upper bound for the channel error exponent and is given by (9.91).

Example 10.5 Now consider the same binary CS Q_{SL} given in Example 9.1 such that

$$E_{s1}(\rho, Q_{SL}) = (1+\rho)\log_2\left\{\left[\left(\frac{2}{3}\right)^{\frac{1}{1+\rho}} + \left(\frac{1}{3}\right)^{\frac{1}{1+\rho}}\right](1-q)^{\frac{1}{1+\rho}} + 2\left(\frac{q}{2}\right)^{\frac{1}{1+\rho}}\right\},$$

and consider the same binary multiple access channel $W_{Y|UX}$ as in Example 9.1 with binary additive noise $P_F(F=1) = \epsilon$ $(0<\epsilon<1/2)$ such that

$$E_{sp}(R_1, R_2, W_{Y|UX}) = \min_{i=1,2}\max_{\rho\ge 0}[\widetilde{E}_i(\rho, W_{Y|UX}) - \rho\widehat{R}_i] = \max_{\rho\ge 0}[\widetilde{E}_i(\rho, W_{Y|UX}) - \rho(R_1+R_2)]$$

where $\widehat{R}_1 = R_1 + R_2$, $\widehat{R}_2 = R_2$, and

$$\widetilde{E}_1(\rho, W_{Y|UX}) = \widetilde{E}_2(\rho, W_{Y|UX}) = \rho - (1+\rho)\log_2\left(\epsilon^{\frac{1}{1+\rho}} + (1-\epsilon)^{\frac{1}{1+\rho}}\right).$$

It follows from (10.56) that upper bound for E_T only depends on the sum rate $R_1 + R_2$ and hence the upper bound can be reduced to

$$E_T(Q_{SL}, W_{YZ|UX}, \tau) \leq \sup_{R>0} \min \left\{ \max_{\rho \geq 0} \left[\rho R - \tau E_{s1}(\rho, Q_{SL})\right], \max_{\rho \geq 0} [\widetilde{E}_1(\rho, W_{Y|UX}) - \rho R \right\}.$$

In Fig. 10.22, we plot the lower bound for E_J from (9.96), and the above upper bound for E_T for different source and channel parameters. It is seen that for a large class of (q, ϵ) pairs with the same transmission rate τ, there is a considerable gap between the upper bound for E_T and the lower bound for E_J, which implies that JSCC can substantially outperforms tandem coding in terms of error exponent for many binary CS-AMAC systems with additive noise. In fact, from Fig. 10.22, we see that E_J almost doubles E_T for many (q, ϵ) pairs. When $E_J \approx 2E_T$ holds, it can be equivalently interpreted that, to achieve the same system error performance, JSCC only requires around half delay of the tandem coding, provided that the coding length is sufficiently large.

