On the Joint Source-Channel Reliability Function for Memoryless Communication Systems

by

Yangfan Zhong

A project submitted to the
Department of Mathematics and Statistics
in conformity with the requirements
for the degree of
Master of Science (Eng.)

Queen's University
Kingston, Ontario, Canada
November 2003

Abstract

In this project we analytically compute Csiszár's random-coding lower bound for the joint source-channel coding reliability function of discrete memoryless systems using basic Lagrange optimization techniques. Given a discrete memoryless source (DMS) with generic distribution Q and alphabet $\mathcal{S}$, and a discrete memoryless channel (DMC) with transition probability W, input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, the lossless joint source-channel reliability function $E_J(Q, W)$ is uniquely determined by Q and W, and when some conditions in terms of Q and W are satisfied, it is shown that $E_J(Q, W)$ can be computed exactly. The rate that achieves $E_J(Q, W)$ is also provided. The expurgated lower bound to $E_J(Q, W)$ is also studied. A byproduct of our results is the surprising observation that Csiszár's joint source-channel random-coding lower bound is identical to Gallager's bound obtained in 1964. We also examine Csiszár's upper bound based on the sphere-packing upper bound to the channel reliability function to $E_J(Q, W)$ using a similar approach. We then compare the reliability function $E_J(Q, W)$ with the tandem coding exponent $E_T(Q, W)$ and give sufficient conditions for which $E_J(Q, W)$ is strictly larger than $E_T(Q, W)$. Numerical examples show that $E_J(Q, W)$ can be substantially larger than $E_T(Q, W)$. Finally, we investigate the computation of the lossy joint source-channel exponent $E_J^{\Delta}(Q, W)$ with distortion threshold Δ under the Hamming distortion measure. For a binary DMS $\{q, 1-q\}$ and a DMC, lower and upper bounds for $E_J^{\Delta}(Q, W)$ are given.

Acknowledgements

First and foremost I would like to thank my parents for the encouragement and support they provided me throughout my life. They fostered my intense interest in mathematics and physics.

I would like to express my sincere gratitude to my supervisor, Dr. Fady Alajaji, and co-supervisor, Dr. L. Lorne Campbell, for their advice, guidance and valuable help throughout this project. Their expertise, understanding, and patience considerably enriched my graduate studies experience.

I would like to thank my friends: Guang-Chong Zhu, Libo Zhong, Yadong Wang, Firouz Behnamfar and Azadeh Moghtaderi, for their generous help when I had any difficulty in using LaTeX.

Finally, I recognize that my study at Queen's University would not have been possible without the financial assistance of my two supervisors, the Department of Mathematics and Statistics and the School of Graduate Studies.

Contents

List of Figures

Chapter 1

Introduction

In this chapter we present a review of the articles upon which our research is based. We then specify the main contributions of the project. Finally, we outline the general flow of the project.

1.1 Literature Review

The joint source-channel coding theorem was first introduced by Gallager [4, Problem 5.16] as a take-home quiz in 1964; also see Jelinek [6, Section 6.8]. This result also indicates that the joint source-channel coding can lead to a larger exponent than the tandem coding exponent, i.e., the exponent obtained by concatenating optimal source and channel codes. In 1980, Csiszár [1] established a lower bound (based on the random-coding channel exponent) and an upper bound for the joint source-channel coding reliability function $E_J(Q, W)$. He showed that the upper bound, which is expressed as the minimum of the sum of $e(R, Q)$ and $E(R, W)$ over R, where $e(R, Q)$ is the source reliability function [1, 6, 7, 8] and $E(R, W)$ is the channel reliability function [1, 4, 6], is tight if the latter minimum is attained for an R not less than the critical rate of the channel. Two years later, Csiszár [2] showed that the lower bound can be improved by an expurgated one under some conditions. He also extended his results to the joint source-channel coding with a distortion threshold. The source coding reliability function with general fidelity criteria was studied in [1, 9, 11, 12].

1.2 Contribution

The contributions of this project are as follows:

- Computing Csiszár's lower and upper bounds for the joint source-channel reliability function for memoryless systems. A parametric solution is provided for arbitrary discrete memoryless source (DMS) and discrete memoryless channel (DMC) pairs.
- Establishing the condition for which the joint source-channel exponent can be exactly determined.
- Providing the condition for which the source-channel lower bound can be improved by an expurgated bound, and determining the parametric solution for the improved bound.
- Revealing the relation between Gallager's joint source-channel coding theorem and Csiszár's lower bound.
- Comparing the joint source-channel exponent with the tandem source-channel coding exponent.
- Computing the lower and upper bounds for the joint source-channel reliability function for memoryless systems with a distortion threshold. A parametric solution is provided for binary DMS over DMC under the Hamming distortion measures.

1.3 Project Overview

We first review some fundamentals such as basic concepts, important properties and useful results for the source and channel reliability functions in Chapter 2.

Joint source-channel coding theory is briefly introduced in Chapter 3. Here we mainly concentrate on introducing Csiszár's result about the (lossless) joint source-channel exponent and the more general exponent with fidelity criteria.

In Chapter 4, we study the analytical computation of the joint source-channel reliability function based on Csiszár's work [1, 2]. Our main objective is to give a general solution for arbitrary DMS and DMC pairs. Some interesting byproduct results are obtained such as exact exponent expressions for DMS's over quasi-symmetric DMC's. We also apply Arimoto's [10] algorithm for the numerical computation of the lower bound to the joint source-channel exponent. After computing the lower bound, the expurgated lower bound and the upper bound of the joint source-channel exponent, we relate Gallager's joint source-channel coding theorem with Csiszár's lower bound, and we evaluate and compare the joint source-channel exponent with the tandem source-channel exponent. In the last part of the chapter, we study the computation of the joint source-channel exponent for binary DMS and DMC pairs under the Hamming distortion measure.

In Chapter 5 we give a summary of the project.

Note that throughout the project, the block lengths of source messages and channel codewords are assumed to be the same merely for the sake of convenience. The results carry over in an obvious way to the case when these lengths are k and n, respectively, with k/n converging to an arbitrary positive constant. Furthermore, all logarithms and exponentials are in base 2.

Chapter 2

Preliminaries: Reliability Function for Source and Channel Coding

In this chapter, some basic and classical results on the source coding and channel coding reliability functions are reviewed.

2.1 Source Reliability Function

Shannon's first coding theorem states that for a DMS with distribution Q only $H(Q)+\varepsilon$ bits per source symbol are needed to encode the source with arbitrarily small probability of error, provided the block length n of encoded source symbols is allowed to be sufficiently large. The source coding exponent was developed to determine the asymptotics of the smallest possible probability of incorrect decoding as a function of the coding rate

$$R \triangleq \frac{\log M}{n}$$

where M is the number of codewords.

2.1.1 Source Coding Exponent

Definition 2.1 *An (n, M) block code for a DMS with generic distribution Q and alphabet $\mathcal{S}$ is a pair of mappings:*

$$f : \mathcal{S}^n \longrightarrow \{1, 2, ..., M\} \quad \text{and} \quad \varphi : \{1, 2, ..., M\} \longrightarrow \mathcal{S}^n.$$

The average error probability of the code is

$$P_e(n, M) \triangleq \sum_{s^n : s^n \neq \varphi(f(s^n))} Q^n(s^n), \tag{2.1}$$

where $s^n \triangleq (s_1 \cdots s_n) \in \mathcal{S}^n$. The code rate is defined as

$$R(n, M) \triangleq \frac{1}{n} \log M \quad \text{bits/source symbol.}$$

Definition 2.2 (*Source Reliability Function*) *For any $R > 0$, the source reliability function $e(R, Q)$ of the source Q is defined as the largest number e for which there exists a sequence of n-length block codes (n, M) with*

$$e \leq \liminf_{n \to \infty} -\frac{1}{n} \log P_e(n, M) \tag{2.2}$$

and

$$R \geq \limsup_{n \to \infty} \frac{1}{n} \log M. \tag{2.3}$$

Theorem 2.1 [8] *Given a DMS with generic distribution Q, the source reliability function satisfies*

$$e(R, Q) = \min_{P : H(P) \geq R} D(P \parallel Q), \tag{2.4}$$

where $H(\cdot)$ is the entropy and $D(\cdot \parallel \cdot)$ is the relative entropy (or information divergence).

Theorem 2.2 *$e(R, Q)$ is strictly increasing, convex and hence a continuous function of R for $H(Q) \leq R \leq \log |\mathcal{S}|$, and is zero for $R \leq H(Q)$, where $|\mathcal{S}|$ is the size of the source alphabet.*

Proof: Suppose $H(Q) \leq R_1 < R_2 \leq \log|\mathcal{S}|$, then,

$$\{P : H(P) \geq R_2\} \subset \{P : H(P) \geq R_1\}.$$

It directly follows that $e(R_1, Q) \leq e(R_2, Q)$, and $e(R, Q)$ is an increasing function of R. Now suppose that in $e(R_1, Q) = \min_{P:H(P)\geq R_1} D(P \parallel Q)$, the minimum is attained when $P = P_1$, and that in $e(R_2, Q) = \min_{P:H(P)\geq R_2} D(P \parallel Q)$, the minimum is attained when $P = P_2$. Then for $\forall\lambda \in (0, 1)$,

$$\begin{aligned}
\lambda e(R_1, Q) + (1-\lambda)e(R_2, Q) &= \lambda D(P_1 \parallel Q) + (1-\lambda)D(P_2 \parallel Q) \\
&= -\sum_{s\in\mathcal{S}}[\lambda P_1(s) + (1-\lambda)P_2(s)]\log Q(s) \\
&\quad -[\lambda H(P_1) + (1-\lambda)H(P_2)] \\
&\geq -\sum_{s\in\mathcal{S}}[\lambda P_1(s) + (1-\lambda)P_2(s)]\log Q(s) \\
&\quad -H(\lambda P_1 + (1-\lambda)P_2) \\
&= D[(\lambda P_1 + (1-\lambda)P_2) \parallel Q],
\end{aligned}$$

where the inequality follows from the concavity of entropy function $H(\cdot)$. Moveover,

$$H(\lambda P_1 + (1-\lambda)P_2) \geq \lambda H(P_1) + (1-\lambda)H(P_2) \geq \lambda R_1 + (1-\lambda)R_2.$$

Since P_1 achieves $e(R_1, Q)$ and P_2 achieves $e(R_2, Q)$, by definition of $e(R, Q)$, we thus obtain that $e(R, Q)$ is convex:

$$\lambda e(R_1, Q) + (1-\lambda)e(R_2, Q) \geq e(\lambda R_1 + (1-\lambda)R_2, Q).$$

When $R \leq H(Q)$, letting $P = Q$, yields that $e(R, Q) = 0$. Since $e(R, Q)$ is increasing and convex and is zero when $R \leq H(Q)$, we then conclude that $e(R, Q)$ is a strictly increasing and continuous function of R when $H(Q) \leq R \leq \log|\mathcal{S}|$. □

Corollary 2.1 *For a DMS with generic distribution Q and alphabet $\mathcal{S}$, the source reliability function satisfies*

$$e(R, Q) = \begin{cases} \min_{P:H(P)=R} D(P \parallel Q), & \textit{if } H(Q) \leq R \leq \log|\mathcal{S}|, \\ 0, & \textit{if } 0 \leq R \leq H(Q). \end{cases} \tag{2.5}$$

Proof: Here we only need to explain the part when $H(Q) \leq R \leq \log |S|$. If $e(R_1, Q)$ is achieved for a P such that $H(P) = R_2 > R_1$, then $e(R_1, Q) = e(R_2, Q)$, which contradicts Theorem 2.2. $\square$

We also give Jelinek's source reliability function.

Corollary 2.2 *For a DMS with generic distribution Q and alphabet $\mathcal{S}$, the source reliability function satisfies*

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

where

$$E_s(\rho) \triangleq (1 + \rho) \log \sum_{s \in \mathcal{S}} Q(s)^{\frac{1}{1+\rho}}$$

is called Gallager's source function.

Remark: It has been shown that the parametric form (2.6) is an equivalent expression of (2.5), see Blahut [8].

2.1.2 Source Coding Exponent with Fidelity Criterion

Definition 2.3 *Let $\mathcal{S}$ be a finite set and $d(\cdot, \cdot)$ be a distortion measure, i.e.,a nonnegative valued function defined on $\mathcal{S} \times \mathcal{S}$ and extended to $\mathcal{S}^n \times \mathcal{S}^n$ by setting*

$$d(s^n, \widetilde{s}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(s_i, \widetilde{s}_i),$$

where $s^n \triangleq (s_1 \cdots s_n) \in \mathcal{S}^n$ and $\widetilde{s}^n \triangleq (\widetilde{s}_1 \cdots \widetilde{s}_n) \in \mathcal{S}^n$.

Definition 2.4 *For a given distortion measure (nonnegative valued function) $d(\cdot, \cdot)$ on $\mathcal{S} \times \mathcal{S}$, a n-length block code (f, φ) for sources with generic distribution Q and alphabet $\mathcal{S}$ is a pair of mappings:*

$$f : \mathcal{S}^n \longrightarrow \{1, 2, ..., M\} \quad \text{and} \quad \varphi : \{1, 2, ..., M\} \longrightarrow \mathcal{S}^n.$$

The probability that the n-length message s^n of a DMS with generic distribution Q is not reproduced within distortion Δ is

$$P_\Delta(n, M) \triangleq \sum_{s^n : d(s^n, \varphi(f(s^n))) > \Delta} Q(s^n), \tag{2.7}$$

where $s^n \triangleq (s_1 \cdots s_n) \in \mathcal{S}^n$. The rate of the code is defined as

$$R(n, M) \triangleq \frac{1}{n} \log M \quad \textit{bits/source symbol.}$$

Definition 2.5 *The rate-distortion function of a DMS with generic distribution Q and alphabet $\mathcal{S}$ under distortion measure $d(s, \widehat{s})$ is defined as*

$$R(Q, \Delta) \triangleq \min_{P(\widehat{s}|s) : \sum_{s,\widehat{s}} Q(s) P(\widehat{s}|s) d(s,\widehat{s}) \leq \Delta} I(S; \widehat{S}), \tag{2.8}$$

where $I(\cdot;\cdot)$ is the mutual information, and Δ is the distortion threshold.

Definition 2.6 (*Source Reliability Function with a Fidelity Criterion*) *For any $R > 0$, the source reliability function $F(Q, R, \Delta)$ of the source with distribution Q under a distortion threshold Δ is defined as the largest number e^* for which there exists a sequence of n-length block codes (n, M) with*

$$e^* \leq \liminf_{n \to \infty} -\frac{1}{n} \log P_\Delta(n, M) \tag{2.9}$$

and

$$R \geq \limsup_{n \to \infty} \frac{1}{n} \log M. \tag{2.10}$$

Theorem 2.3 [*9*] *Given a DMS with generic distribution Q with distortion measure $d(\cdot, \cdot)$, the source reliability function under a distortion threshold Δ satisfies*

$$F(R, Q, \Delta) = \inf_{P : R(P, \Delta) \geq R} D(P \parallel Q). \tag{2.11}$$

Corollary 2.3 *$F(R, Q, \Delta)$ is an increasing function in R.*

Remark: Unlike $e(R, Q)$, $F(R, Q, \Delta)$ is not necessarily convex or even continuous in R.

2.2 Channel Reliability Function

By Shannon's Channel Coding Theorem, block codes with arbitrarily small probability of block decoding error exist at any code rate smaller than the channel capacity C. However, the result does not carefully describe the relation between the rate of convergence or decay for the error probability and the code rate (specifically for rates $R \leq C$).

Like the source coding reliability function $e(R, Q)$, a channel coding error exponent (resulting in the decay of the error probability) is needed in order to know how large the code length should be in order to achieve a certain level of error probability.

For a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, let $W(y^n|x^n) \triangleq P_{Y^n|X^n}(y^n|x^n)$ denote the conditional probability that sequence $y^n \triangleq (y_1, ..., y_n)$ is received given sequence $x^n \triangleq (x_1, ...x_n)$ was transmitted.

Definition 2.7 *An (n, M) block code for a discrete channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and block transition probability $W(y^n|x^n) \triangleq P_{Y^n|X^n}(y^n|x^n), \{n = 1, 2, ...\}$ is a pair of mappings:*

$$f : \{1, 2, ..., M\} \longrightarrow \mathcal{X}^n \quad \text{and} \quad \varphi : \mathcal{Y}^n \longrightarrow \{1, 2, ..., M\}.$$

The average error probability of this code is

$$P_e(n, M) \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{y^n : \varphi(y^n) \neq m} P_{Y^n|X^n}(y^n|f(m)) \tag{2.12}$$

and the code rate is defined as

$$R(n, M) \triangleq \frac{1}{n} \log M \quad \text{bits/channel use.}$$

Definition 2.8 (*Channel reliability function*) *For any $R > 0$, the channel reliability function $E(R, W)$ of the channel W is defined as the largest number E for which there* *exists a sequence of (n, M) block codes with*

$$E \leq \liminf_{n\to\infty} -\frac{1}{n} \log P_e(n, M) \tag{2.13}$$

and

$$R \leq \liminf_{n\to\infty} \frac{1}{n} \log M. \tag{2.14}$$

The properties of $E(R, W)$ are as follows:

• From the above definition, we directly obtain that the asymptotic behavior in n of $P_e(n, M)$ is upper bounded by

$$P_e(n, M) \leq 2^{-nE(R,W)+o(n)}$$

where $o(n) \to 0$ as $n \to 0$.

• for $R > C$, $E(R, W) = 0$, where C is the channel capacity. For a DMC, $C = \max_{p(x)} I(X; Y)$, where the maximum is taken over all possible channel input distribution $p(x)$.

• Like channel capacity C, $E(R, W)$ is a quantity that depends on the channel characteristics.

But unlike the source reliability function, the reliability function of a DMC is not known for all R. The most familiar bounds to $E(R, W)$ are the random-coding and expurgated lower bounds, and the sphere-packing upper bound due to Fano (1961) and Shannon-Gallager-Berlerkamp (1967), respectively (see [4, 6]).

2.2.1 Random-Coding Exponent

Definition 2.9 (*Random-Coding Exponent $E_r(R, W)$*) *The random-coding exponent for a DMC with generic distribution $W \triangleq P_{Y|X}$ is defined by*

$$E_r(R, W) \triangleq \max_{0\leq\rho\leq 1} \left[-\rho R + E_o(\rho)\right], \tag{2.15}$$

where

$$E_o(\rho) \triangleq \max_{P_X} \left[-\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y|x) \right)^{1+\rho} \right] \tag{2.16}$$

is called Gallager's channel function.

Remark: If we denote $E_o(\rho, P_X) \triangleq -\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y|x) \right)^{1+\rho}$, then

$$E_r(R, W) = \max_{0 \leq \rho \leq 1} \left[-\rho R + \max_{P_X} E_o(\rho, P_X) \right] = \max_{P_X} \max_{0 \leq \rho \leq 1} \left[-\rho R + E_o(\rho, P_X) \right],$$

since the function $[-\rho R + E_o(\rho, P_X)]$ is concave over ρ and P_X, which are independent of each other, see [4, p. 143-144].

Theorem 2.4 *[4] (Random-Coding Lower Bound) For a DMC with generic distribution probability $W \triangleq P_{Y|X}$,*

$$E(R, W) \geq E_r(R, W) \tag{2.17}$$

for $R > 0$.

Lemma 2.1 *[4] (Properties of $E_r(R, W)$)*

1. There exists a critical rate R_{cr} such that for $0 \leq R \leq R_{cr}$, $E_r(R, W)$ is a line with slope equal to -1, and with $E_r(0, W) = E_o(1)$.

2. When $R_{cr} \leq R \leq C$, $E_r(R, W)$ is strictly convex and decreasing in R, where C is the channel capacity such that $E_r(C, W) = 0$, and $E_r(C - \delta, W) > 0$, for all $0 \leq \delta < C$.

3. The critical rate R_{cr} and the channel capacity C can be related to $E_o(\rho)$ by

$$R_{cr} = \left. \frac{\partial E_o(\rho)}{\partial \rho} \right|_{\rho=1} \quad \textit{and} \quad C = \left. \frac{\partial E_o(\rho)}{\partial \rho} \right|_{\rho=0}.$$

2.2.2 Expurgated Exponent

The proof of the random-coding exponent is based on the argument of random coding, which selects codewords independently according to some input distribution. Since the codewords selection is unbiased, the "good" codewords (i.e, with small $P_{e|m}(C_{M_n})$, where $P_{e|m}(C_{M_n})$ is the conditional error probability given that message m is sent for a code with codebook C_{M_n} and M_n codewords.) and "bad" codewords (with large $P_{e|m}(C_{M_n})$) contribute equally to the overall average error probability. Therefore, if, to some extent, we can "expurgate" the contribution of the "bad" codewords, a better bound on the channel reliability function may be found.

Instead of randomly selecting M_n codewords, the expurgated approach first draws $2M_n$ codewords to form a codebook $\mathcal{C}_{2M_n}$ and sorts these codewords in ascending order in terms of $P_{e|m}(\mathcal{C}_{2M_n})$ to form a new code book $\mathcal{C}_{M_n}$. It can be expected that for $1 \le m \le M_n$ and $M_n \le m' \le 2M_n$,

$$P_{e|m}(\mathcal{C}_{M_n}) \le P_{e|m}(\mathcal{C}_{2M_n}) \le P_{e|m'}(\mathcal{C}_{2M_n})$$

provided that the Maximum-Likelihood (ML) decoder is always employed at the receiver side. Hence, a better codebook is obtained.

Definition 2.10 (*Expurgated Exponent $E_{ex}(R,W)$*) *The expurgated exponent for a DMC with generic distribution $W \triangleq P_{Y|X}$ is defined by*

$$E_{ex}(R,W) \triangleq \max_{\rho \ge 1}[-\rho R + E_x(\rho)], \tag{2.18}$$

where

$$E_x(\rho) \triangleq \max_{P_X}\left[-\rho \log \sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} P_X(x)P_X(x')\left(\sum_{y\in\mathcal{Y}}\sqrt{P_{Y|X}(y|x)P_{Y|X}(y\mid x')}\right)^{1/\rho}\right]. \tag{2.19}$$

Remark: The maximization of $E_{ex}(R, W)$ is quite similar to that of the random-coding exponent. The order of the two maximizations in $E_{ex}(R, W)$ can also be exchanged, see [4, p. 155].

Theorem 2.5 [4] (*Expurgated Lower Bound*) *For a DMC with generic transtion probability* $W \triangleq P_{Y|X}$,

$$E(R, W) \geq E_{ex}(R, W) \tag{2.20}$$

for $R > 0$.

Lemma 2.2 [4] (*Properties of* $E_{ex}(R, W)$)
1. There exists an R'_{cr} $(R'_{cr} \leq R_{cr})$ *given by*

$$R'_{cr} = \left.\frac{\partial E_x(\rho)}{\partial \rho}\right|_{\rho=1},$$

such that for $R'_{cr} \leq R \leq E_x(1)$, $E_{ex}(R, W)$ *is a line with slope equal to* -1, *where* $E_x(1)$ *is the smallest rate where* $E_{ex}(R, W) = 0$, *i.e,* $E_{ex}\left(E_x(1), W\right) = 0$, *and* $E_{ex}\left(E_x(1) - \delta, W\right) > 0$, *for all* $0 \leq \delta < E_x(1)$. *Furthermore,* $E_x(1) = E_o(1)$.
2. $E_{ex}(R, W))$ *is strictly convex and decreasing in* $R \in (R_{x,\infty}, R'_{cr})$, *where* $R_{x,\infty}$ *is the biggest rate where* $E_{ex}(R, W)$ *is infinite, given by*

$$R_{x,\infty} = \lim_{\rho \to \infty} \frac{\partial E_x(\rho)}{\partial \rho}.$$

3. If $R_{x,\infty} = 0$, $\lim_{R \to 0} E(R, W) = \lim_{R \to 0} E_{ex}(R, W)$, *i.e., the expurgated bound is tight as R goes to 0.*

Remark: By the first property, we observe that the expurgated bound $E_{ex}(R, W)$ cannot improve the random-coding lower bound $E_r(R, W)$ for $R \geq R'_{cr}$.

2.2.3 Sphere-Packing Exponent

Definition 2.11 *(Sphere Packing Exponent $E_{sp}(R,W)$) The sphere-packing exponent for a DMC with generic transition probability $W \triangleq P_{Y|X}$ is defined by*

$$E_{sp}(R,W) \triangleq \max_{\rho \geq 0} \left[-\rho R + E_o(\rho)\right], \tag{2.21}$$

where $E_o(\rho)$ is defined in (2.16).

It should be noted that the sphere-packing exponent is defined in almost the same way as the random-coding exponent, the only difference being in the range of ρ over which the maximization is performed.

Theorem 2.6 [4] *(Sphere-Packing Upper Bound) For a DMC with generic transition probability $W \triangleq P_{Y|X}$,*

$$E(R,W) \leq E_{sp}(R,W) \tag{2.22}$$

for $R > 0$.

Lemma 2.3 [4] *(Properties of $E_{sp}(R,W)$)*

1. $E_{sp}(R,W)$ is infinite for all rates less than a constant R_∞, if $R_\infty > 0$, where

$$R_\infty = \lim_{\rho \to \infty} \frac{E_o(\rho)}{\rho}.$$

2. $E_{sp}(R,W))$ is strictly convex and decreasing in $R \in (R_\infty, C)$.

3. $E(R,W) = E_{sp}(R,W) = E_r(R,W)$ for $R_{cr} \leq R \leq C$, i.e, the channel reliability function is exactly known at $R \geq R_{cr}$, where R_{cr} is the critical rate of the channel and C is channel capacity.

Chapter 3

Reliability Function for Joint Source-Channel Coding

The results of this chapter are obtained from Csiszár's work in [1, 2].

Definition 3.1 *A joint source-channel code for a DMS with generic distribution Q and finite alphabet $\mathcal{S}$ and a DMC with transition probability $W \triangleq P_{Y|X}$ input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ with block length n is a pair of mappings:*

$$f : \mathcal{S}^n \longrightarrow \mathcal{X}^n \quad \textit{and} \quad \varphi : \mathcal{Y}^n \longrightarrow \mathcal{S}^n.$$

The average error probability of the code is

$$P_e^{(n)}(Q, W) \triangleq \sum_{(s^n, y^n): \varphi(y^n) \neq s^n} Q(s^n) P(y^n | f(s^n)). \tag{3.1}$$

Definition 3.2 (*Joint Source-Channel Reliability Function*) *The joint source-channel reliability function $E_J(Q, W)$ for source Q and channel W is defined as the largest number E^* for which there exists a sequence of n-length joint source-channel codes (f, φ) with*

$$E^* \leq \liminf_{n \to \infty} -\frac{1}{n} \log P_e^{(n)}(Q, W). \tag{3.2}$$

We know from the joint source-channel coding theorem that $E_J(Q, W)$ can be positive if and only if $H(Q) < C$. Note that $E_J(Q, W)$ is a quantity that only depends on the source and channel distributions.

3.1 Csiszár's Joint Source-Channel Coding Theorems

Theorem 3.1 [1] *Given a DMS with generic distribution Q and finite alphabet $\mathcal{S}$ (denoted by $\{Q : \mathcal{S}\}$) and a DMC with transition probability $W \triangleq P_{Y|X}$ input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ ((denoted by $\{W : \mathcal{X} \to \mathcal{Y}\}$)), the joint source-channel reliability function satisfies*

$$E_J(Q, W) \geq \min_{H(Q) \leq R \leq \min(\log |\mathcal{S}|, C)} [e(R, Q) + E_r(R, W)] \tag{3.3}$$

and

$$\begin{aligned} E_J(Q, W) &\leq \min_{H(Q) \leq R \leq \min(\log |\mathcal{S}|, C)} [e(R, Q) + E(R, W)] \\ &\leq \min_{H(Q) \leq R \leq \min(\log |\mathcal{S}|, C)} [e(R, Q) + E_{sp}(R, W)], \end{aligned} \tag{3.4}$$

where $e(R, Q)$ is the DMS reliability function and $E(R, W)$ is the DMC reliability function. By Lemma 2.3 (3), if the minimum in (3.3) or (3.4) is attained at a rate R_m such that $R_m \geq R_{cr}$, where R_{cr} is the critical rate of the channel, then the lower bound in (3.3) and the upper bound in (3.4) coincide with each other, and we can determine the joint source-channel coding reliability function exactly:

$$E_J(Q, W) = e(R_m, Q) + E_r(R_m, W) = e(R_m, Q) + E(R_m, W). \tag{3.5}$$

Of course, the two bounds are nontrivial if and only if $H(Q) < C$.

The proof of (3.4) involves a method of types approach, and the proof of (3.3) is based on a method of types and a random-coding argument, see [1, 3, 12].

Remark: We will call the lower bound to $E_J(Q,W)$ in (3.3) the "random-coding lower bound" since it is based on the random-coding lower bound to the channel reliability function; Similarly, we will call the upper bound in (3.4) the "sphere-packing upper bound".

Definition 3.3 *(Bhattacharya Distance d_B) The Bhattacharya distance for a DMC $P_{Y|X}$ between elements x and $\widetilde{x}$ is defined by*

$$d_B(x,\widetilde{x}) \triangleq -\log \sum_{y\in\mathcal{Y}} \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\widetilde{x})}. \tag{3.6}$$

With the definition of Bhattacharya distance, we can re-write

$$E_{ex}(R,W) = \max_P E_{ex}(R,P,W), \tag{3.7}$$

where

$$E_{ex}(R,P,W) \triangleq \max_{\rho\geq 1} \left\{ -\rho R - \rho \log \sum_{x\in\mathcal{X}} \sum_{\widetilde{x}\in\mathcal{X}} P_X(x)P_X(\widetilde{x}) \exp\left\{ -\frac{1}{\rho} d_B(x,\widetilde{x}) \right\} \right\}.$$

Csiszár and Körner [3] showed that $E_{ex}(R,P,W)$ is equivalent to

$$E_{ex}(R,P,W) = \min_{I(X,\widetilde{X})\leq R, P_X = P_{\widetilde{X}} = P} \left[\mathbf{E} d_B(X,\widetilde{X}) + I(X,\widetilde{X}) - R \right], \tag{3.8}$$

where

$$\mathbf{E} d_B(X,\widetilde{X}) = \sum_{x,\widetilde{x}} P_{X\widetilde{X}}(x,\widetilde{x}) d_B(x,\widetilde{x}).$$

Definition 3.4 *(Zero-error Capacity) The zero-error capacity of a channel is defined as the largest rate at which data can be transmitted over the channel with zero error probability (as opposed to arbitrarily small error probability).*

From [1] we know that for a DMS $\{Q:\mathcal{S}\}$ and a DMC $\{W:\mathcal{X}\to\mathcal{Y}\}$ with positive zero-error capacity, we have

$$E_J(Q,W) = \min_R [e(R,Q) + E(R,W)].$$

In the rest of the section, we suppose that $\{W : X \to Y\}$ is a given DMC with zero-error capacity equal to 0. Then $d_B(x, \widetilde{x})$ is finite for every x and $\widetilde{x}$ in $\mathcal{X}$, and so is $E_{ex}(R, W)$ for every $R > 0$, i.e., $R_{x,\infty} = 0$ ($R_{x,\infty}$ is described in Lemma 2.2).

Theorem 3.2 *[2] Given a DMS $\{Q : \mathcal{S}\}$ and a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, for an arbitrary input distribution P on $\mathcal{X}$, the joint source-channel reliability function satisfies*

$$E_J(Q, W) \geq \min_{0 \leq R \leq \min(E_x(1), \log|\mathcal{S}|)} [e(R, Q) + E_{ex}(R, P, W)]. \tag{3.9}$$

The above inequality also holds with $E_{ex}(R, P, W)$ replaced by

$$\widetilde{E}_{ex}(R, W) \triangleq \max_{P} \min_{I(X, \widetilde{X}) \leq R, P_X = P} \left[\mathbf{E} d_B(X, \widetilde{X}) + I(X, \widetilde{X}) - R \right], \tag{3.10}$$

where the minimum is taken for a fixed X with $P_X = P$, with respect to all random variables (RV's) $\widetilde{X}$ with values in $\mathcal{X}$.

The proof of Theorem 3.2 is similar to that of (3.3), and differs only in the choice of decoder, see [2].

From (3.7), (3.8) and (3.10), we note that $\widetilde{E}_{ex}(R, W) \leq E_{ex}(R, W)$. Thus we have the following result.

Corollary 3.1 *[2] If $E_{ex}(R, W) = \max_P E_{ex}(R, P, W)$ is attained for a P not depending on R then*

$$E_J(Q, W) \geq \min_{H(Q) \leq R \leq \min(E_x(1), \log|\mathcal{S}|)} [e(R, Q) + E_{ex}(R, W)]. \tag{3.11}$$

Remark: We will call the lower bound to $E_J(Q, W)$ in (3.11) the “expurgated lower bound” since it is based on the expurgated lower bound to the channel reliability function.

3.2 General Fidelity Criteria

Definition 3.5 *Suppose the message $s^n \in \mathcal{S}^n$ is to be transmitted over a DMC with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and transition probability $W \triangleq P_{Y|X}$ with a threshold Δ of tolerated distortion. Assume a joint source-channel n-length block code with*

$$f : \mathcal{S}^n \longrightarrow \mathcal{X}^n \quad \text{and} \quad \varphi : \mathcal{Y}^n \longrightarrow \mathcal{S}^n.$$

The average probability of the code of exceeding the threshold Δ is

$$P_\Delta^{(n)}(Q, W) \triangleq \sum_{(s^n, y^n): d(\varphi(y^n), s^n) > \Delta} Q(s^n) P(y^n | f(s^n)), \tag{3.12}$$

where $d(\cdot, \cdot)$ is the distortion measure described in Definition 2.3.

Note that when $d(s^n, \widetilde{s^n}) = 0$ if and only if $s^n = \widetilde{s^n}$ and $\Delta = 0$, then we write $P_e^{(n)}(Q, W)$ instead of $P_\Delta^{(n)}(Q, W)$, and the former one is the probability of error.

Definition 3.6 (*Joint Source-Channel Reliability Function with a Distortion Threshold*) *The joint source-channel reliability function $E_J^\Delta(Q, W)$ of source $\{Q : \mathcal{S}\}$ over channel $\{W : \mathcal{X} \to \mathcal{Y}\}$ under distortion threshold Δ is defined as the largest number E^Δ for which there exists a sequence of n-length joint source-channel codes (f, φ) with*

$$E^\Delta \leq \liminf_{n \to \infty} -\frac{1}{n} \log P_\Delta^{(n)}(Q, W). \tag{3.13}$$

We know from the joint source-channel coding theorem that $E_J^\Delta(Q, W)$ can be positive if and only if $R_{Q,\Delta} < C$.

Theorem 3.3 [2] *Given a DMS $\{Q : \mathcal{S}\}$ and a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, the joint source-channel reliability function under distortion threshold Δ satisfies*

$$E_J^\Delta(Q, W) \geq \inf_{R(Q,\Delta) \leq R \leq \min(\log |\mathcal{S}|, C)} [F(R, Q, \Delta) + E_r(R, W)] \tag{3.14}$$

and

$$E_J^{\Delta}(Q,W) \le \inf_{R(Q,\Delta)\le R\le \min(\log|\mathcal{S}|,C)} [F(R,Q,\Delta) + E_{sp}(R,W)]. \tag{3.15}$$

If the infimum in (3.14) or (3.15) is attained for $R_m \ge R_{cr}$, *where* R_{cr} *is the critical rate of the channel, then the lower bound and the upper bound coincide with each other, and we can determine the joint source-channel coding reliability function exactly:*

$$E_J^{\Delta}(Q,W) = F(R_m,Q,\Delta) + E_r(R_m,W) = F(R_m,Q,\Delta) + E_{sp}(R_m,W). \tag{3.16}$$

Of course, the two bounds are nontrivial if and only if $R(Q, Delta) < C$.