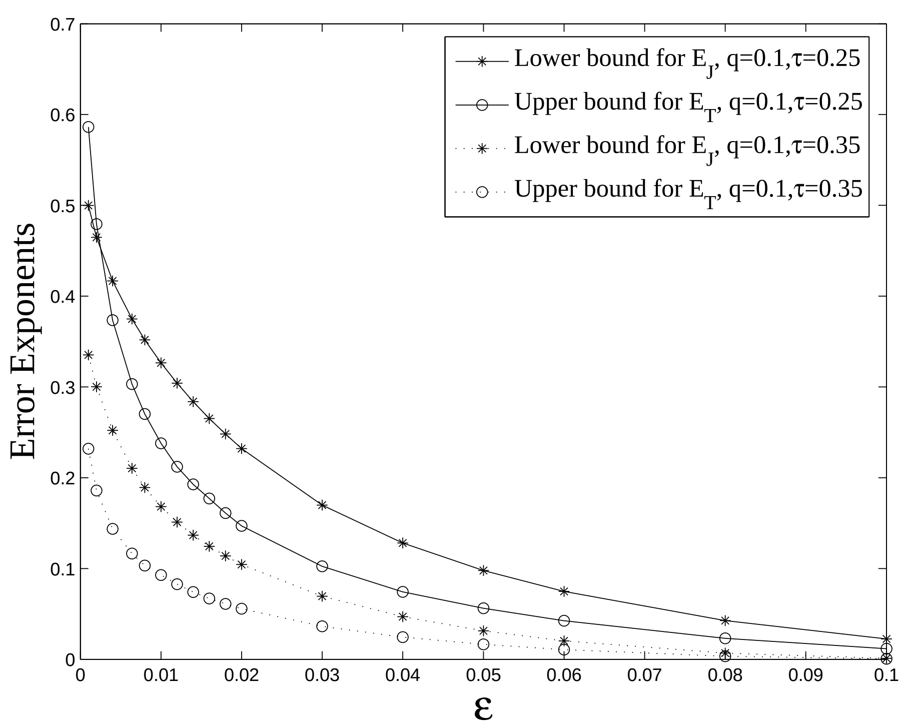

Figure 10.22: The lower bound of E_J vs the upper bound of E_T.

10.7 Conclusion

We derived a conceptual formula for the tandem coding error exponent for arbitrary discrete systems. We hence can provide a systematic comparison between E_J and E_T for DMS-DMC pairs and SEM source-channel pairs. We showed that JSCC can at most double the error exponent vis-a-vis tandem coding for these discrete source-channel pairs. However, we are not able to show that the same relation is still valid for the lossy discrete case. We established sufficient explicit conditions under which $E_J > E_T$ for discrete memoryless systems and SEM systems. Numerical results indicate that $E_J \approx 2E_T$ for a large class of DMS-DMC pairs and SEM source-channel pairs, hence illustrating the substantial potential benefit of JSCC over tandem coding. This benefit is also shown to result into a power saving gain of more than 2 dB for a binary DMS and a BPSK-modulated AWGN/Rayleigh channel with finite output quantization. We next partially extended our results to discrete memoryless systems with channel output feedback and source SI. Numerical results demonstrated that the JSCC is superior to the corresponding tandem coding error exponent for many cases.

We also derived an expression for the tandem coding exponent for Gaussian source-channel pairs provided that SDR ≥ 4 ($\approx$ 6dB). The tandem Gaussian exponent has a similar form as the discrete tandem error exponent. As in the discrete cases, the JSCC exponent is observed to be considerable larger than the tandem exponent for a large class of Gaussian source-channel pairs.

We then derive a formula for the tandem coding error exponent for the 2-user asymmetric source-channel systems in terms of the 2-user source and channel exponents. To exploit the advantage of JSCC over tandem coding for the 2-user system, we show that the tandem coding exponent can never be larger than the JSCC exponent, and we numerically show that there is a considerable gain of the JSCC error exponent over the tandem coding error exponent for a large class of binary CS-AMAC systems with additive noise.

Chapter 11

Conclusion

11.1 Contributions

The main purpose of the thesis was to study the JSCC reliability function for single-user and multi-user systems. The following techniques were used to establish and analyze lower and upper bounds for the JSCC reliability function for different systems.

(a) We generalized Csiszár's type packing lemma for a 2-dimensional type setting and we proposed a new continuous-type class, the Laplacian-type class, to analyze the memoryless Laplacian sources.

(b) Several approaches were used for upper bounding the JSCC reliability function:

- We employed the type counting argument to obtain an upper bound in terms of the source and channel error exponents (e.g. Theorems 6.1, 7.5 and 9.3);
- We employed the strong converse JSCC theorem and the artificial source-channel pairs to lower bound the probability of decoding error/exceeding distortion (e.g., Theorems 7.2, 8.2, and 8.6).

(c) To establish a lower bound for the JSCC reliability function, the following basic bounding techniques were used:

- We employed a joint type packing lemma combined with generalized mutual information decoders/conditional mutual information decoders to upper bound the probability of error (e.g., Theorems 6.4 and 9.1);
- We modified Zigangirov's iteration algorithm to upper bound the probability of error in the presence of feedback (e.g., Theorem 6.2).
- We employed the random-binning encoder together with a minimum conditional entropy decoder in the two-stage JSCC to upper bound the probability of error (e.g., Theorem 6.4);
- We employed and modified Gallager's classical random-coding bound for lossless JSCC with channel input cost constraints (e.g., Theorems 8.3, 8.6 and 8.7);
- We employed and extended the two-stage (Δ-admissible quantization plus lossless JSCC) JSCC scheme to upper bound the probability of excess distortion (e.g., Theorems 8.3, 8.6 and 8.7);
- We employed superposition coding for the transmission of two CS over a 2-user channel to enlarge achievable error exponents (e.g., Theorem 9.1).