Theorem 3.4 [2] *Given a DMS* $\{Q : \mathcal{S}\}$ *and a DMC* $\{W : \mathcal{X} \to \mathcal{Y}\}$, *for an arbitrary input distribution* P *on* $\mathcal{X}$, *the joint source-channel reliability function under distortion threshold* Δ *satisfies*

$$E_J^{\Delta}(Q,W) \ge \inf_{R(Q,\Delta)\le R\le \min(E_x(1),\log|\mathcal{S}|)} [F(R,Q,\Delta) + E_{ex}(R,P,W)] \tag{3.17}$$

The above inequality also holds with $E_{ex}(R,P,W)$ *replaced by* $\widetilde{E}_{ex}(R,W)$ *defined in (3.10).*

Corollary 3.2 [2] *If* $E_{ex}(R,W) = \max_P E_{ex}(R,P,W)$ *is attained for a* P *not depending on* R *then*

$$E_J^{\Delta}(Q,W) \ge \inf_{R(Q,\Delta)\le R\le \min(E_x(1),\log|\mathcal{S}|)} [F(R,Q,\Delta) + E_{ex}(R,W)]. \tag{3.18}$$

Chapter 4

Computation of the Joint Source-Channel Coding Exponent

In this chapter, we will employ Csiszár's results to analytically compute the source-channel coding exponent. Without loss of generality, we express the source distribution $Q(s)$, $s \in \mathcal{S}$, as $Q = \big(q_1, q_2, \cdots, q_{|\mathcal{S}|}\big)$.

4.1 The Joint Source-Channel Random-Coding Lower Bound

Definition 4.1 *Consider the twice differentiable function of n real variables*

$$f(x_1, x_2, ..., x_n) = f(\boldsymbol{x}).$$

The point with co-ordinates $(x_1, x_2, ..., x_n)$ in the n dimensional Euclidean space is denoted by the column vector $\boldsymbol{x}$. $f(\cdot)$ has a local minimum at $\boldsymbol{x}_0$ if there exists a neighbourhood of $\boldsymbol{x}$ such that $f(\boldsymbol{x})$ is at least as large as $f(\boldsymbol{x}_0)$ for all points in the neighbourhood, i.e., there exists a positive δ such that for $|\boldsymbol{x} - \boldsymbol{x}_0| < \delta$, $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$. Furthermore, $f(\cdot)$ has a global minimum at $\boldsymbol{x}^*$ if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x}$.

Lemma 4.1 *[5] The necessary condition for $f(\cdot)$ having a local minimum at $\boldsymbol{x}_0$ is*

$$\nabla f(\boldsymbol{x}_0) = \boldsymbol{0}, \tag{4.1}$$

where $\nabla f(\cdot)$ is the gradient of the function. The point $\boldsymbol{x}_0$ satisfying (4.1) is called a stationary point.

Note that (4.1) is not a necessary condition for a global minimum. For example, for the one variable function $f(x)$ for which the variable x is defined on $[a, b]$, the global minimum of $f(x)$ is $\min\{f(a), f(x_1), f(x_2), ..., f(x_n), f(b)\}$, where $x_1, x_2, ..., x_n$ are all the stationary points or local minimum points within $[a, b]$. Specially, if $f(x)$ is convex, then it has at most one stationary point in $[a, b]$, and the global minimum is

$$\min_{a \leq x \leq b} f(x) = \begin{cases} f(x_0), & \text{if there exists a stationary point } x_0 \text{ in [a,b]}, \\ \min\{f(a), f(b)\}, & \text{if there is no stationary point in [a,b]}. \end{cases}$$

4.1.1 A General Derivation

We begin by recalling some useful properties of the random-coding lower bound for the channel reliability function of a DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$. Since

$$E_r(R, W) = \max_{0 \leq \rho \leq 1} [-\rho R + E_o(\rho)],$$

by setting the derivative of $[-\rho R + E_o(\rho)]$[1] with respect to ρ to zero, we obtain parametric expressions for R and $E_r(R, W)$ in terms of ρ^* (ρ^* maximizes $[-\rho R + E_o(\rho)]$ in [0,1], see [4, p. 142]). When

$$R_{cr} = \left.\frac{\partial E_o(\rho^*)}{\partial \rho^*}\right|_{\rho^*=1} \leq R \leq C = \left.\frac{\partial E_o(\rho^*)}{\partial \rho^*}\right|_{\rho^*=0},$$

[1]We remark that there is a slight difference from Gallager's textbook [4, p. 142], where he takes the derivative of $[-\rho R + E_o(\rho, P_X)]$ instead of $[-\rho R + E_o(\rho)]$, where $E_o(\rho, P_X)$ is defined in Section 2.2.1. We note that this change does not effect our results since the following manipulations generally hold for any given P_X, and in Section 2.2.1 we have mentioned that the order of the two maximizations (over ρ and over P_X) in $E_r(R, W)$ can be exchanged. In fact, $E_o(\rho)$ can also be seen as $E_o(\rho, \mathbf{P})$, where $\mathbf{P}$ maximizes $E_o(\rho, P_X)$.

$R(\rho^*) = \frac{\partial E_o(\rho^*)}{\partial \rho^*}$ is a decreasing function of ρ^* $(0 \leq \rho^* \leq 1)$ and

$$E_r(R, W) = E_o(\rho^*) - \rho^* \frac{\partial E_o(\rho^*)}{\partial \rho^*}. \tag{4.2}$$

In fact, $-\rho^*$ is the slope of $E_r(R, W)$ since

$$\frac{\partial E_r(R, W)}{\partial R} = \frac{\frac{\partial E_r(R,W)}{\partial \rho^*}}{\frac{\partial R}{\partial \rho^*}} = \frac{-\rho^* \frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*}}{\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*}} = -\rho^*.$$

Remark: We notice that $E_o(\rho^*)$ and $R(\rho^*) = \frac{\partial E_o(\rho^*)}{\partial \rho^*}$ are differentiable[2] with respect to ρ^*, and (4.2) implies that $E_r(R(\rho^*), W)$ is also differentiable with respect to ρ^*. It immediately follows that $E_r(R, W)$ is differentiable with respect to R.

A special case occurs if $\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*} = 0$, then $\frac{\partial E_o(\rho^*)}{\partial \rho^*}$ is constant and $R_{cr} = C$. When $0 \leq R \leq R_{cr}$, the slope is -1 and hence

$$E_r(R, W) = E_o(1) - R, \tag{4.3}$$

which implies that for the special case of $\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*} = 0$, $E_o(1) = C$.

The source reliability function for a DMS $\{Q : \mathcal{S}\}$ is given by

$$e(R, Q) = \begin{cases} \min_{P:H(P)=R} D(P \parallel Q), & \text{if } H(Q) \leq R \leq \log |\mathcal{S}|, \\ 0, & \text{if } 0 \leq R \leq H(Q). \end{cases} \tag{4.4}$$

Now our goal is to compute

$$\min_{H(Q) \leq R \leq \min(C, \log |\mathcal{S}|)} [E_r(R, W) + e(R, Q)]. \tag{4.5}$$

Since $E_J(Q, W) = 0$ when $H(Q) > C$, we only consider the following two cases:

$$R_{cr} \leq H(Q) \leq \min(C, \log |\mathcal{S}|) \quad \text{and} \quad H(Q) \leq R_{cr}.$$

Note that if $\log |\mathcal{S}| \leq R_{cr}$, then $H(Q) \leq R_{cr}$.

Case 1: When $R_{cr} \leq H(Q) \leq \min\{C, \log |\mathcal{S}|\}$, we need to compute

$$\begin{aligned} &\min_{H(Q) \leq R \leq \min(C, \log |\mathcal{S}|)} [E_r(R, W) + e(R, Q)] \\ = &\min_{H(Q) \leq R \leq \min(C, \log |\mathcal{S}|)} \left[E_r(R, W) + \min_{R=H(P)} D(P \parallel Q) \right]. \end{aligned} \tag{4.6}$$

[2]All the elementary functions are differentiable.

Setting $P = (p_1, p_2, ..., p_{|\mathcal{S}|})$, we may write (4.6) as

$$\min G(R,P) = \min \left[E_r(R,W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{q_i} \right] \tag{4.7}$$

subject to

$$\begin{array}{ll} (1) & H(P) - R = 0, \\ (2) & \sum_{i=1}^{|\mathcal{S}|} p_i = 1, \quad p_i > 0, \\ (3) & H(Q) \le R \le \min(C, \log|\mathcal{S}|). \end{array}$$

Note that here we assume that $R_{cr} < C$, i.e., $\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*} < 0$. If $\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*} = 0$, $E_r(R,W)$ is a line with a slope equal to -1 for all the $R \in [0, C]$, and this case can be included in Case 2. Since

$$\begin{aligned} \frac{\partial^2 G(R,P)}{\partial^2 R} &= \frac{\partial^2 E_r(R,W)}{\partial^2 R} = -\frac{1}{\frac{\partial^2 E_o(\rho^*)}{\partial^2 \rho^*}} > 0 \\ \text{and} \quad \frac{\partial^2 G(R,P)}{\partial^2 p_i} &= \frac{\log e}{p_i} > 0, \quad \forall \quad i = 1, ..., |\mathcal{S}|, \end{aligned}$$

the Hessian matrix [5] of $G(R,P)$,

$$\nabla^2 G(R,P) = \begin{bmatrix} \frac{\partial^2 E_r(R,W)}{\partial^2 R} & 0 & \cdots & 0 \\ 0 & \frac{\log e}{p_1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \frac{\log e}{p_{|\mathcal{S}|}} \end{bmatrix},$$

is positive definite. Thus $G(R,P)$ is strictly convex in (R,P) and it admits a global minimum. Actually the existence of the global minimum can also be seen by noticing that $E_r(R,W) + e(R,W)$ is strictly convex in R, which is obvious because the sum of two strictly convex function is still convex. Observing the constraint (3) in (4.7), we see that if there exists a stationary point in the range provided by constraint (3), then the stationary point is the global minimum of the sum over R within the range $[H(Q), \min(C, \log|\mathcal{S}|)]$, otherwise the minimum must be attained at the boundary points, $R = H(Q)$ or $R = \min(C, \log|\mathcal{S}|)$. The Lagrangian function [5] $F(R, P, \lambda, \mu)$ is then given by

$$\begin{aligned} F(R,P,\lambda,\mu) &= E_r(R,W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{q_i} \\ &\quad + \lambda \left(R + \sum_{i=1}^{|\mathcal{S}|} p_i \log p_i \right) + \mu \left(\sum_{i=1}^{|\mathcal{S}|} p_i - 1 \right). \end{aligned} \tag{4.8}$$

Setting the derivative with respect to R equal to zero yields

$$-\lambda = \frac{\partial E_r(R,W)}{\partial R}, \tag{4.9}$$

which implies that $-\lambda$ is the slope of $E_r(R,W)$ with $\lambda \geq 0$. Therefore $\lambda = -\rho^*$ and

$$R = \frac{\partial E_o(\lambda)}{\partial \lambda}. \tag{4.10}$$

Setting the derivative with respect to p_i equal to zero and solving for μ under constraint (2) yields

$$p_i = \frac{q_i^{\frac{1}{1+\lambda}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda}}}, \quad i = 1, 2, ..., |\mathcal{S}|. \tag{4.11}$$

According to constraint (1), the stationary point $(R(\lambda_1), \mathbf{p}(\lambda_1))$ for $G(R,P)$ should satisfy

$$H(\mathbf{p}(\lambda_1)) - R(\lambda_1) = 0, \tag{4.12}$$

where $\mathbf{p}(\lambda_1)$ stands for the distribution $\big(p_1(\lambda_1), p_2(\lambda_1), ..., p_{|\mathcal{S}|}(\lambda_1)\big)$ and

$$H(\mathbf{p}(\lambda_1)) \triangleq H(P(\lambda_1)) = H\left(p_1(\lambda_1), p_2(\lambda_1), ..., p_{|\mathcal{S}|}(\lambda_1)\right) = -\sum_{i=1}^{|\mathcal{S}|} p_i(\lambda_1) \log p_i(\lambda_1).$$

For the sake of convenience, we will sometimes allow a slight abuse of notation by denoting $H(P)$ or $H(\mathbf{p})$ by $H(p_i)$.

Lemma 4.2 *$f(\beta) \triangleq H\left(\frac{x_i^{\frac{1}{1+\beta}}}{\sum_{i=i}^{N} x_i^{\frac{1}{1+\beta}}}\right)$ is increasing in β ($\beta > -1$), where $x_i > 0$, $i = 1, ..., N$, N is a positive integer. Moreover, $f(\beta)$ is strictly increasing in β if there exist β_1, β_2 ($\beta_1 < \beta_2$) such that $f(\beta_1) < f(\beta_2)$.*

Proof: Differentiating with respect to β, we have

$$\begin{aligned}\frac{\partial f(\beta)}{\partial \beta} &= \frac{1}{(1+\beta)^3} \\ &\quad \frac{\sum_{i=1}^{N}[x_i^{\frac{1}{1+\beta}}\log^2(x_i)]\sum_{i=1}^{N}[x_i^{\frac{1}{1+\beta}}] - [\sum_{i=1}^{N} x_i^{\frac{1}{1+\beta}}\log(x_i)]^2}{[\log e \sum_{i=1}^{N} x_i^{\frac{1}{1+\beta}}]^2} \geq 0.\end{aligned}$$

The last inequality follows from the *Cauchy-Schwartz* inequality, with equality if and only if x_i = constant. If $f(\beta_1) < f(\beta_2)$ holds for some $\beta_1 < \beta_2$, it implies that $x_i \neq$ constant, and thus $f(\beta)$ is strictly increasing. $\square$

We next show such a stationary point satisfies constraint (3), and hence it is the global minimum point of (4.7). Setting the auxiliary function

$$\varphi(\lambda) \triangleq H(\mathbf{p}(\lambda)) - R(\lambda), \tag{4.13}$$

we see by Lemma 4.2 that $\varphi(\lambda)$ is increasing in λ since $R(\lambda) = \frac{\partial E_o(\lambda)}{\partial \lambda}$ is decreasing in λ.

- If $\log|\mathcal{S}| \geq C$, by the assumption of Case 1, $H(Q) \leq \min(C, \log|\mathcal{S}|) = C$, we have that
$$\varphi(0) = H(Q) - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=0} = H(Q) - C \leq 0,$$
and since $H(Q) \geq R_{cr}$, we also have that
$$\varphi(1) = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=1} \geq H(Q) - R_{cr} \geq 0.$$
We stress that $\varphi(0) < \varphi(1)$, i.e., $\varphi(0)$ and $\varphi(1)$ cannot be 0 simultaneously. It follows by the strictly increasing property of function $\varphi(\cdot)$ that $0 \leq \lambda_1 \leq 1$. For such λ_1, it is easy to verify that the stationary point satisfies
$$H(Q) - R(\lambda_1) \leq \varphi(\lambda_1) = 0 \Longrightarrow R \geq H(Q)$$
and
$$H(\mathbf{p}(\lambda_1)) - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=0} \leq \varphi(\lambda_1) = 0 \Longrightarrow R = H(\mathbf{p}(\lambda_1)) \leq C.$$

- If $R_{cr} < \log|\mathcal{S}| < C$, by the assumption of Case 1, $H(Q) \leq \min(C, \log|\mathcal{S}|) = \log|\mathcal{S}|$. Since $R(\lambda)$ is decreasing in λ, we claim that there exist a λ_0 ($0 < \lambda_0 < 1$) such that $R(\lambda_0) = \log|\mathcal{S}|$. Furthermore, we note that

$$\varphi(\lambda_0) = H(\mathbf{p}(\lambda_0)) - \log|\mathcal{S}| \leq 0.$$

  Observing that (as above) $\varphi(1) \geq 0$, it follows that $\lambda_0 \leq \lambda_1 \leq 1$. For such λ_1, it is easy to verify that the stationary point satisfies

$$H(Q) - R(\lambda_1) \leq \varphi(\lambda_1) = 0 \Longrightarrow R \geq H(Q)$$

  and

$$H(\mathbf{p}(\lambda_1)) - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_0} \leq \varphi(\lambda_1) = 0 \Longrightarrow R = H(\mathbf{p}(\lambda_1)) \leq \log|\mathcal{S}|.$$

Accordingly, we can conclude that the stationary point is located in $[H(Q), \min(C, \log|\mathcal{S}|)]$, in other words, the global minimum point of (4.7) is $(R(\lambda_1), \mathbf{p}(\lambda_1))$. Substituting it into the objective function $G(R, P)$, yields

$$\begin{aligned}
&\min_{R,P: H(Q) \leq R = H(P) \leq \min(C, \log|\mathcal{S}|)} \left[E_r(R, W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{q_i} \right] \\
&= E_o(\lambda_1) - \lambda_1 R(\lambda_1) + \sum_{i=1}^{|\mathcal{S}|} \frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}} \log \frac{\frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}}}{q_i} \\
&= E_o(\lambda_1) - (1 + \lambda_1) \log \left(\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}} \right), \qquad (4.14)
\end{aligned}$$

where λ_1 is the root of equation $\varphi(\lambda) = 0$. Note that this lower bound to $E_J(Q, W)$ is attained at rate $R = H\left(\frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}} \right)$ or $\left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_1}$.

Case 2: When $H(Q) \leq R_{cr}$, we may write

$$\begin{aligned} & \min_{H(Q)\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)] \\ = \quad & \min\left\{ \min_{R_{cr}\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)], \min_{H(Q)\leq R\leq R_{cr}} [E_r(R,W)+e(R,Q)] \right\} \end{aligned} \tag{4.15}$$

if $\log|\mathcal{S}| > R_{cr}$. Otherwise,

$$\min_{H(Q)\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)] = \min_{H(Q)\leq R\leq \log|\mathcal{S}|} [E_r(R,W)+e(R,Q)]. \tag{4.16}$$

Note that the computation of (4.16) is actually the same with that of the second part of (4.15). For the sake of convenience, we first assume $\log|\mathcal{S}| > R_{cr}$ and compute (4.15). There are two subcases to consider:

$$H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr} \quad and \quad H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \leq R_{cr}.$$

a) Assume $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$.

As in Case 1, we employ Lagrange multipliers to compute

$$\min_{R_{cr}\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)], \tag{4.17}$$

and obtain the stationary point

$$\begin{cases} R = \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_1}, \\ p_i = \frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}}, \quad i=1,...,|\mathcal{S}|, \end{cases} \tag{4.18}$$

where λ_1 is the root of the equation

$$\varphi(\lambda) \triangleq H(\mathbf{p}(\lambda)) - R(\lambda) = 0. \tag{4.19}$$

where $H(\mathbf{p}(\lambda)) = H(p_1(\lambda),...,p_{|\mathcal{S}|}(\lambda))$.

If $C \leq \log|\mathcal{S}|$, we note that

$$\varphi(0) = H(Q) - \frac{\partial E_o(\lambda)}{\partial \lambda}|_{\lambda=0} = H(Q) - C \leq 0,$$

and

$$\varphi(1) = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - \frac{\partial E_o(\lambda)}{\partial \lambda}\,|_{\lambda=1} = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - R_{cr} \geq 0,$$

which implies that $0 \leq \lambda_1 \leq 1$. For such λ_1, it is easy to verify that the stationary point is located in the interval $[H(Q), C]$. If $\log|\mathcal{S}| < C$, we also can show that the stationary point is in $[H(Q), \log|\mathcal{S}|]$. Substituting (4.18) into (4.17), yields

$$\begin{aligned} &\min_{R,P:R_{cr}\leq R=H(P)\leq \min(C,\log|\mathcal{S}|)} \left[E_r(R,W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log\frac{p_i}{q_i}\right] \\ &= E_o(\lambda_1) - (1+\lambda_1)\log\left(\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}\right), \end{aligned} \tag{4.20}$$

where λ_1 is the root of equation $\varphi(\lambda) = 0$. The expression in (4.17) is attained at rate $R = H\left(\frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}}\right)$ or $R = \frac{\partial E_o(\lambda)}{\partial\lambda}\Big|_{\lambda=\lambda_1}$.

Next we note that

$$\begin{aligned} &\min_{H(Q)\leq R\leq R_{cr}} [E_r(R,W) + e(R,Q)] \qquad (4.21)\\ &= \min_{H(Q)\leq R\leq R_{cr}} \left[(E_o(1) - R) + \min_{R=H(P)} D(P \parallel Q)\right] \\ &= \min_{R,P:H(Q)\leq R=H(P)\leq R_{cr}} \left[E_o(1) - R + \sum_{i=1}^{|\mathcal{S}|} p_i\log\frac{p_i}{q_i}\right] \\ &= \min_{P:H(Q)\leq H(P)\leq R_{cr}} \left[E_o(1) - H(P) + \sum_{i=1}^{|\mathcal{S}|} p_i\log\frac{p_i}{q_i}\right] \\ &= E_o(1) - 2\left\{\max_{P:H(Q)\leq H(P)\leq R_{cr}} \sum_{i=1}^{|\mathcal{S}|} p_i\log\frac{\sqrt{q_i}}{p_i}\right\}. \end{aligned}$$

Consider the concave function $g(x) = \log x$. By *Jensen*'s inequality, we obtain

$$\sum_{i=1}^{|\mathcal{S}|} p_i \log\frac{\sqrt{q_i}}{p_i} \leq \log\left(\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}\right)$$

with equality if and only if

$$\frac{\sqrt{q_i}}{p_i} = \text{constant} \quad \text{and} \quad \sum_{i=1}^{|\mathcal{S}|} p_i = 1 \quad \Longrightarrow p_i = \frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}.$$

It looks that the minimum of (4.21) is attained at $R = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$. However, it cannot be achieved because $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$.

Lemma 4.3 *If $f(x)$ is strictly convex and the stationary point of $f(x)$ is attained at $x = x_0$, then when $x \leq x_0$, $f(x)$ is strictly decreasing, and when $x \geq x_0$, $f(x)$ is strictly increasing.*

Consider the strictly convex objective function in (4.21) with respect to R with its stationary point lies outside the interval $[H(Q), R_{cr}]$. It follows by Lemma 4.3 that

$$\min_{H(Q)\leq R\leq R_{cr}} [E_r(R,W) + e(R,Q)] = E_r(R_{cr},W) + e(R_{cr},Q), \tag{4.22}$$

i.e., the global minimum of (4.21) is actually attained at R_{cr} instead of $R = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$.
Comparison of (4.20) with (4.22) reveals that

$$\min_{H(Q)\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W) + e(R,Q)] = E_o(\lambda_1) - (1+\lambda_1)\log\left(\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}\right), \tag{4.23}$$

where λ_1 is determined by $\varphi(\lambda_1) = 0$.
b) Assume $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \leq R_{cr}$.
We employ Lagrange multipliers to compute

$$\min_{R_{cr}\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W) + e(R,Q)], \tag{4.24}$$

and obtain the stationary point

$$\begin{cases} R = \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_1}, \\ p_i = \frac{q_i^{\frac{1}{1+\lambda_1}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\lambda_1}}}, \quad i = 1, ..., |\mathcal{S}|, \end{cases} \tag{4.25}$$

where λ_1 is the root of the equation

$$\varphi(\lambda) \triangleq H(\mathbf{p}(\lambda)) - R(\lambda) = 0. \tag{4.26}$$

Note that

$$\varphi(1) = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=1} = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - R_{cr} \leq 0,$$

and

$$\left.\varphi(\lambda)\right|_{\lambda\to\infty} = \log|\mathcal{S}| - \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda\to\infty} > \log|\mathcal{S}| - R_{cr} \geq 0,$$

which implies that $\lambda_1 \geq 1$. Thus the local minimum $R(\lambda_1)$ is not located in the interval $(R_{cr}, \min(C, \log|\mathcal{S}|)]$, since $\left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_1} \leq \left.\frac{\partial E_o(\lambda)}{\partial \lambda}\right|_{\lambda=1} = R_{cr}$. By Lemma 4.3, we see that

$$\min_{R_{cr}\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W) + e(R,Q)] = E_r(R_{cr},W) + e(R_{cr},Q), \tag{4.27}$$

i.e., the global minimum of (4.24) is attained at R_{cr}. On the other hand,

$$\begin{aligned} &\min_{H(Q)\leq R\leq R_{cr}} [E_r(R,W) + e(R,Q)] \\ = \quad & E_o(1) - 2\left\{\max_{P:H(Q)\leq H(P)\leq R_{cr}} \sum_{i=1}^{|\mathcal{S}|} p_i \log\frac{\sqrt{q_i}}{p_i}\right\} \\ = \quad & E_o(1) - 2\log\left(\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}\right), \end{aligned} \tag{4.28}$$

where the minimum is attained at $R = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < R_{cr}$.

Comparison of (4.27) with (4.28) reveals that

$$\min_{H(Q)\leq R\leq \min(C,\log|\mathcal{S}|)} [E_r(R,W) + e(R,Q)] = E_o(1) - 2\log\left(\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}\right), \tag{4.29}$$

where the minimum is attained at $R_m \triangleq H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$.

Finally, we complete the computation by considering $\log|\mathcal{S}| \leq R_{cr}$. In this case, it is seen that the minimization is the same as in (4.21) since

$$H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < \log|\mathcal{S}| \leq R_{cr},$$

hence the result is the same as in part b) of Case 2. Recall that

$$E_s(\rho) = (1+\rho)\log\left(\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho}}\right),$$

and examining (4.14), (4.23) and (4.29), we remark that

$$\min_{H(Q)\le R\le \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)] = E_o(\rho) - E_s(\rho), \tag{4.30}$$

where $\rho = \min(\lambda_1, 1)$ and λ_1 is the root of the equation

$$\varphi(\lambda) \triangleq H(\mathbf{p}(\lambda)) - R(\lambda) = 0. \tag{4.31}$$

$\min(\lambda_1, 1)$ means that if $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$, then the root λ_1 exists in [0,1), otherwise if $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \le R_{cr}$ then the root λ_1 is no less than 1 and we let $\rho = 1$ in (4.30). In fact, differentiating $E(\lambda) \triangleq E_o(\lambda) - E_s(\lambda)$ with respect to λ, we obtain

$$\begin{aligned} \frac{\partial E(\lambda)}{\partial \lambda} &= \frac{\partial E_o(\lambda)}{\partial \lambda} - \frac{\partial E_s(\lambda)}{\partial \lambda} \\ &= R(\lambda) - H(p_i(\lambda)). \end{aligned} \tag{4.32}$$

Moreover,

$$\frac{\partial^2 E(\lambda)}{\partial^2 \lambda} = \frac{\partial R(\lambda)}{\partial \lambda} - \frac{\partial H(p_i)(\lambda)}{\partial \lambda} = -\frac{\partial \varphi(\lambda)}{\partial \lambda} < 0. \tag{4.33}$$

Comparison of (4.31) with (4.32) yields

$$\min_{H(Q)\le R\le \min(C,\log|\mathcal{S}|)} [E_r(R,W)+e(R,Q)] = \max_{0\le\lambda\le 1} E_o(\lambda) - E_s(\lambda). \tag{4.34}$$

Remark: If the maximum of $E_o(\lambda) - E_s(\lambda)$ over $\lambda \ge 0$ is attained at $0 \le \lambda \le 1$, which means that the root λ_1 of (4.31) is in $[0,1]$, and hence $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \ge R_{cr}$, the lower bound is attained at $R_m \ge R_{cr}$. Hence, the joint source-channel exponent is determined exactly.

Theorem 4.1 *Given a DMS $\{Q : \mathcal{S}\}$ and an arbitrary DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, the joint-source channel coding reliability function satisfies*

$$E_J(Q, W) \geq \max_{0 \leq \rho \leq 1} [E_o(\rho) - E_s(\rho)] \tag{4.35}$$

if $H(Q) < C = \frac{\partial E_o(\rho)}{\partial \rho}\Big|_{\rho=0}$, where equality holds if $\max_{\rho \geq 0}[E_o(\rho) - E_s(\rho)]$ is attained at $\rho \in [0, 1]$.

Proposition 4.1 *$E(\rho) \triangleq E_o(\rho) - E_s(\rho)$ has the following properties:*

1) $E(\rho)$ is strictly concave in ρ ($\rho \geq 0$).

2) The global maximum of $E(\rho)$ is attained at $0 < \rho_{max} < +\infty$ if and only if $H(Q) < C$ and $R_\infty = \lim_{\rho \to \infty} \frac{E_o(\rho)}{\rho} < \log |\mathcal{S}|$.

3) Denoting $\widetilde{\rho} \triangleq \arg\max_{\rho \geq 0} E(\rho)$, then

$$\begin{cases} H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|} \sqrt{q_i}}\right) \geq R_{cr} \Leftrightarrow \widetilde{\rho} \leq 1, & \textit{and hence } E_J(Q, W) = E(\widetilde{\rho}), \\ H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|} \sqrt{q_i}}\right) < R_{cr} \Leftrightarrow \widetilde{\rho} > 1 & \textit{and hence } E_J(Q, W) \geq E(1). \end{cases}$$

4) Denoting $\widehat{\rho} \triangleq \min(\widetilde{\rho}, 1)$, then the lower bound of joint source-channel exponent is achieved at rate $R_m = H\left(\frac{q_i^{\frac{1}{1+\widehat{\rho}}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\widehat{\rho}}}}\right)$.

Proof: 1), 3) and 4) have been shown in the computation of the joint source-channel exponent lower bound. We note that

$$E(0) = 0, \quad \frac{\partial E(\rho)}{\partial \rho}\bigg|_{\rho=0} = C - H(Q) \quad \text{and} \quad \frac{\partial E(\rho)}{\partial \rho}\bigg|_{\rho=\infty} = R_\infty - \log |\mathcal{S}|,$$

hence 2) follows immediately by considering the concavity of $E(\rho)$. $\square$

We stress that Property 3 is very useful to judge if $E_J(Q, W)$ can be determined exactly or not. If the lower bound to $E_J(Q, W)$ is attained at a rate smaller than R_{cr}, then it has a very simple form: $E(1)$. These conclusions will be used frequently in Section 4.4.2.

Observation: The random-coding lower bound to $E_J(Q, W)$ was originally expressed as the minimum of the sum of $E_r(R,W)$ and $e(R,Q)$ over R. By Theorem 4.1, we show that this minimization problem is equivalent to a more concrete maximization problem of the function $E(\rho)$ with respect to the parameter $0 \le \rho \le 1$, where $E(\rho)$ is the difference between Gallager's channel function $E_o(\rho)$ and source function $E_s(\rho)$. Thus, solving the maximization of $E(\rho)$, whose properties are provided in Proposition 4.1, boils down to the determination of $E_o(\rho)$. Although $E_o(\rho)$ does not admit analytical expression for arbitrary DMCs, it can always be obtained numerically via Arimoto's algorithm [10]. Therefore, we can always compute numerically (as we will show in Section 4.1.3) the lower bound to $E_J(Q,W)$. Furthermore, for class of DMCs satisfying a certain symmetry, $E_o(\rho)$ can be analytically solved, hence yielding a parametric expression for the lower bound to $E_J(Q,W)$. This is examined next.

4.1.2 DMS over Symmetric Channels

Definition 4.2 *[7] A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ is said to be symmetric if the rows of the channel transition matrix $P_{Y|X}(y|x)$ are permutations of each other, and the columns are permutations of each other.*

Definition 4.3 *A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $W \triangleq P_{Y|X}(y|x)$ is said to be quasi-symmetric*[3] *if W can be partitioned along its columns into symmetric sub-arrays $W_1, W_2, ..., W_s$, each W_i having size $|\mathcal{X}| \times |\mathcal{Y}_i|$, where $\bigcup_{i=1}^{s} \mathcal{Y}_i = \mathcal{Y}$ and $\mathcal{Y}_i \bigcap \mathcal{Y}_j = \emptyset$ $\forall i \neq j$.*

[3] Our notion of "quasi-symmetry" is actually identical to Gallager's notion of "symmetry" (see [4, p. 94]). We use the term "quasi-symmetric channel" since we have already denoted the term "symmetric channel" in Definition 4.2.

For example, the following channels with transition matrices

$$W = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}, \quad W = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix}, \quad \text{and} \quad W = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{bmatrix}$$

are all quasi-symmetric. Note that a symmetric DMC must be quasi-symmetric, and we have the following important property for a quasi-symmetric DMC.

Lemma 4.4 *For any quasi-symmetric DMC, $E_o(\rho)$ (defined in (2.16)) is maximized by a uniform input distribution.*

The proof of this lemma is lengthy, and we defer it to the Appendix. Consider a quasi-symmetric DMC with channel transition matrix $W \triangleq P_{Y|X}(y|x)$. Suppose W can be partitioned along its columns into symmetric sub-matrices $W_1, W_2, ..., W_s$, each W_i having size $|\mathcal{X}| \times |\mathcal{Y}_i|$. Now denote the transition probabilities in any column of the symmetric sub-matrix W_i by $\{p_{i1}, p_{i2}, ..., p_{i|\mathcal{X}|}\}$, $i = 1, 2, ..., s$. By Lemma 4.4, we can write $E_o(\rho)$ as

$$E_o(\rho) = (1+\rho)\log|\mathcal{X}| - \log\left\{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\}. \tag{4.36}$$

Differentiating $E_o(\rho)$ with respect to ρ yields

$$\frac{\partial E_o(\rho)}{\partial \rho} = \log|\mathcal{X}| - \frac{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho} H\left(\frac{p_{ij}^{\frac{1}{1+\rho}}}{\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}}}\right)}{\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}}. \tag{4.37}$$

Consider a DMS with distribution Q and alphabet $\mathcal{S}$ that is to be sent over this quasi-symmetric DMC W. By Theorem 4.1, if

$$H(Q) < C = \left.\frac{\partial E_o(\rho)}{\partial \rho}\right|_{\rho=0} = \log|\mathcal{X}| - \frac{1}{|\mathcal{X}|}\sum_{i=1}^{s} |\mathcal{Y}_i| \left(\sum_{j=1}^{|\mathcal{X}|} p_{ij}\right) H\left(\frac{p_{ij}}{\sum_{j=1}^{|\mathcal{X}|} p_{ij}}\right),$$

then

$$E_J(Q,W) \geq \max_{0\leq\rho\leq 1}\left[(1+\rho)\log|\mathcal{X}| - \log\left\{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\}\right.$$
$$\left. -(1+\rho)\log\left(\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho}}\right)\right]. \tag{4.38}$$

By Proposition 4.1, we can deduce the following result.

Corollary 4.1 *Consider a DMS $\{Q:\mathcal{S}\}$ and a quasi-symmetric DMC $\{W:X\to Y\}$,*
1) If

$$\begin{cases} \frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H\left(\frac{p_{ij}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}}\right)+H(Q)\leq\log|\mathcal{X}|, \\ \frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2 H\left(\frac{\sqrt{p_{ij}}}{\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}}\right)}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2}+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)\geq\log|\mathcal{X}|, \end{cases}$$

then the joint source-channel coding exponent is exactly given by

$$E_J(Q,W) = (1+\rho^*)\log|\mathcal{X}| - \log\left\{\left[\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho^*}}\right)^{1+\rho^*}\right]\left(\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho^*}}\right)^{1+\rho^*}\right\}, \tag{4.39}$$

where ρ^* is the root of equation

$$\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}H\left(\frac{p_{ij}^{\frac{1}{1+\rho}}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}}\right)}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}} + H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho}}}\right) = \log|\mathcal{X}|.$$

2) If

$$\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2 H\left(\frac{\sqrt{p_{ij}}}{\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}}\right)}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2}+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)\leq\log|\mathcal{X}|,$$

then the lower bound of joint source-channel coding exponent is given by

$$E_J(Q,W)\geq 2\log|\mathcal{X}| - \log\left\{\left[\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}\sqrt{p_{ij}}\right)^2\right]\left(\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}\right)^2\right\}. \tag{4.40}$$

Note that for $\frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H\left(\frac{p_{ij}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}}\right)+H(Q)>\log|\mathcal{X}|$, $E_J(Q,W)=0$.