(d) Finally, we employed the properties of conjugate functions and Fenchel duality theorem to evaluate these upper and lower bounds, and to show the relation between the results of Gallager and Csiszár (Observation 5.1).

Now let us, chapter by chapter, go through the main contributions of the thesis.

In Chapter 3, we generalized Csiszár's type packing lemma from a single type setting to a 2-dimensional joint type setting. We developed Laplacian-type class and formulated a type covering lemma for this continuous type class.

In Chapter 4, we illustrated the (Fenchel transform) relations between source/channel reliability functions (error exponents and excess distortion exponents) and the source/channel functions. Several Fenchel transforms were also derived; see Table 4.1.

In Chapter 5, by employing the Fenchel duality theorem, we established equivalent expressions in terms of the difference of source and channel functions for Csiszár's random-coding source-channel exponent $\underline{E}_{Jr}$, sphere-packing source-channel exponent $\overline{E}_{Jsp}$, and expurgated source-channel exponent $\underline{E}_{Jex}$ for discrete memoryless systems. A sufficient and necessary condition for which $\underline{E}_{Jr} = \overline{E}_{Jsp}$ was given. We also examined this condition to DMS-DMC pairs when the channel admits a symmetric distribution. We derived a sufficient and necessary condition under which $\underline{E}_{Jex}$ strictly improves $\underline{E}_{Jr}$. This condition has been examined for equidistant channels. Using a similar optimization technique, we also derived equivalent expressions for Csiszár's upper and lower bounds for the lossy JSCC excess distortion exponent for binary input channels under Hamming source distortion.

In Chapter 6, we first established upper and lower bounds for the JSCC error exponent $E_{J,fb}$ for DMS-DMC system with perfect feedback. It was demonstrated by numerical examples that feedback can strictly increase the JSCC error exponent. The source side information was next considered. When the source side information is available at the decoder only, we established a lower bound for the JSCC error exponent E_J^{SID}. We established the JSCC theorem and proved that the separation principle holds for this scenario. We showed that, in the presence of the source side information at the decoder, the JSCC error exponent can be strictly larger than the one without any side information. A sufficient condition for which the source side information can improve the JSCC error exponent for a binary source and a symmetric channel has been driven.

In Chapter 7, we mainly focused on establishing upper bounds for the JSCC error exponent for SEM source-channel pairs. We first established a sphere-packing type computable upper bound in terms of Rényi entropy rates of artificial Markovian source and noise processes. We then established a conceptual upper bound in terms of SEM source and SEM channel error exponents by introducing Markov types. By comparing the sphere-packing type upper bound with Gallager's lower bound, when the later one is specialized to SEM source-channel systems, we obtained a sufficient and necessary condition for which the JSCC error exponent is exactly determined by the upper and lower bounds. By using the Fenchel duality theorem, equivalent expressions for these bounds were derived as in the case of memoryless systems. As by-products, we obtained upper bounds for the SEM source error exponent and the SEM channel error exponent.

In Chapter 8, we first established upper and lower bounds for the JSCC excess distortion $E_J^{\Delta,\mathcal{E}}$ for an MGS (under the squared-error distortion) and an MGC (with the quadratic power constraint). Equivalent expressions for these bounds were given, via which we obtained a sufficient and necessary condition for the upper and lower bounds to coincide. We next extended the results regarding the upper and lower bounds for the JSCC excess distortion exponent to a system consisting of an MLS under magnitude-error distortion and an MGC. We also proved a lower bound for the JSCC exponent for a general class of continuous source-channel pairs when the distortion is a metric and if there exists an element $s_o \in \mathbb{R}$ with $\mathbb{E}\exp[td(s, s_o)] < \infty$ for all $t \in (-\infty, +\infty)$, where the expectation is taken over the source distribution.

In Chapter 9, we established universally achievable error exponent pairs for transmitting two correlated sources over a 2-user asymmetric discrete channel. Lower and upper bounds for the system overall JSCC error exponent E_J as well as the JSCC theorem were established. We proved that the separation principle holds for the asymmetric 2-user scenario. We also applied these results to CS-AMAC systems and CS-ABC systems. We finally evaluated our bounds for E_J for certain CS-AMAC systems when the channel admits a symmetric distribution by deriving equivalent expressions for the lower and upper bounds for the system JSCC error exponent in terms of source and channel functions.