Observation: Symmetric Channels

A special case of Corollary 4.1 is that if the DMC is symmetric, then each column of the probability transition matrix W is the permutation of the same set of transition probabilities, denoted by $\{p_1, p_2, ..., p_{|\mathcal{X}|}\}$, and Corollary 4.1 is simplified to

1) If $H\left(\frac{p_i}{\sum_{i=1}^{|\mathcal{X}|}p_i}\right)+H(Q)\le\log|\mathcal{X}|$ and $H\left(\frac{\sqrt{p_i}}{\sum_{i=1}^{|\mathcal{X}|}\sqrt{p_i}}\right)+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)\ge\log|\mathcal{X}|$, then the joint source-channel coding exponent is exactly given by

$$E_J(Q,W)=(1+\rho^*)\log|\mathcal{X}|-\log|\mathcal{Y}|-(1+\rho^*)\log\left[(\sum_{i=1}^{|\mathcal{X}|}p_i^{\frac{1}{1+\rho^*}})(\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho^*}})\right], \quad (4.41)$$

where ρ^* is the root of equation $H\left(\frac{p_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|\mathcal{X}|}p_i^{\frac{1}{1+\rho}}}\right)+H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho}}}\right)=\log|\mathcal{X}|$.

2) If $H\left(\frac{\sqrt{p_i}}{\sum_{i=1}^{|\mathcal{X}|}\sqrt{p_i}}\right)+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)\le\log|\mathcal{X}|$, then the lower bound of joint source-channel coding exponent is given by

$$E_J(Q,W)\ge\log\frac{|\mathcal{X}|^2}{|\mathcal{Y}|}-2\log\left[(\sum_{i=1}^{|\mathcal{X}|}\sqrt{p_i})(\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i})\right]. \quad (4.42)$$

For $H\left(\frac{p_i}{\sum_{i=1}^{|\mathcal{X}|}p_i}\right)+H(Q)>\log|\mathcal{X}|$, $E_J(Q,W)=0$.

Example 4.1 (*Binary DMS over Binary Symmetric Channel* (*BSC*)) *A BSC has 2 inputs, 2 outputs with error transition probability* ε, *as shown in Figure 4.1. Consider a binary source with probability distribution* $\{q, 1-q\}$.

1) If $h_b(\varepsilon)+h_b(q)\le 1$ *and* $h_b\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right)+h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right)\ge 1$, *where* $h_b(\cdot)$ *is the binary entropy function, then*

$$E_J(Q,W)=\rho^*-(1+\rho^*)\log\left[\left(\varepsilon^{\frac{1}{1+\rho^*}}+(1-\varepsilon)^{\frac{1}{1+\rho^*}}\right)\left(q^{\frac{1}{1+\rho^*}}+(1-q)^{\frac{1}{1+\rho^*}}\right)\right], \quad (4.43)$$

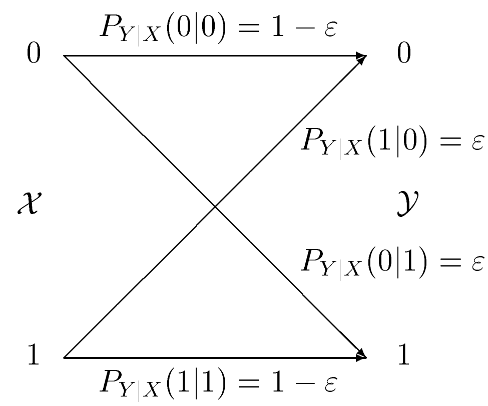

Figure 4.1: Binary symmetric channel with parameter ε.

where ρ^* *is the root of equation*

$$h_b\left(\frac{\varepsilon^{\frac{1}{1+\rho}}}{\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}}\right) + h_b\left(\frac{q^{\frac{1}{1+\rho}}}{q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}}\right) = 1.$$

2) If $h_b\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right) + h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right) < 1$ *then*

$$E_J(Q,W) \geq 1 - 2\log\left[(\sqrt{\varepsilon}+\sqrt{1-\varepsilon})(\sqrt{q}+\sqrt{1-q})\right]. \tag{4.44}$$

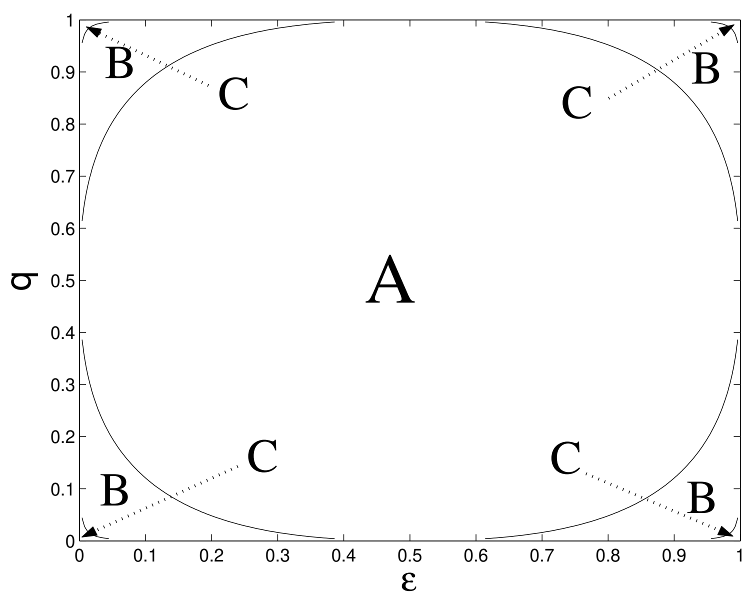

Figure 4.2: The regions for the (ε, q) pairs in the binary DMS and BSC system.

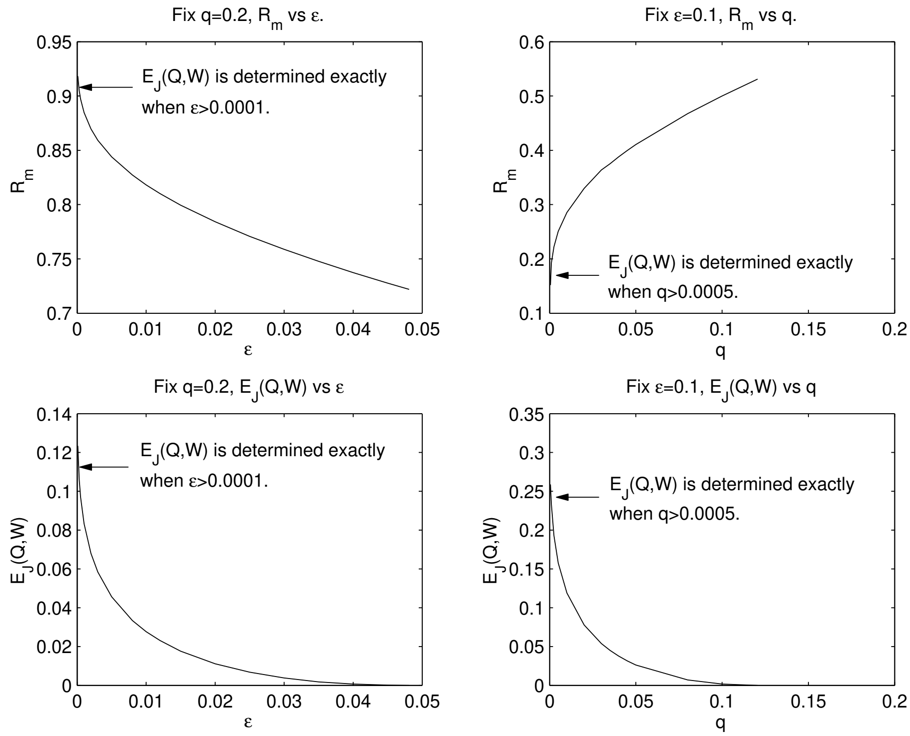

Figure 4.3: The joint source-channel exponent lower bound vs BSC parameter ε (left side), and the joint source-channel exponent lower bound vs binary source distribution q (right side). R_m is the rate attaining the lower bound of $E_J(Q,W)$.

In Figure 4.2, we see that if the pair (ε, q) is located in region B, then the corresponding joint source-channel exponent can be determined exactly by (4.43). If (ε, q) is located in region C, then (4.44) holds. When $(\varepsilon, q) \in A$, $E_J(Q,W)$ is zero, and the error probability of this communication system will converge to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in B \cup C$.

From Figure 4.3 we observe that in practical communication systems we always hope that the noise distribution ε is as small as possible. When ε decreases, we can attain a higher reliability exponent. However, once ε is fixed, as we decrease q, the joint source-channel exponent increases, i.e., when q (or $1-q$) becomes very small ("skewed" source), we can attain a higher exponent.

Example 4.2 (*DMS over N-ary Strong Symmetric Channel (N-ary SSC)*) *An N-ary SSC has N inputs, N outputs with error transition probability matrix*

$$P_{Y|X}(y|x) = \begin{bmatrix} 1-\varepsilon & \frac{\varepsilon}{N-1} & \frac{\varepsilon}{N-1} \cdots & \frac{\varepsilon}{N-1} \\ \frac{\varepsilon}{N-1} & 1-\varepsilon & \frac{\varepsilon}{N-1} \cdots & \frac{\varepsilon}{N-1} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \frac{\varepsilon}{N-1} & \frac{\varepsilon}{N-1} & \frac{\varepsilon}{N-1} \cdots & 1-\varepsilon \end{bmatrix}. \tag{4.45}$$

Consider an M-ary source with probability distribution $Q=\{q_1,...q_M\}$.

1) If

$$\begin{cases} h_b(\varepsilon)+\varepsilon\log(N-1)+H(Q)\le \log N, \\ h_b\left(\frac{\sqrt{\varepsilon(N-1)}}{\sqrt{\varepsilon(N-1)}+\sqrt{1-\varepsilon}}\right)+\frac{\sqrt{\varepsilon(N-1)}}{\sqrt{\varepsilon(N-1)}+\sqrt{1-\varepsilon}}\log(N-1)+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^M\sqrt{q_i}}\right)\ge \log N, \end{cases}$$

then

$$E_J(Q,W)=\rho^*\log N-(1+\rho^*)\log\left[\left((1-\varepsilon)^{\frac{1}{1+\rho^*}}+(N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho^*}}\right)\left(\sum_{i=1}^M q_i^{\frac{1}{1+\rho^*}}\right)\right], \tag{4.46}$$

where ρ^* is the root of

$$h_b\left(\frac{(N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho}}}{(1-\varepsilon)^{\frac{1}{1+\rho}}+(N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho}}}\right)+H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^M q_i^{\frac{1}{1+\rho}}}\right)=\log N.$$

2) If

$$h_b\left(\frac{\sqrt{\varepsilon(N-1)}}{\sqrt{\varepsilon(N-1)}+\sqrt{1-\varepsilon}}\right)+\frac{\sqrt{\varepsilon(N-1)}}{\sqrt{\varepsilon(N-1)}+\sqrt{1-\varepsilon}}\log(N-1)+H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^M\sqrt{q_i}}\right)\le\log N,$$

then

$$E_J(Q,W)\ge\log N-2\log\left[\left(\sqrt{1-\varepsilon}+\sqrt{(N-1)\varepsilon}\right)\left(\sum_{i=1}^M\sqrt{q_i}\right)\right]. \tag{4.47}$$

In Figure 4.4, we see that if the pair (ε,q) is located in region B, then the corresponding joint source-channel exponent can be determined exactly by (4.46). If (ε,q)

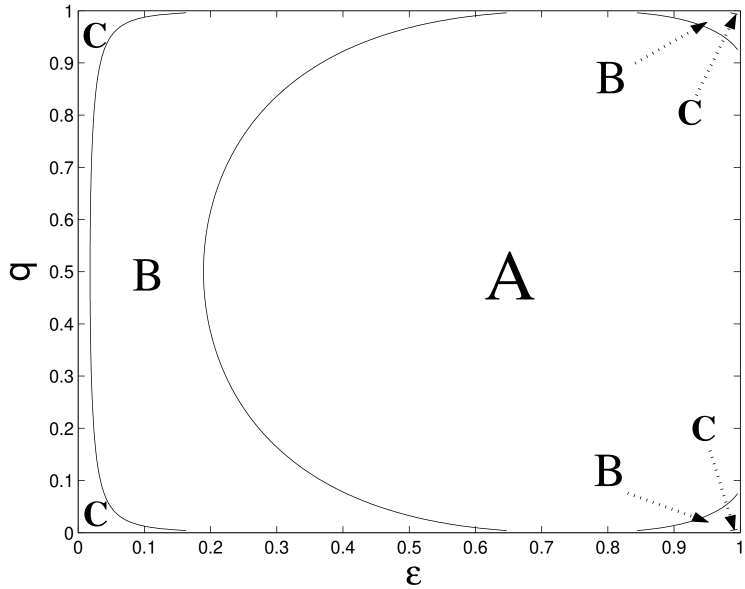

Figure 4.4: The regions for the (ε, q) pairs in the binary DMS and 4-ary SSC system.

is located in region C, then (4.47) holds. When $(\varepsilon, q) \in A$, the error probability of this communication system will converge to 1 for n sufficiently large. Figure 4.5 shows for various values of (ε, q) the joint source-channel exponent lower bound.

Example 4.3 *Consider a DMS $\{Q : \mathcal{S}\}$ and a binary erasure channel (BEC) $\{W : X \to Y\}$ with transition probability $W \triangleq P_{Y|X}$ with parameter α as shown in Figure 4.6. The channel probability transition matrix*

$$W = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}, \tag{4.48}$$

is quasi-symmetric since it can be partitioned into two symmetric sub-matrices

$$W_1 = \begin{bmatrix} 1-\alpha & 0 \\ 0 & 1-\alpha \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \tag{4.49}$$

In sub-matrix W_1, $p_{11} = 1-\alpha$, $p_{12} = 0$, $|\mathcal{Y}_1| = 2$. In sub-matrix W_2, $p_{11} = p_{22} = \alpha$, $|\mathcal{Y}_2| = 1$. Applying Corollary 4.1, we have

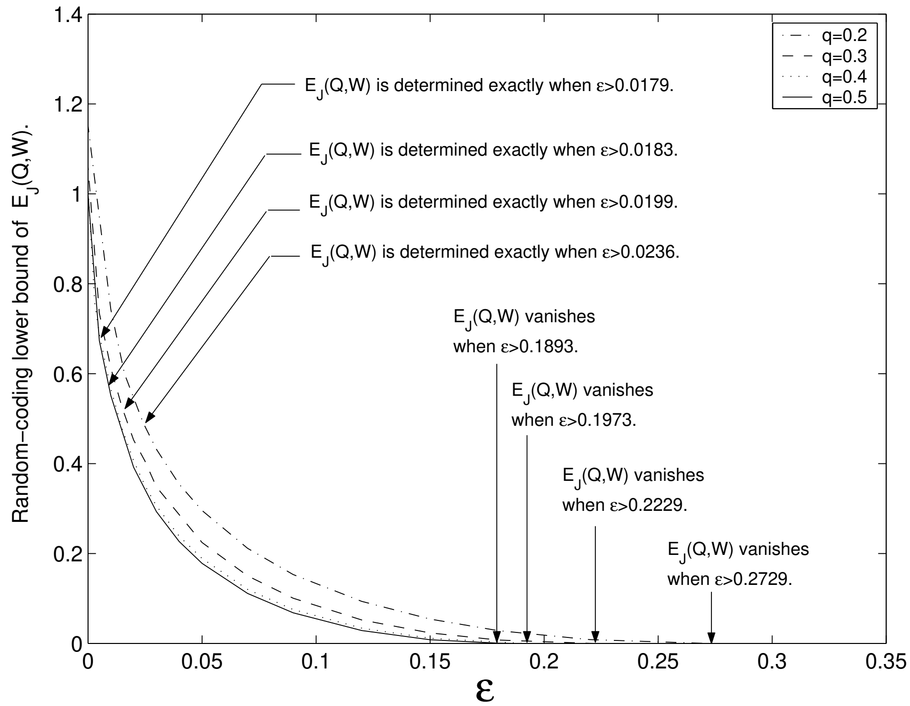

Figure 4.5: The joint source-channel exponent lower bound of the binary DMS $\{q, 1-q\}$ and 4-ary SSC system.

1) If $H(Q)+\alpha \leq 1$ and $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|S|}\sqrt{q_i}}\right) \geq \frac{1-\alpha}{1+\alpha}$, then the joint source-channel exponent is exactly given by

$$E_J(Q,W) = \rho^* - \log[(1-\alpha)+\alpha 2^{\rho^*}] - (1+\rho^*)\log\left(\sum_{i=1}^{|S|} q_i^{\frac{1}{1+\rho^*}}\right), \tag{4.50}$$

where ρ^* is the root of equation

$$H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho}}}\right) = \frac{1-\alpha}{(1-\alpha)+\alpha 2^{\rho}}.$$

2) If $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|S|}\sqrt{q_i}}\right) \leq \frac{1-\alpha}{1+\alpha}$, then the lower bound of joint source-channel exponent is

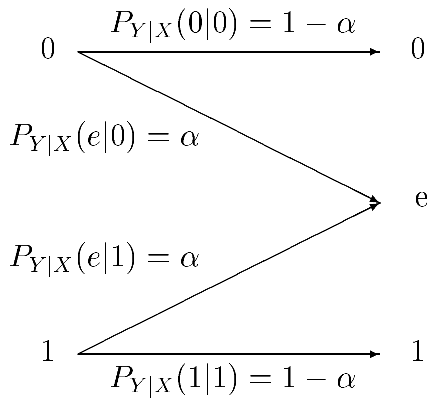

Figure 4.6: Binary erasure channel with parameter α.

given by

$$E_J(Q,W) \geq 1 - \log(1+\alpha) - 2\log\left(\sum_{i=1}^{|S|}\sqrt{q_i}\right). \tag{4.51}$$

For $H(Q) + \alpha \geq 1,\ E_J(Q,W) = 0.$

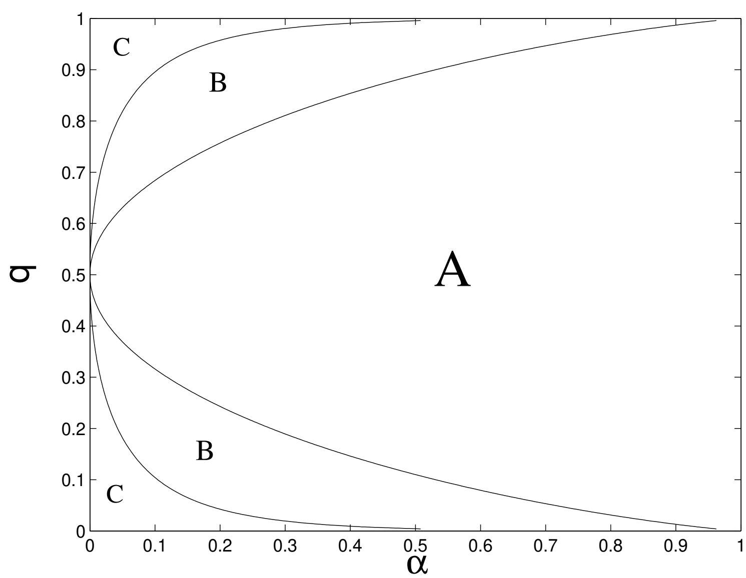

Figure 4.7: The regions for the (α, q) pairs in the binary DMS and BEC system.