In Chapter 10, we first derived a formula for the tandem coding error exponent E_T for discrete system with arbitrary memory. We then compared the JSCC error exponent E_J with the tandem coding error exponent E_T for DMS-DMC pairs and SEM source-channel pairs. For both cases, we have shown that E_J can at most double E_T. Several computable sufficient conditions for which $E_J > E_T$ was established for the discrete memoryless and Markovian systems. The numerical results demonstrated that these conditions are satisfied by a large class of source-channel pairs, and for many cases E_J can be close to twice E_T. Such exponent improvement due to JSCC translates into a power saving gain of more than 2 dB for a binary DMS and a BPSK-modulated AWGN/Rayleigh channel with finite output quantization. We next derived the tandem excess distortion exponent $E_T^{\Delta,\mathcal{E}}$ for MGS-MGC pairs when the distortion threshold is less than 1/4 of the source variance. It was seen that

$E_T^{\Delta,\mathcal{E}}$ admits a similar expression as E_T in the discrete case. By numerically comparing the lower bound of the JSCC excess distortion exponent $E_J^{\Delta,\mathcal{E}}$ with $E_T^{\Delta,\mathcal{E}}$, we observed that $E_J^{\Delta,\mathcal{E}}$ substantially outperforms $E_T^{\Delta,\mathcal{E}}$ for many MGS-MGC pairs. We derived a formula for the tandem coding error exponent of the asymmetric 2-user system. Numerical examples show that for a large class of systems consisting of two correlated sources and an asymmetric multiple-access channel with additive noise, the JSCC error exponent considerably outperforms the corresponding tandem coding error exponent.

11.2 Suggestions for Future Research

Since the JSCC reliability function reflects the best (asymptotic) performance of transmitting a single source (or multiple sources) over a communication channel, it provides an important information-theoretic limit that points out certain systems for which a search for good joint codes might prove fruitful. On the other hand, determining and bounding the JSCC reliability function is one of the most challenging problems in Shannon theory. In this section, we will touch on some open problems and indicate suggestions for future research.

First of all, important work has to be done with respect to pursuing tighter upper and lower bounds for E_J (or E_J^{Δ}, $E_J^{\Delta,\mathcal{E}}$). Following the conceptual upper bound for the JSCC error exponent for DMS-DMC and SEM source-channel systems (which states that E_J is upper bounded by the smallest sum of the source and channel error exponents), upper bounding the JSCC error exponent for the two discrete systems is strongly related to upper bounding the channel error exponent, since a new upper bound for the channel error exponent leads to a new JSCC error exponent upper bound. For instance, recent works (see [14], [23] and the references therein) have proved tighter upper bounds for the BSC. It is shown that these sharpened upper bounds coincide with the DMC random-coding lower bound $E_r(R, W_{Y|X})$ for some interval directly below the channel critical rate (in other words, it is shown that for the BSC with its ε above a certain threshold, $E_r(R, W_{Y|X}) = E(R, W_{Y|X})$ for $R_1 \leq R \leq C(W_{Y|X})$ where R_1 can be *less* than $R_{cr}(W_{Y|X})$). Therefore, by using the new upper bound for BSC error exponent in the bound

$$E_J(Q_S, W_{Y|X}, \tau) \le \inf_R \left[\tau e\left(\frac{R}{\tau}, Q_S\right) + E(R, W_{Y|X})\right]$$

we can enlarge Region **B** in Fig. 5.4., i.e., we can determine E_J for more binary source and BSC pairs.

In Chapter 5, we examined Csiszár's expurgated lower bound $\underline{E}_{Jex}$ for a DMS and a DMC with zero-error capacity equal to 0 if $E_{ex}(R, W_{Y|X}) = \max_{P_X} E_{ex}(R, P_X, W_{Y|X})$ is attained for a P_X not depending on R. We wonder whether the expurgated source-channel lower bound holds for arbitrary DMS-DMC pairs, and more generally, whether there exists an expurgated-type bound for other systems, say SEM source-channel systems, or MGS-MGC systems.