For example, if the source is binary $\{q, 1-q\}$, Figure 4.7 shows that if the pair (ε, q) in the binary DMS and BEC system is located in region B, then the corresponding joint source-channel exponent can be determined exactly. If (ε, q) is located in region C, then (4.51) holds. When $(\varepsilon, q) \in A$, the error probability of this communication system will converge to 1 for n sufficiently large.

4.1.3 Arimoto's Algorithm: DMS over Arbitrary DMC

As we have mentioned before, for arbitrary DMC, we cannot write out the parametric form of $E_o(\rho)$ because we do not know under what channel input distribution the maximization in (2.16) can be achieved. Thus we are not able to directly maximize $E_o(\rho) - E_s(\rho)$ for an arbitrary DMC, if the transition matrix is not quasi-symmetric or has a large input or output alphabet.

To investigate how to maximize the right side of (4.35), we write

$$\begin{aligned} &\max_{0\le\rho\le 1}[E_o(\rho) - E_s(\rho)] \\ = \quad &\max_{0\le\rho\le 1}\max_{P_X}\left\{-(1+\rho)\log\left[\sum_{i=1}^{|\mathcal{S}|} Q_i^{\frac{1}{1+\rho}}\right] - \log\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y \mid x)\right)^{1+\rho}\right\}. \end{aligned}$$

It is possible to rewrite the above maximization as

$$\begin{aligned} &\max_{0\le\rho\le 1}\max_{P_X}\max_{\Phi_{X|Y}}\left\{-(1+\rho)\log\left[\sum_{i=1}^{|\mathcal{S}|} Q_i^{\frac{1}{1+\rho}}\right] - \right. \\ &\left. \log\left[\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_X^{1+\rho}(x) P_{Y|X}^{\frac{1}{1+\rho}}(y \mid x)\Phi_{X|Y}(x|y)^{-\rho}\right]\right\} \\ = \quad &\max_{0\le\rho\le 1}\left\{-(1+\rho)\log\left[\sum_{i=1}^{|\mathcal{S}|} Q_i^{\frac{1}{1+\rho}}\right] + \rho\left[\max_{P_X}\max_{\Phi_{X|Y}} G_\rho(P_X, \Phi_{X|Y})\right]\right\}, \quad (4.52) \end{aligned}$$

where

$$G_\rho(P_X, \Phi_{X|Y}) = \frac{-1}{\rho}\log\left[\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y \mid x)\left\{\frac{\Phi_{X|Y}(x \mid y)}{P_X(x)}\right\}^{-\rho}\right],$$

and $\Phi_{X|Y}(x|y)$ is an arbitrary transition probability matrix from the channel $\{W^{-1} : Y \to X\}$, often called a dummy "backward" channel. The form (4.52) was introduced by Blahut [8] and Arimoto [10]. To compute the joint source-channel exponent lower bound (4.35), we therefore can apply Arimoto's algorithm [10].

From [10], we know that Arimoto's iterative algorithm is to maximize $G_\rho(P_X, \Phi_{X|Y})$ over P_X and $\Phi_{X|Y}$ based on a given ρ, but our problem is to maximize (4.52) over ρ from 0 to 1. Although we do not have the parametric expression of $E_o(\rho)$, we note by Proposition 4.1 that $E(\rho)$ is concave and must has a global maximum at a finite $\rho \geq 0$. Therefore, we can apply these properties of $E(\rho)$ to search for the maximizer ρ^* as follows: at each iteration s, we choose $t+1$ ($t \geq 3$) points $\rho_1^{(s)}, \rho_1^{(s)}, ..., \rho_{t+1}^{(s)}$ to divide the interval $\mathcal{I}^{(s)}$ ($\mathcal{I}^{(1)} = [0,1]$) into t consecutive sub-intervals and compute each $E(\rho_i^{(s)})$ ($i = 1, ..., t+1$). Suppose $E(\rho_j^{(s)})$ is the biggest among the $t+1$ numbers, then ρ^* must be in the new interval

$$\mathcal{I}^{(s+1)} = \begin{cases} \left[\rho_1^{(s)}, \rho_2^{(s)}\right], & \text{if } j = 1, \\ \left[\rho_{j-1}^{(s)}, \rho_{j+1}^{(s)}\right], & \text{if } 1 < j < t+1, \\ \left[\rho_t^{(s)}, \rho_{t+1}^{(s)}\right], & \text{if } j = t+1. \end{cases}$$

Specially, if $E(\rho_j^{(s)}) = E(\rho_{j+1}^{(s)})$ both are the biggest among the $t+1$ numbers, then ρ^* must be in the new interval

$$\mathcal{I}^{(s+1)} = \left[\rho_j^{(s)}, \rho_{j+1}^{(s)}\right].$$

Note that whatever the positive integer t is, we always can choose $t+1$ points in the closed interval $[0,1]$, such that for the first (initial) iteration, we compute $E(\rho_i^{(s)})$ for $t+1$ points, but for the next iteration we only need to compute $E(\rho_i^{(s)})$ for $t-1$ points if $j = 1$ or $j = t+1$ and for $t-2$ points if $1 < j < t+1$. Define the ρ-fidelity error $\delta \triangleq \left|\rho_j^{(s)} - \rho^*\right|$. If we want $\delta < 10^{-k}$, we need to compute approximately at most $\left(t+1+(t-1) \times \left\lceil \frac{3.3219k}{\log(t+1)} \right\rceil\right)$ points in total, where $\lceil x \rceil$ is equal to the smallest integer which is no less than x. To reduce the computation time, it is easy to see that t should be as small as possible, that is, t is chosen to be 3. Considering how to choose the 4 points for each iteration, we can employ *golden section* [5, Section 2.3] search[4], so that

[4] *Golden section* search is also called a limiting form of *Fibonacci* search in some books.

the computation iterations is approximately equal to $(\lceil 4.784k \rceil)$ for a ρ-fidelity error δ such that $\delta < 10^{-k}$.

The algorithm for computing the source-channel coding exponent lower bound (4.35) is as follows,

1) Initialize $E^{(0)} = 0$, let $s = 1$, and set $\rho^{(s)}(1) = 0$, $\rho^{(s)}(2) = 0.382$, $\rho^{(s)}(3) = 0.618$, $\rho^{(s)}(4) = 1$;

2) Use Arimoto's iteration algorithm (see Appendix B) to compute

$$E(\rho) = -(1+\rho)\log\left[\sum_{i=1}^{|\mathcal{S}|} Q_i^{\frac{1}{1+\rho}}\right] + \rho\left[\max_{P_X}\max_{\Phi_{X|Y}} G_\rho(P_X, \Phi_{X|Y})\right]$$

for each $\rho^{(s)}(i)$, $i = 1,2,3,4$, and denote the results as $E^{(s)}(i)$, $i = 1,2,3,4$.

3) Let $E^{(s)} = \max\{E^{(s)}(1), E^{(s)}(2), E^{(s)}(3), E^{(s)}(4)\}$. If

$$\frac{E^{(s)} - E^{(s-1)}}{E^{(s)}} < \gamma,$$

stop; else continue.

4) If there exists a j such that $E^{(s)} = E^{(s)}(j) = E^{(s)}(j+1)$, $j = 1,2,3$, then $E^{(s+1)}(1) = E^{(s)}(j)$, $E^{(s+1)}(4) = E^{(s)}(j+1)$, $\rho^{(s+1)}(1) = \rho^{(s)}(j)$, $\rho^{(s+1)}(4) = \rho^{(s)}(j+1)$. Compute $E^{(s+1)}(i), i = 2,3$ for

$$\rho^{(s+1)}(2) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.382,$$
$$\rho^{(s+1)}(3) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.618;$$

else continue.

5) If $E^{(s)} = E^{(s)}(1)$, then $E^{(s+1)}(1) = E^{(s)}(1)$, $E^{(s+1)}(4) = E^{(s)}(2)$, $\rho^{(s+1)}(1) = \rho^{(s)}(1)$, $\rho^{(s+1)}(4) = \rho^{(s)}(2)$. Compute $E^{(s+1)}(i), i = 2,3$ for

$$\rho^{(s+1)}(2) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.382,$$
$$\rho^{(s+1)}(3) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.618.$$

If $E^{(s)} = E^{(s)}(2)$, then $E^{(s+1)}(1) = E^{(s)}(1)$, $E^{(s+1)}(3) = E^{(s)}(2)$, $E^{(s+1)}(4) = E^{(s)}(3)$, $\rho^{(s+1)}(1) = \rho^{(s)}(1)$, $\rho^{(s+1)}(3) = \rho^{(s)}(2)$, $\rho^{(s+1)}(4) = \rho^{(s)}(3)$. Compute $E^{(s+1)}(2)$ for

$$\rho^{(s+1)}(2) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.382.$$

If $E^{(s)} = E^{(s)}(3)$, then $E^{(s+1)}(1) = E^{(s)}(2)$, $E^{(s+1)}(2) = E^{(s)}(3)$, $E^{(s+1)}(4) = E^{(s)}(4)$, $\rho^{(s+1)}(1) = \rho^{(s)}(2)$, $\rho^{(s+1)}(2) = \rho^{(s)}(3)$, $\rho^{(s+1)}(4) = \rho^{(s)}(4)$. Compute $E^{(s+1)}(3)$ for

$$\rho^{(s+1)}(3) = \rho^{(s+1)}(1) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.618.$$

If $E^{(s)} = E^{(s)}(4)$, then $E^{(s+1)}(1) = E^{(s)}(3)$, $E^{(s+1)}(4) = E^{(s)}(4)$, $\rho^{(s+1)}(1) = \rho^{(s)}(3)$, $\rho^{(s+1)}(4) = \rho^{(s)}(4)$. Compute $E^{(s+1)}(i)$, $i = 2, 3$ for

$$\rho^{(s+1)}(2) = \rho^{(s+1)}(4) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.382,$$
$$\rho^{(s+1)}(3) = \rho^{(s+1)}(4) + (\rho^{(s+1)}(4) - \rho^{(s+1)}(1)) \times 0.618.$$

6) Let $s = s + 1$, goto 3).

In Figure 4.8, we determine the joint source-channel exponent lower bound using (4.43), (4.44), for q=0.1, 0.15 and 0.2, and compare with the lower bound obtained by our algorithm. We see that the two group of curves coincide with each other. In [10], it is assumed that $\Phi_{X|Y}(x|y) \geq 0$, but we see that $G_\rho(P_X, \Phi_{X|Y})$ in any given iteration may go to infinite if $\Phi_{X|Y}(x|y)$ is zero, since $\Phi_{X|Y}^{-\rho}(\rho \geq 0)$ appears in (4.52). From the iterative equations (see [10]) uesd to compute $\Phi_{X|Y}$ and P_X, we note that the necessary and sufficient condition for $\Phi_{X|Y} \neq 0$ is $P_{Y|X} \neq 0$. This leads us to suggest a modification if there are zero entries in the channel transition matrix $P_{Y|X}$. Because the reliability function is continuous in each $P_{Y|X}(y|x)$, one could add a very small quantity δ to obtain a new channel $W' \triangleq P'_{Y|X}$ without 0 entries

$$P'_{Y|X}(y_j|x_i) = \begin{cases} P_{Y|X}(y_j|x_i) - \frac{n_i\delta}{m-n_i}, & \text{if } P_{Y|X}(y_j|x_i) \neq 0, \\ \delta, & \text{if } P_{Y|X}(y_j|x_i) = 0, \end{cases} \tag{4.53}$$

such that $E_J(Q, W')$ converges to $E_J(Q, W)$ if δ goes to 0, where n_i is the number of zero's in row i, for each $i = 1, 2, ..., |\mathcal{X}|$ and $j = 1, 2, ..., |\mathcal{Y}|$. For example, the BEC can

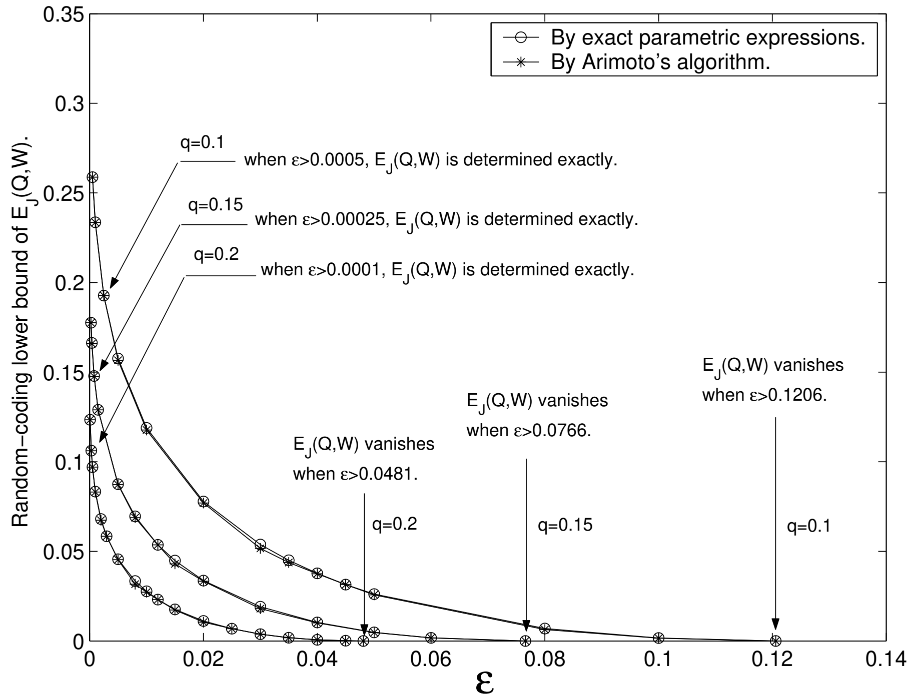

Figure 4.8: The joint source-channel exponent lower bound of binary DMS $\{q, 1-q\}$ and BSC pairs: Arimoto's algorithm vs exact expressions given in Example 4.1

be replaced by the new channel

$$P'_{Y|X} = \begin{bmatrix} 1-\alpha-0.5\delta & \alpha-0.5\delta & \delta \\ \delta & \alpha-0.5\delta & 1-\alpha-0.5\delta \end{bmatrix}, \tag{4.54}$$

then we can apply our algorithm to compute the joint source-channel exponent lower bound for the binary DMS and BEC pairs, see Figure 4.9.

From Figure 4.9 we remark that when we set $\delta = 10^{-6}$, the error resulting from involving the "noise" δ is very small. From Figure 4.11, we show the joint source-channel exponent lower bound for binary DMS and Z-DMC (described in Figure 4.10) pairs, computed using Arimoto's algorithm with $\delta = 10^{-6}$. Note that in the case, no analytical exact expression for the exponent lower bound exists.

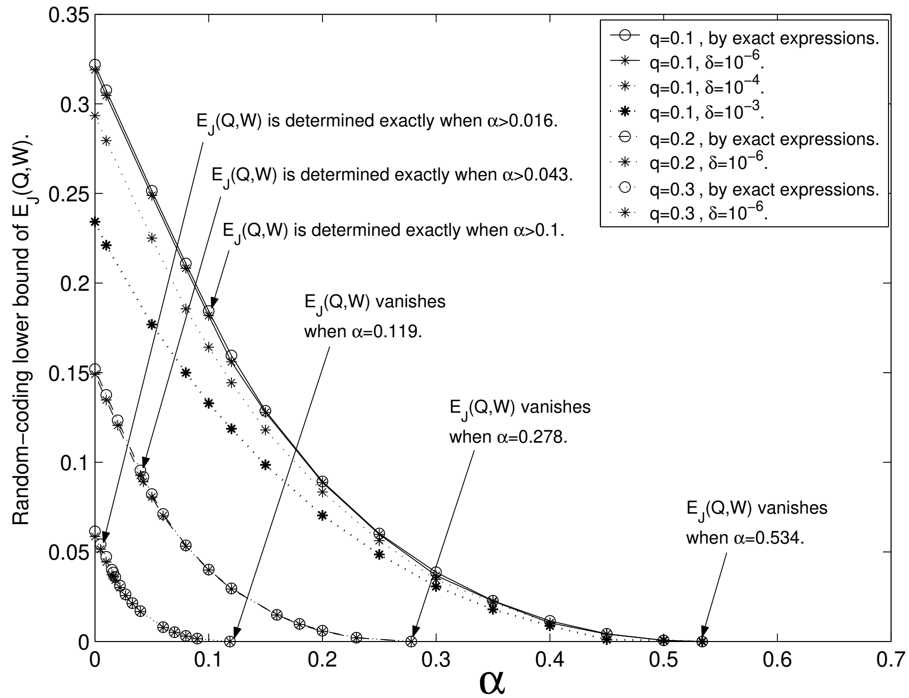

Figure 4.9: The joint source-channel exponent lower bound of binary source $\{q, 1-q\}$ and BEC pairs: Arimoto's algorithm vs exact expressions given in Example 4.3

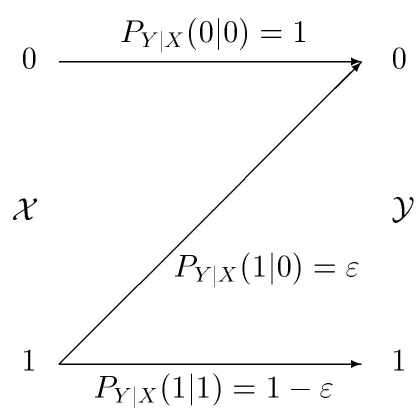

Figure 4.10: Z-DMC with parameter ε.