In Chapter 6, we mentioned two interesting problems for future study: how to establish an upper bound for the JSCC error exponent with source SI at decoder, and we wonder whether the lower bound $\underline{E}_J^{SID}$ still holds if the random-coding channel exponent $E_r(\tau H_{P_S}(S), W_{Y|X})$ is replaced by the expurgated channel exponent $E_{ex}(\tau H_{P_S}(S), W_{Y|X})$?

We also mention the need for further study with respect to the tandem coding excess distortion exponent. After all, most of the applications used today are designed for tandem systems. In Chapter 10, we proved that the formula for the tandem error exponent for discrete systems, which is expressed by the max min of the source and channel exponents, is still valid for MGS-MGC systems when the distortion threshold Δ is less than 1/4 of the source variance. It might be not difficult to extend this result to MLS-MGC systems due to the similarity of Gaussian and Laplacian distributions, but the proof for the converse part of the exponent (Theorem 10.12), which relies on the geometric property of Gaussian density function, cannot be applied to other sources, say DMS's with Hamming distortion measure. We expect that a unified approach irrespective of the distortion measure and the source distribution can be used to determine the tandem excess distortion exponent.

Additional results might be expected for other source-channel systems, e.g., for transmitting a DMS through a time-varying channel, and for transmitting an MGS over a fading channel. As Shannon's source-channel separation theorem breaks down and the achievable rate region as well as the proper method of coding are unknown for general multi-terminal source-channel coding systems, it is important to investigate the JSCC reliability function for other multi-terminal systems. For example, since we do not have a single-letter characterization of the JSCC theorem for transmitting correlated sources over a multi-terminal channel, say transmitting three sources (one common source plus two private sources) over a symmetric multiple-access channel, establishing an achievable lower bound for the JSCC reliability function is particularly meaningful.

Finally, note that the JSCC reliability function can be used as a tool (particularly when it admits a simple analytical expression) for the construction of high-performing JSCC techniques and JSC modulation constellations for communication systems, (e.g., see [52] for such a study involving only the channel random-coding error exponent).

Bibliography

[1] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, May 1983.

[2] R. Ahlswede, "Extremal properties of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 36, pp. 166–171, Jan. 1990.

[3] F. Alajaji, N. Phamdo, N. Farvardin, and T. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inform. Theory*, vol. 42, No. 1, pp. 230–239, Jan. 1996.

[4] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech and Audio Processing*, vol. 4, no. 5, pp. 325–336, Sept. 1996.

[5] F. Alajaji and N. Phamdo, "Soft-decision COVQ for Rayleigh-fading channels," *IEEE Commun. Lett.*, vol. 2, pp. 162–164, June 1998.

[6] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. 44, No. 3, pp. 1041–1056, May 1998.

[7] E. Arikan and N. Merhav, "Joint source-channel coding and guessing with application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 44, No. 5, pp. 1756–1769, Sep. 1998.

[8] S. Arimoto, "'On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 19, pp. 357–359, May 1973.

[9] S. Arimoto, "Computation of random coding exponent functions," *IEEE Trans. Inform. Theory*, vol. 22, pp. 665–671, Nov. 1976.

[10] E. A. Arutyunyan (Haroutunian), "Lower bound for the error probability of multiple-access channels," *Probl. Pered. Inform.*, vol. 11, pp. 23–36, April–June 1975.

[11] E. A. Arutyunyan (Haroutunian), "Lower bound for error probobility in channel with feedback," *Probl. Contr. Inform. Theory*, Vol. 13, No. 2, pp. 36-44, 1977.

[12] A. E. Ashikhmin, A. Barg, and S. N. Litsyn, "A new upper bound on the reliability function of the Gaussian channel,"*IEEE Trans. Inform. Theory*, vol. 46, No. 6, pp. 1945–1961, Sep. 2000.

[13] E. Ayanoğlu and R. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, vol. 33, pp. 855–865, Nov. 1987.

[14] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4237–4246, Dec. 2005.

[15] T. Berger, *Rate Distortion Theory*, Prentice-Hall, New Jersey, 1971.

[16] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, pp. 197–206, Mar. 1973.

[17] D. P. Bertsekas, with A. Nedić and A. E. Ozdagler, *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA, 2003.

[18] P. Billingsley, *Ergodic Theory and Information*, Huntington, New York, 1978.

[19] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, July 1974.

[20] K. De Bruyn, V. V. Prelov, and E. Van Der Meulen, “Reliable transmission of two correlated sources over an asymmetric multiple-acess channel,” *IEEE Trans. Inform. Theory*, vol. 33, pp. 716–718, Sep. 1987.

[21] Brian D. Bunday, *Basic Optimisation Methods*. London: Arnold, 1984.

[22] M. V. Burnashev, “Reliability function of a binary symmetrical channel with feedback,” *Probl. Contr. Inform. Theory*, Vol. 24, No. 1, pp. 3-10, 1985.

[23] M. V. Burnashev, “Sharpening of an upper bound for the reliability function of a binary symmetric channel,” *Probl. Contr. Inform. Theory*, Vol. 41, No. 4, pp. 901-918, 2005.

[24] P.-N. Chen and F. Alajaji, “Optimistic Shannon coding theorems for arbitrary single-user systems,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 2623–2629, Nov. 1999.

[25] P.-N. Chen and F. Alajaji, “Csiszár’s cutoff rates for arbitrary discrete sources,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 330–337, Jan. 2001.

[26] Q. Chen and T. R. Fisher, “Image coding using robust quantization for noisy digital transmission,” *IEEE Trans. Image Processing*, vol. 7, pp. 496–505, Apr. 1998.

[27] S. Choi and S. S. Pradhan, “A graph-based framework for transmission of correlated sources over broadcast channels,” *IEEE Trans. Inform. Theory*, to appear 2008.

[28] T. M. Cover, A. El Gamal and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Trans. Inform. Theory*, vol. 26, no. 6, Nov. 1980.

[29] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: Wiley, 2nd Ed. 2006.

[30] I. Csiszár, “Joint source-channel error exponent,” *Probl. Contr. Inform. Theory*, vol. 9, pp. 315–328, 1980.

[31] I. Csiszár, “On the error exponent of source-channel transmission with a distortion threshold,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 823–828, Nov. 1982.

[32] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[33] I. Csiszár, "The method of types," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2505–2522, Nov. 1998.

[34] L. D. Davisson, G. Longo and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inform. Theory*, vol. 27, no. 4, pp. 431–438, July 1981.

[35] G. Dueck, "A note on the multiple access channel with correlated sources," *IEEE Trans. Inform. Theory*, vol. 27, no. 2, Mar. 1981.

[36] J. G. Dunham and R. M. Gray, "Joint source and noisy channel trellis encoding," *IEEE Trans. Inform. Theory*, vol. 27, pp. 516–519, July 1981.

[37] A. G. D'yachkov, "Upper bounds on the error probability for discrete memoryless channels with feedback," *Probl. Contr. Inform. Theory*, Vol. 11, No. 4, pp. 13-28, 1975.

[38] A. A. EL Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, pp. 166–169, March. 1979.

[39] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799–809, July 1990.

[40] T. Fine, "Properties of a optimum digital system and applications," *IEEE Trans. Inform. Theory*, vol. 10, pp. 443–457, Oct. 1964.

[41] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no.1, pp. 3–18, Jan. 1965.

[42] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[43] R. G. Gallager, "Source coding with side information and universal coding", Technical Report LIDS-P-937, *Mass. Instit. Tech.*, 1976.

[44] M. Gastpar, B. Rimoldi and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1147–1158, May 2003.

[45] J. D. Gibson and T. R. Fisher, "Alphabet-constrained data compression," *IEEE Trans. Inform. Theory*, vol. 28, pp. 443–457, May 1982.

[46] R. M. Gray and D. S. Ornstein, "Sliding-block joint source/noisy-channel coding theorems," *IEEE Trans. Inform. Theory*, vol. 22, pp. 682–690, Nov. 1976.

[47] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sep. 1995.

[48] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrary correlated sources," *IEEE Trans. Inform. Theory*, vol. 33, no. 5, pp. 641–650, Sep. 1987.

[49] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2003.

[50] M. E. Hellman, "Convolutional source encoding," *IEEE Trans. Inform. Theory*, vol. 21, pp. 651–656, Nov. 1975.

[51] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sep. 1997.

[52] J. Huang, S. Meyn, M. Médard, "Error exponents for channel coding with application to signal constellation design," *IEEE J. Selected Areas in Comm.*, vol. 24, no. 8, pp. 1647–1661, Aug. 2006.

[53] S. Ihara, *Information Theory for Continuous Syetems*, World Scientific, 1993.