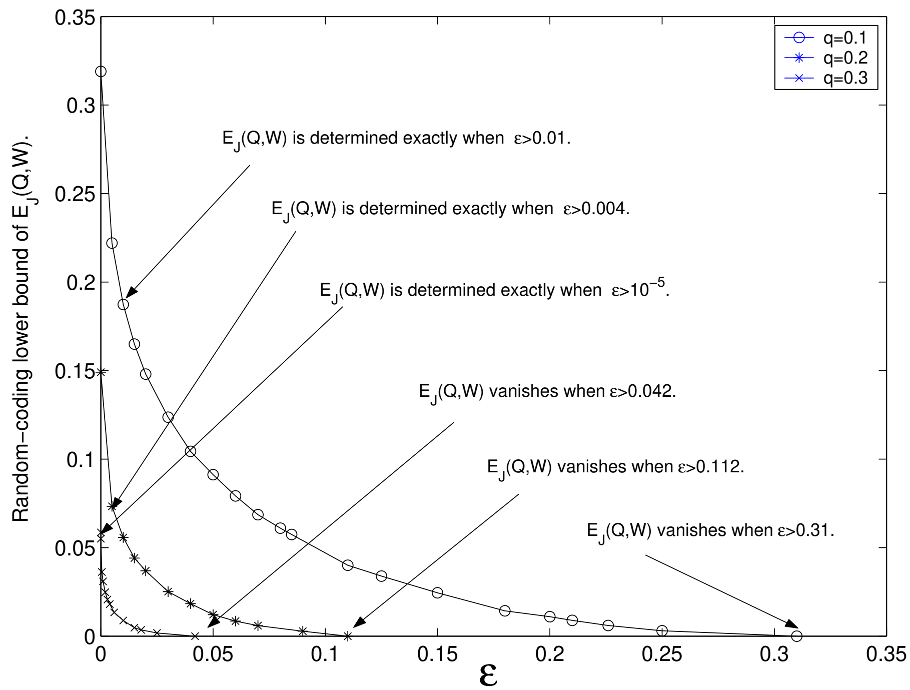

Figure 4.11: The joint source-channel exponent lower bound of the binary source $\{q, 1-q\}$ and Z-DMC pairs, computed using Arimoto's algorithm with $\delta = 10^{-6}$.

4.2 The Joint Source-Channel Expurgated Lower Bound

4.2.1 A General Derivation

In Section 3.2, it is mentioned that if some condition about the channel characteristics holds (see Corollary 3.1), then the joint source-channel coding lower bound satisfies

$$E_J(Q,W) \geq \min_{0 \leq R \leq \min(E_x(1), \log|\mathcal{S}|)} [E_{ex}(R,W) + e(R,Q)].$$

Based on this corollary, we next investigate the analytical computation of

$$\min_{H(Q) \leq R \leq \min(\log|\mathcal{S}|, E_x(1))} \left[E_{ex}(R,W) + e(R,Q)\right], \tag{4.55}$$

for any DMS and DMC with zero-error capacity equal to 0. Like the random-coding lower bound $E_r(R,W)$, since

$$E_{ex}(R,W) = \max_{\rho \geq 1} [-\rho R + E_x(\rho)],$$

by setting the derivative of $[-\rho R + E_x(\rho)]$ with respect to ρ to zero, we obtain the parametric expressions of R and $E_{ex}(R,W)$ in terms of ρ^* (ρ^* maximizes $[-\rho R + E_x(\rho)]$ in [1,+∞), see [4, p. 155]). When $0 < R \leq R'_{cr} = \frac{\partial E_x(\rho^*)}{\partial \rho^*}\Big|_{\rho^*=1}$,

$$\begin{aligned} E_{ex}(R,W) &= E_x(\rho^*) - \rho^* \frac{\partial E_{ex}(\rho^*)}{\partial \rho^*}, & (4.56) \\ \frac{\partial E_x(R,W)}{\partial R} &= -\rho^*, & (4.57) \end{aligned}$$

i.e., $-\rho^*$ is the slope of $E_{ex}(R,W)$, where $R(\rho^*) = \frac{\partial E_x(\rho^*)}{\partial \rho^*}$ is a decreasing function in ρ^* ($1 \leq \rho^* < +\infty$). When $R'_{cr} < R < E_x(1)$, the slope is -1 and hence

$$E_{ex}(R,W) = E_x(1) - R. \tag{4.58}$$

R'_{cr} is described in Lemma 2.2. It is very similar to the critical rate R_{cr} but it is smaller than R_{cr}. In words, for $0 \le R \le R'_{cr}$ (or $(1 \le \rho^* < +\infty)$), $E_{ex}(R,W)$ is strictly convex in R and the slope is $-\rho^*$. For $R'_{cr} \le R \le E_x(1)$ (or $\rho^* = 1$), $E_{ex}(R,W)$ is a straight line with slope -1. First, we assume that $\log|\mathcal{S}| > R'_{cr}$. For the sake of convenience, we may write (4.55) as

$$\begin{aligned}
& \min_{H(Q)\le R\le \min(\log|\mathcal{S}|, E_x(1))} [E_{ex}(R,W) + e(R,Q)] \\
= & \min_{H(Q)\le R\le \min(\log|\mathcal{S}|, E_x(1))} \left[E_{ex}(R,W) + \min_{P:H(P)=R} D(P \parallel Q)\right] \\
= & \min_{P,R:H(Q)\le R=H(P)\le \min(E_x(1),\log|\mathcal{S}|)} [E_{ex}(R,W) + D(P \parallel Q)] \\
= & \min\left\{\min_{P,R:H(Q)\le R=H(P)\le R'_{cr}} [E_{ex}(R,W) + D(P \parallel Q)],\right. \\
& \left.\min_{P,R:R'_{cr}\le R=H(P)\le \min(E_x(1),\log|\mathcal{S}|))} [E_x(1) - R + D(P \parallel Q)]\right\} \\
= & \min\left\{\min_{P,R:H(Q)\le R=H(P)\le R'_{cr}} [E_{ex}(R,W) + D(P \parallel Q)],\right. \\
& \left.\min_{R'_{cr}\le H(P)\le \min(E_x(1),\log|\mathcal{S}|))} [E_x(1) - H(P) + D(P \parallel Q)]\right\}. \qquad (4.59)
\end{aligned}$$

By applying Lagrange multipliers, we obtain

$$\min_{P,R:R=H(P)} \left[E_{ex}(R,W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{q_i}\right] = \max_{\rho}[E_x(\rho) - E_s(\rho)],$$

where the minimum of the left side is attained at $p_i = \frac{q_i^{\frac{1}{1+\rho^*}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho^*}}}$, $i = 1, ..., |\mathcal{S}|$, and rate $R'_m \triangleq H\left(\frac{q_i^{\frac{1}{1+\rho^*}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho^*}}}\right)$, where ρ^* maximizes $E_x(\rho) - E_s(\rho)$. On the other hand, it is easily seen that

$$E_x(1) - H(P) + D(P \parallel Q) = E_x(1) - 2\sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{\sqrt{q_i}}$$

is a convex function in the distribution $P = \{p_1, p_2, ..., p_{|\mathcal{S}|}\}$ and the minimization of the above formula with respect to P is attained at $p_i = \frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}$, $i = 1, ..., |\mathcal{S}|$, and hence the minimum of the left side is attained at rate $R'_m \triangleq H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$. Now we consider the following two cases: $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \leq R'_{cr}$ and $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R'_{cr}$.

Case 1: If $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \leq R'_{cr}$, then it can be shown that $\rho^* \geq 1$ and hence $R_m \leq R'_{cr}$. Therefore,

$$\begin{aligned}
&\min_{P,R:H(Q)\leq R=H(P)\leq R'_{cr}} [E_{ex}(R,W) + D(P \parallel Q)] \\
&= \min_{P,R:R=H(P)} \left[E_{ex}(R,W) + \sum_{i=1}^{|\mathcal{S}|} p_i \log \frac{p_i}{q_i}\right] \\
&= E_x(\rho^*) - E_s(\rho^*) \\
&= E_{ex}(R_m, W) + e(R_m, Q). \qquad (4.60)
\end{aligned}$$

Since $R'_m = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \leq R'_{cr}$, it follows by Lemma 4.3 that

$$\min_{R'_{cr}\leq H(P)\leq \min(E_x(1),\log|\mathcal{S}|)} [E_x(1) - H(P) + D(P \parallel Q)] = E_{ex}(R'_{cr}, W) + e(R'_{cr}, Q). \qquad (4.61)$$

Comparison of (4.60) with (4.61) reveals that

$$\min_{H(Q)\leq R\leq \min(\log|\mathcal{S}|, E_x(1))} [E_{ex}(R,W) + e(R,Q)] = \max_{\rho\geq 1}[E_x(\rho) - E_s(\rho)]. \qquad (4.62)$$

Case 2: If $R'_{cr} \leq H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$, then it can be shown that $\rho^* \leq 1$ and hence $R_m \geq R'_{cr}$. By Lemma 4.3,

$$\min_{P,R:H(Q)\leq R=H(P)\leq R'_{cr}} [E_{ex}(R,W) + D(P \parallel Q)] = E_{ex}(R'_{cr}, W) + e(R'_{cr}, Q). \qquad (4.63)$$

Conversely, since $R'_m = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R'_{cr}$,

$$\min_{R'_{cr}\leq H(P)\leq \min(E_x(1),\log|\mathcal{S}|)} [E_x(1) - H(P) + D(P \parallel Q)] = E_{ex}(R'_m, W) + e(R'_m, Q), \qquad (4.64)$$

Comparison of (4.63) with (4.64) yields

$$\begin{aligned} & \min_{H(Q)\le R\le \min(\log|\mathcal{S}|, E_x(1))} [E_{ex}(R,W)+e(R,Q)] \\ = \;& E_{ex}(R'_m,W)+e(R'_m,Q) \\ = \;& E_x(1)-E_s(1). \end{aligned} \tag{4.65}$$

Finally, recall that we first assumed that $\log|\mathcal{S}| > R'_{cr}$. If $\log|\mathcal{S}| \le R'_{cr}$, we still have

$$\begin{aligned} & \min_{H(Q)\le R\le \min(\log|\mathcal{S}|, E_x(1))} [E_{ex}(R,W)+e(R,Q)] \\ = \;& \min_{H(Q)\le R\le \log|\mathcal{S}|} [E_x(1)-R+e(R,Q)] \\ = \;& E_x(1)-E_s(1). \end{aligned} \tag{4.66}$$

Accordingly, we have the following theorem.

Theorem 4.2 *Consider a DMS and a DMC with zero-error capacity equal to 0, if* $E_{ex}(R,W) = \max_P E_{ex}(R,P,W)$ *is attained for a* P *not depending on* R*, then the joint source-channel exponent expurgated lower bound of Corollary 3.1 is equal to*

$$E_J(Q,W) \ge \max_{\rho\ge 1}[E_x(\rho)-E_s(\rho)], \tag{4.67}$$

attaining at rate $R'_m = H\left(\frac{q_i^{\frac{1}{1+\widehat{\rho}}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\widehat{\rho}}}}\right)$*, where* $\widehat{\rho}$ *achieves (4.67).*

Proposition 4.2 (*Comparison of Random-Coding Lower Bound with Expurgated Lower Bound to* $E_J(Q,W)$) *The random-coding lower bound is better than the expurgated lower bound to* $E_J(Q,W)$ *iff* $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$*; The random-coding lower bound is equal to the expurgated lower bound to* $E_J(Q,W)$ *iff* $R'_{cr} \le H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \le R_{cr}$*; The random-coding lower bound is improved by the expurgated lower bound to* $E_J(Q,W)$ *iff* $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < R'_{cr}$*, i.e.,*

$$E_J(Q,W) \ge \max_{\rho\ge 1}[E_x(\rho)-E_s(\rho)] \ge \max_{0\le\gamma\le 1}[E_o(\gamma)-E_s(\gamma)],$$

iff $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < R'_{cr} = \left.\frac{\partial E_x(\rho)}{\partial \rho}\right|_{\rho=1}$.

Proof: When $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < R'_{cr}$, it also implies that $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) < R_{cr}$. Then the random-coding lower bound is attained at $R_m = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$ and the expurgated lower bound is attained at $R'_m = H\left(\frac{q_i^{\frac{1}{1+\widehat{\rho}}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\widehat{\rho}}}}\right)$ where $\widehat{\rho} > 1$. It follows that the random-coding lower bound to $E_J(Q,W)$ is given by

$$\begin{aligned} E_r(R_m,W) + e(R_m,Q) &< E_r(R'_m,W) + e(R'_m,Q) \\ &< E_{ex}(R'_m,W) + e(R'_m,Q), \end{aligned}$$

which is equal to the expurgated lower bound to $E_J(Q,W)$.

When $R'_{cr} \le H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \le R_{cr}$, the random-coding lower bound is equal to $E_o(1) - E_s(1)$, and the expurgated lower bound is equal to $E_x(1) - E_s(1)$. Thus the two lower bounds are identical since $E_o(1) = E_x(1)$.

When $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$, it also implies that $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R'_{cr}$. Then the random-coding lower bound is attained at $R_m = H\left(\frac{q_i^{\frac{1}{1+\widetilde{\rho}}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\widetilde{\rho}}}}\right)$ where $0 < \widetilde{\rho} < 1$ and the expurgated lower bound is attained at $R'_m = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right)$. It follows that the random-coding lower bound to $E_J(Q,W)$ is

$$\begin{aligned} E_r(R_m,W) + e(R_m,Q) &> E_{ex}(R_m,W) + e(R_m,Q) \\ &> E_{ex}(R'_m,W) + e(R'_m,Q), \end{aligned}$$

which is the expurgated lower bound to $E_J(Q,W)$. $\square$

4.2.2 DMS over Equidistant DMC

From [2], we know that a channel input distribution P_0 attains the $E_x(\rho)$ (described in (2.19)) if and only if it maximizes $E_{ex}(R, P_0, W)$ (described in (3.8)), and such a P_0 with the corresponding ρ_0 (ρ_0 achieves $E_{ex}(R,W)$) satisfies the Kuhn-Tucker conditions

$$\sum_x P_0(x) \exp\left\{-\frac{1}{\rho_0} d_B(x,\widetilde{x})\right\} = c \tag{4.68}$$

for every $\widetilde{x} \in X$ with $P(\widetilde{x}) > 0$, where c does not depend on $\widetilde{x}$ (see [3, p. 193]).

Definition 4.4 (*Equidistant Channel*) *A DMC $W \triangleq P_{Y|X}$ is called equidistant if there exists a number $\beta > 0$ such that for all pairs of inputs $x \neq \widetilde{x}$,*

$$\sum_y \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\widetilde{x})} = \beta. \tag{4.69}$$

For example, all binary input channels are equidistant.

Remark: By *Hölder*'s inequality [4] it is easy to see that

$$\beta \leq [\sum_y P_{Y|X}(y|x)]^{\frac{1}{2}}[\sum_y P_{Y|X}(y|\widetilde{x})]^{\frac{1}{2}} = 1.$$

This implies that all equidistant DMC's have zero-error capacity equal to 0 since

$$\begin{aligned}
E_{ex}(0,W) &= \lim_{R\to 0} E_{ex}(R,W) \\
&= -\sum_{x,\widetilde{x}} P_X(x)P_X(\widetilde{x}) \log\left[\sum_y \sqrt{P_{Y|X}(y|x)P_{Y|X}(y|\widetilde{x})}\right] \\
&= -\sum_{x,\widetilde{x}:\widetilde{x}\neq x} P_X(x)P_X(\widetilde{x}) \log\beta - \sum_x P_X^2(x)\log\left[\sum_y P_{Y|X}(y|x)\right] \\
&= -\sum_{x,\widetilde{x}:\widetilde{x}\neq x} P_X(x)P_X(\widetilde{x}) \log\beta < +\infty,
\end{aligned}$$

for every P_X such that $P_X(x) > 0$ for every $x \in \mathcal{X}$, which is a sufficient and necessary condition for a DMC with zero-error capacity equal to 0, see [3, p. 187]. Moreover, for an equidistant channel, it can be easily verified that Kuhn-Tucker conditions (4.68) are satisfied by uniform input distributions.

Lemma 4.5 *For an equidistant channel, $E_x(\rho)$ is achieved in the range $\rho \geq 1$ by a channel input distribution*

$$P_X(x) = \left(\frac{1}{N}\right), \qquad \forall \quad x \in \mathcal{X}.$$

The above property of equidistant channels was first given by Jelinek [6, p.231].

Lemma 4.6 [2] *For a given DMS $\{Q : \mathcal{S}\}$ and an equidistant DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$, the joint source-channel exponent satisfies*

$$E_J(Q, W) \geq \min_{H(Q) \leq R \leq \min(\log|\mathcal{S}|, C)} [E_{ex}(R, W) + e(R, Q)].$$

By Lemma 4.5, we can write $E_x(\rho)$ (described in (2.19)) as

$$E_x(\rho) = -\rho \log\left(\frac{N-1}{N}\beta^{\frac{1}{\rho}} + \frac{1}{N}\right) \quad for \quad \rho \geq 1, \tag{4.70}$$

where N is the input alphabet size. Now applying Theorem 4.2, yields the following result.

Corollary 4.2 *For a given DMS $\{Q : \mathcal{S}\}$ and an equidistant DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$ with parameter β, if*

$$H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|S|}\sqrt{q_i}}\right) + \log\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta + \frac{1}{|\mathcal{X}|}\right) \leq \frac{\beta\log\beta}{\beta + \frac{1}{|\mathcal{X}|-1}}, \tag{4.71}$$

the improved joint source-channel coding lower bound satisfies

$$E^*(Q, W) \geq E_x(\rho_1) - E_s(\rho_1), \tag{4.72}$$

where ρ_1 is the root of the equation

$$H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|S|} q_i^{\frac{1}{1+\rho}}}\right) + \log\left(\frac{|\mathcal{X}|-1}{|\mathcal{X}|}\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|}\right) = \frac{(\rho-1)\beta^{\frac{1}{\rho}}\log\beta}{\beta^{\frac{1}{\rho}} + \frac{1}{|\mathcal{X}|-1}}, \tag{4.73}$$

and $E_x(\rho)$ is given in (4.70).