[54] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with fidelity criterion," *IEICE Tran. Fundamentals,* vol. E83-A, No. 10, pp 1891-1897, Oct. 2000.

[55] S. Ihara and M. Kubo, "Error exponent of coding for stationary memoryless sources with a fidelity criterion," *IEICE Tran. Fundamentals,* vol. E88-A, no. 5, pp 1339-1345, May. 2005.

[56] K. Iwata and Y. Oohama, "Information-spectrum characterization of multiple-access channels with correlated sources," *IEICE Trans. Fundamentals*, vol. E88-A, no. 11, pp. 3196–3202, Nov. 2005.

[57] F. Jelinek, *Probabilistic Information Theory*, New York, McGraw Hill, 1968.

[58] W. Kang and S. Ulukus, "An outer bound for multiple access channels with correlated sources," *Proceedings of the 2006 IEEE Int'l. Symp. Inform. Theory*, Seattle, USA, July 2006.

[59] J. Körner and K. Marton, "Genreal broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. 23, pp. 60–64, Jan. 1977.

[60] J. Körner and A. Sgarro, "Univerally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. 26, pp. 670–679, Nov. 1980.

[61] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Eng. Jpn.*, vol. 67-B, no. 4, pp. 39–47, 1984.

[62] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Transactions on Communication Technology*, vol. COM-17, pp. 291–302, Apr. 1969.

[63] G. Longo and A. Sgarro, "The source coding theorem revisited: a combinatorial approach," *IEEE Trans. Inform. Theory*, vol. 25, pp. 544–548, Sept. 1979.

[64] J. Lim and D. L. Neuhoff, "Joint and tandem source-channel coding with complexity and delay constraints," *IEEE Trans. Commun.*, vol. 51, pp. 757–766, May 2003.

[65] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1969.

[66] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

[67] J. L. Massey, "Joint source and channel coding," in *Communications and Random Process Theory,* J. K. Skwirzynski, ed., The Netherlands: Sijthoff and Nordhoff, pp. 279–293, 1978.

[68] D. Miller and K. Rose, "Combined source-channel vector quantization using deterministic annealing," *IEEE Trans. Commun.*, vol. 42, pp. 347–356, Feb.-Apr. 1994.

[69] N. Merhav and S. Shamai (Shitz), "On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2844–2855, Nov. 2003.

[70] U. Mittal and N. Phamdo, "Duality theorems for joint source-channel coding," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1263–1275, July 2000.

[71] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. 27, pp. 1644-1659, Nov. 1979.

[72] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inform. Theory*, vol. 31, no. 3, pp. 360–365, May. 1985.

[73] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1908–1919, Nov. 1994.

[74] Nam Phamdo and Fady Alajaji, "Soft-decision demodulation design for COVQ over white, colored, and ISI Gaussian channels," *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1499–1506, Sep. 2000.

[75] C. Pimentel, T. H. Falk and L. Lisbôa, "Finite-state Markov modeling of correlated Rician fading channels," *IEEE Trans. Vehicular Technology*, vol. 53, pp. 1491–1501, Set. 2004.

[76] S. S. Pradhan, S. Choi and K. Ramchandran, “A graph-based framework for transmission of correlated sources over multiple access channels,” *IEEE Trans. Inform. Theory*, vol. 53, No. 12, pp. 4583–4604, Dec. 2007.

[77] Z. Rached, F. Alajaji and L. L. Campbell, “Rényi’s divergence and entropy rates for finite alphabet Markov sources,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 1553–1561, May. 2001.

[78] R. C. Reininger and J. D. Gibson, “Distributions of the two-dimensional DCT coefficients for images,” *IEEE Trans. Commun.*, vol. 31, pp. 835–839, June 1983.

[79] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[80] H. L. Royden, *Real Analysis*, Third Edition, New York, 1988.

[81] K. Sayood and J. C. Borkenhagen, “Use of residual redundancy in the design of joint source/channel coders,” *IEEE Trans. Commun.*, vol. 39, pp. 838–846, June 1991.

[82] E. Seneta, *Non-negative Matrices and Markov Chains*, New York: Springer-Verlag, 1981.

[83] C. E. Shannon,“A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and pp. 623-656, Jul. and Oct. 1948.

[84] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611-656, 1959.

[85] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65-103, Feb. 1967.

[86] S. Shamai, and S. Verdú, “Capacity of channels with uncoded side information,” *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 587–600, Sep.-Oct. 1995.

[87] S. Shamai, S. Verdú, and R. Zamir, “Systematic lossy source/channel coding,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[88] A. Yu. Sheverdyaev, "Lower bound for error probability in a discrete memoryless channel with feedback," *Probl. Contr. Inform. Theory*, Vol. 18, No. 4, pp. 5-15, 1982.

[89] M. Skoglund and P. Hedelin, "Hadamard-based soft decoding for vector quantization over noisy channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 515–532, Mar. 1999.

[90] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1293–1307, May 1999.

[91] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926-943, June 1992.

[92] G. Taricco, "On the capacity of the binary input Gaussian and Rayleigh fading channels," *Eur. Trans. Telecommun.*, vol. 7, no. 2, Mar.-Apr. 1996.

[93] V. A. Vaishampayan and N. Farvardin, "Joint design of block source codes and modulation signal sets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1230–1248, July 1992.

[94] Karol Vašek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, pp. 318–329, Nov. 1980.

[95] S. Vembu, S. Verdú and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, pp. 44–54, Jan. 1995.

[96] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[97] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, Inc., 1979.

[98] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. J. Weinberger, "Universal discrete denoising: known channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 5–28, Jan. 2005.

[99] K. A. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.

[100] L. Zhong, F. Alajaji, and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2851–2840, Aug. 2007.

[101] L. Zhong, F. Alajaji, and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Vehicular Technology*, vol. 57, no. 1, pp. 79–89, Jan. 2008.

[102] Y. Zhong, F. Alajaji, and L. L. Campbell, "When is joint source-channel coding worthwhile: an information theoretic perspective," in *Proc. 22nd Bienn. Symp. Commun.*, Canada, pp. 121–123, June 2004.

[103] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the computation of the joint source-channel error exponent for memoryless systems," in *Proc. 2004 IEEE Int'l. Symp. Inform. Theory,* p. 477, June-July 2004.

[104] Y. Zhong, F. Alajaji, and L. L. Campbell, "When does joint source-channel coding outperform tandem coding for systems with Markovian memory?" in *Proc. 2005 Canadian Workshop on Information Theory*, June 2005.

[105] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for systems with memory," in *Proc. 2005 IEEE Int'l. Symp. Inform. Theory*, Adelaide, Australia, pp. 1386–1390. Sept. 2005.

[106] Y. Zhong, F. Alajaji and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems: computation and comparison with separate coding," Mathematics and Engineering Technical Report, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, December 2005.

[107] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1450–1468, April 2006.

[108] Y. Zhong, F. Alajaji, and L. L. Campbell, "A type covering lemma and the excess distortion exponent for coding memoryless Laplacian sources," in *Proc. 23nd Bienn. Symp. Commun.*, Kingston, Canada, pp. 100–103, June 2006.

[109] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent of discrete communication systems with Markovian memory," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4457–4472, Dec. 2007.

[110] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the excess distortion exponent for memoryless Gaussian source-channel pairs," in *Proc. 2006 IEEE Int'l. Symp. Inform. Theory,* Seattle, USA, July 2006.

[111] Y. Zhong, F. Alajaji and L. L. Campbell, "Coding error exponents for asymmetric two-user discrete memoryless source-channel systems," Mathematics and Engineering Technical Report, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, March 2007.

[112] Y. Zhong, F. Alajaji, and L. L. Campbell, "Source side information can increase the joint source-channel coding error exponent" in *Proc. 2007 Canadian Workshop on Information Theory*, Edmonton, Canada, pp. 140–143, June 2007.

[113] Y. Zhong, F. Alajaji, and L. L. Campbell, "Error exponents for asymmetric two-user discrete memoryless source-channel systems," in *Proc. 2007 IEEE International Symposium on Information Theory (ISIT'07)*, Nice, France, pp. 1736–1740, June 2007.

[114] G.-C. Zhu, F. Alajaji, J. Bajcsy and P. Mitran, "Transmission of non-uniform memoryless sources via non-systematic Turbo codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1344–1354, Aug. 2004.

[115] K. Sh. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Contr. Inform. Theory*, Vol. 6, No. 2, pp. 87-92, 1970.