Example 4.4 *Consider a binary DMS $\{q, 1-q\}$ and a BEC with parameter α (note that for the BEC, the parameter β is equal to α),*

$$E_x(\rho) = \rho - \rho\log\left(\alpha^{\frac{1}{\rho}} + 1\right). \tag{4.74}$$

From Figure 4.12, we see that if the pair (α, q) is located only in region C_2, then

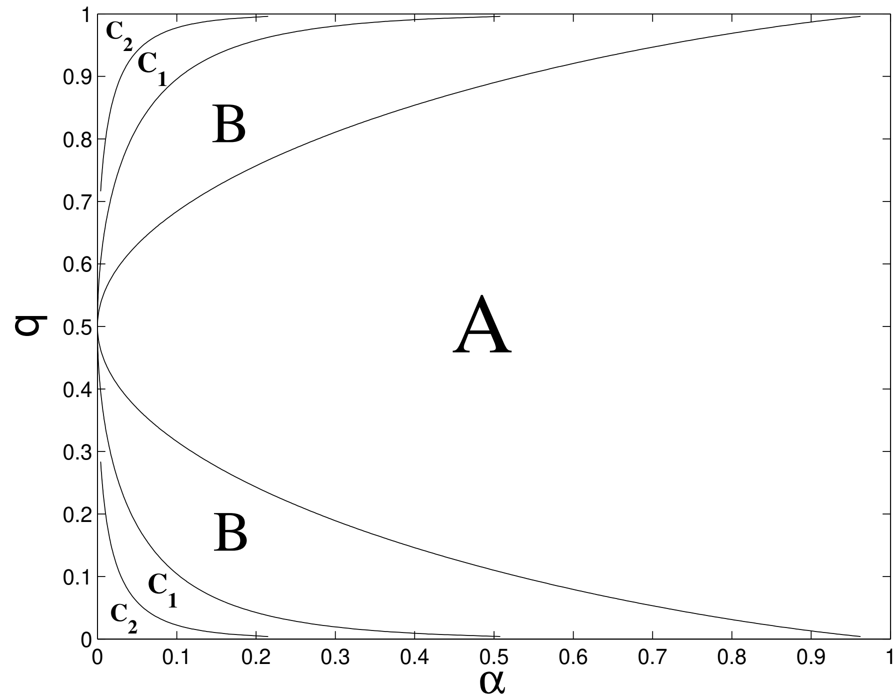

Figure 4.12: The regions for the (α, q) pairs in the binary DMS and BEC system.

the joint source-channel exponent lower bound can be improved by (4.72) (regions A and B were already described in Figure 4.7; also, $C_1 \bigcup C_2 = C$ *where C is described in Figure 4.7). For example, if the source distribution is Q={0.1,0.9}, the lower bound is improved for* $\alpha < 0.0297$*, and if Q={0.2,0.8}, it is improved for* $\alpha < 0.0102$*, see Figure 4.13.*

4.3 The Joint Source-Channel Sphere-Packing Upper Bound

The computation for the joint source-channel coding upper bound is similar to that for the lower bound (of Section 4.1), so we directly state the results.

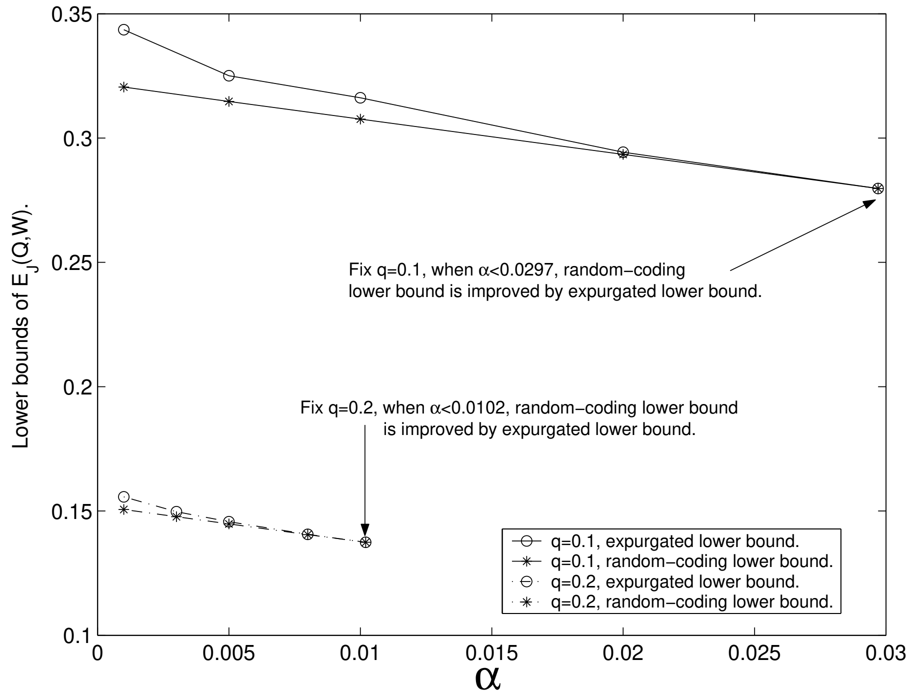

Figure 4.13: Improved expurgated lower bound for the binary DMS and BEC system.

Theorem 4.3 *Given a DMS* $\{Q : \mathcal{S}\}$ *and a DMC* $\{W : \mathcal{X} \to \mathcal{Y}\}$ *such that* $H(Q) < C$, *the joint-source channel coding reliability function satisfies*

$$E_J(Q, W) \leq \max_{\rho \geq 0}[E_o(\rho) - E_s(\rho)], \tag{4.75}$$

where equality holds if $\max_{\rho \geq 0}[E_o(\rho) - E_s(\rho)]$ *is attained at* $\rho \in [0, 1]$.

Note that if $R_\infty = \lim_{\rho \to \infty} \frac{E_o(\rho)}{\rho} < \log|\mathcal{S}|$, then the upper bound in Theorem 4.3 is finite; otherwise, it is infinite. Also, by Proposition 4.1, $H(Q) < C$ and $R_\infty < \log|\mathcal{S}|$ guarantee that $E(\rho) \triangleq E_o(\rho) - E_s(\rho)$ has a global maximum at a finite positive ρ.

Proposition 4.3 *Denoting* $\widehat{\rho} \triangleq \arg\max_{\rho \geq 0} E(\rho)$, *then the sphere-packing upper bound of* $E_J(Q, W)$ *is attained at rate* $R_m = H\left(\frac{q_i^{\frac{1}{1+\widehat{\rho}}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\widehat{\rho}}}}\right)$.

Corollary 4.3 *Consider a DMS* $\{Q : \mathcal{S}\}$ *and a quasi-symmetric DMC* $\{W : \mathcal{X} \to \mathcal{Y}\}$. W *can be partitioned along its columns into symmetric sub-matrices* $W_1, W_2, ..., W_s$, *each* W_i *having size* $|\mathcal{X}| \times |\mathcal{Y}_i|$. *Denote the transition probabilities in any column of the symmetric sub-matrix* W_i *by* $\left\{p_{i,1}, p_{i,2}, ..., p_{i,|\mathcal{X}|}\right\}$, $i = 1, 2, ..., s$.
If $\frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H\left(\frac{p_{ij}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}}\right)+H(Q) \le \log|\mathcal{X}|$, *then the joint source-channel coding exponent upper bound is given by*

$$E_J(Q,W) \le (1+\rho^*)\log|\mathcal{X}| - \log\left\{\left[\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho^*}}\right)^{1+\rho^*}\right]\left(\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho^*}}\right)^{1+\rho^*}\right\}, \tag{4.76}$$

where ρ^* *is the root of equation*

$$\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}H\left(\frac{p_{ij}^{\frac{1}{1+\rho}}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}}\right)}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}} + H\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{|\mathcal{S}|}q_i^{\frac{1}{1+\rho}}}\right) = \log|\mathcal{X}|.$$

Equality holds in (4.76) if and only if $0 \le \rho^* \le 1$.

Note that for $\frac{1}{|\mathcal{X}|}\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}\right)H\left(\frac{p_{ij}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}}\right) + H(Q) \ge \log|\mathcal{X}|$, $E_J(Q,W) = 0$.
Remark: For the quasi-symmetric DMC,

$$\begin{aligned} R_\infty &= \lim_{\rho\to\infty}\frac{(1+\rho)\log|\mathcal{X}| - \log\left\{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\}}{\rho} \\ &= \log|\mathcal{X}| - \lim_{\rho\to\infty}\frac{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}H\left(\frac{p_{ij}^{\frac{1}{1+\rho}}}{\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}}\right)}{\sum_{i=1}^{s}|\mathcal{Y}_i|\left(\sum_{j=1}^{|\mathcal{X}|}p_{ij}^{\frac{1}{1+\rho}}\right)^{1+\rho}} \\ &= 0. \end{aligned}$$

Example 4.5 (*Binary DMS over BSC*) *For a binary DMS and a BSC with error transition probability* ε, *the joint source-channel exponent upper bound is*

$$E_J(Q,W) \le \rho^* - (1+\rho^*)\log\left[\left(\varepsilon^{\frac{1}{1+\rho^*}} + (1-\varepsilon)^{\frac{1}{1+\rho^*}}\right)\left(q^{\frac{1}{1+\rho^*}} + (1-q)^{\frac{1}{1+\rho^*}}\right)\right],$$

if $h_b(q) + h_b(\varepsilon) \le 1$, where ρ^* is the root of

$$h_b\left(\frac{\varepsilon^{\frac{1}{1+\rho}}}{\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}}\right) + h_b\left(\frac{q^{\frac{1}{1+\rho}}}{q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}}\right) = 1.$$

In Figure 4.14, we show the above upper bound and compare it to the lower bound of (4.35).

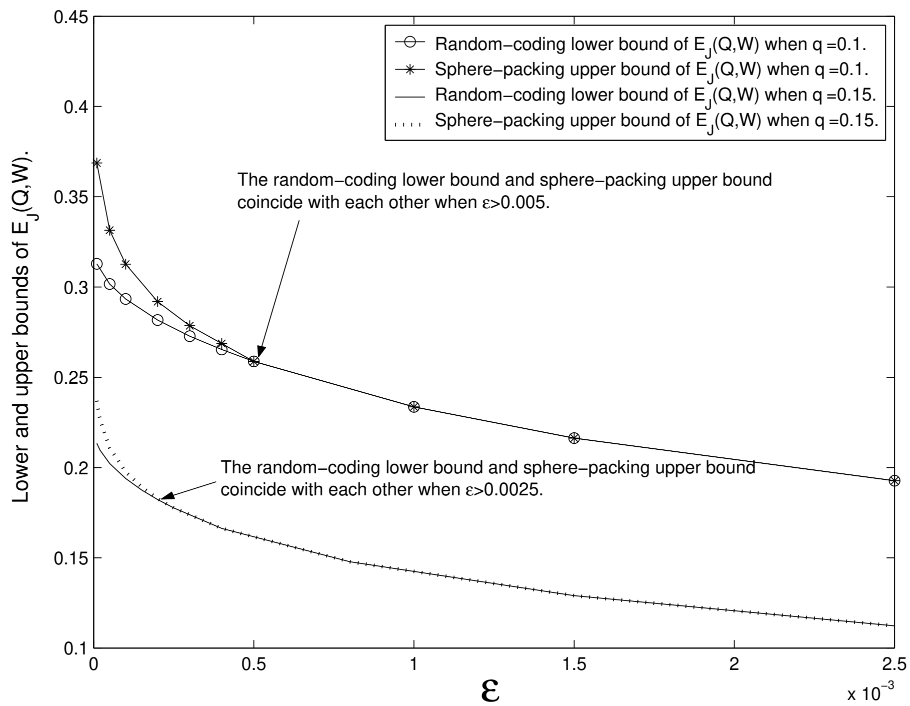

Figure 4.14: The random-coding lower bound and the sphere-packing upper bound for the joint source-channel reliability function for the binary DMS and BSC system.

Example 4.6 (*DMS over N-ary SSC*) *For a M-ary DMS and N-ary SSC with error parameter ε, the joint source-channel exponent upper bound is given by*

$$E_J(Q,W) \le \rho^* \log N - (1+\rho^*) \log \left[\left((1-\varepsilon)^{\frac{1}{1+\rho^*}} + (N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho^*}}\right)\left(\sum_{i=1}^{M} q_i^{\frac{1}{1+\rho^*}}\right)\right]$$

if $h_b(\varepsilon) + \varepsilon \log(N-1) + h_b(Q) \leq \log N$, *where* ρ^* *is the root of*

$$h_b\left(\frac{(N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho}}}{(1-\varepsilon)^{\frac{1}{1+\rho}} + (N-1)(\frac{\varepsilon}{N-1})^{\frac{1}{1+\rho}}}\right) + h_b\left(\frac{q_i^{\frac{1}{1+\rho}}}{\sum_{i=1}^{M} q_i^{\frac{1}{1+\rho}}}\right) = \log N.$$

Example 4.7 *For any binary DMS and BEC pairs, the joint source-channel exponent upper bound is given by*

$$E_J(Q,W) \leq \rho^* - \log[(1-\alpha) + \alpha 2^{\rho^*}] - (1+\rho^*)\log[q^{\frac{1}{1+\rho^*}} + (1-q)^{\frac{1}{1+\rho^*}}] \tag{4.77}$$

where ρ^* *is the root of*

$$h_b\left(\frac{q^{\frac{1}{1+\rho}}}{q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}}\right) = \frac{1-\alpha}{(1-\alpha) + \alpha 2^{\rho}}. \tag{4.78}$$

This upper bound is shown in Figure 4.15 along with the lower bound of (4.35).

4.4 Discussion

4.4.1 Equivalence of Csiszár's and Gallager's Random-Coding Lower Bounds

Definition 4.5 (*Gallager Exponent* $E_G(Q,W)$) *The Gallager joint source-channel exponent for a given DMS with generic distribution* Q *and finite alphabet* $\mathcal{S}$ *over a DMC with transition probability* $W \triangleq P_{Y|X}$ *input alphabet* $\mathcal{X}$ *and output alphabet* $\mathcal{Y}$ *is defined by*

$$E_G(Q,W) = \max_{0 \leq \rho \leq 1} \left[E_o(\rho) - E_s(\rho)\right], \tag{4.79}$$

where $E_o(\rho)$ *and* $E_s(\rho)$ *are Gallager's channel function and source function.*

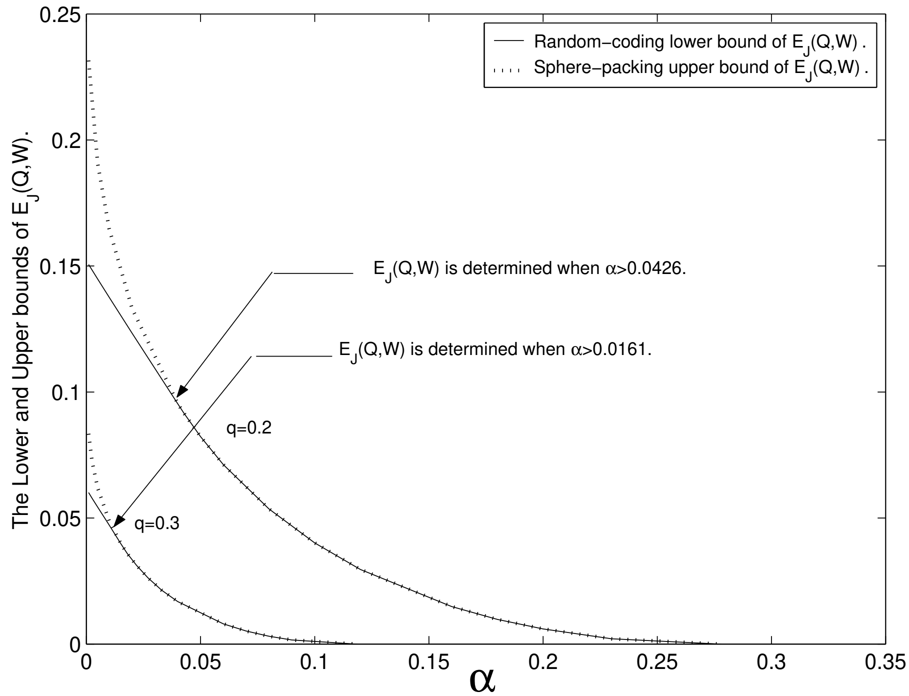

Figure 4.15: The random-coding lower bound and the sphere-packing upper bound for the joint source-channel reliability function for the binary DMS and BEC system.

Theorem 4.4 [*4, Problem* 5.16] *For given DMS with input alphabet S and generic distribution Q and DMC W with generic transition distribution $W \triangleq P_{Y|X}$,*

$$E_J(Q,W) \geq E_G(Q,W). \tag{4.80}$$

In [1], Csiszár states that Gallager's joint source-channel coding lower bound can be larger than the tandem exponent (defined in (4.81)). However, he does not seem to be aware that Gallager's bound is indeed equal to $\min_R[e(R) + E_r(R)]$. This is indeed the case since the two lower bounds in Theorem 4.4 and in Theorem 4.1 are surprisingly identical. Note however that the technique employed by Csiszár to prove Theorem 3.1 in [1] is quite different from Gallager's method to prove Theorem 4.4. Furthermore, Csiszár did not only give the lower bound, but he also established an upper bound. With this two bounds, we can hence determine the exponent exactly for a large class of source-channel pairs (see Figures 4.2, 4.4 and 4.7).

4.4.2 Comparison with the Tandem Source-Channel Exponent

By concatenating optimal source and channel codes, one can achieve exponential convergence of the error probability to 0 with exponent

$$E_T(Q,W) \triangleq \max_R \min \{e(R,Q), E(R,W)\}, \tag{4.81}$$

where $E_T(Q,W)$ is called the tandem source-channel reliability function. We note that $E_T(Q,W)$ is positive if and only if $H(Q) < C$. Otherwise, $E_T(Q,W)$ is zero. Since $e(R,Q)$ is increasing in R and $E(R,W)$ is decreasing in R, and noting that

$$\begin{aligned}
\max_R e(R,Q) &= \max_{H(Q)\le R\le \log|\mathcal{S}|} \min_{P:H(P)=R} D(P \parallel Q) \\
&= \min_{P:H(P)=\log|\mathcal{S}|} D(P \parallel Q) \\
&= \sum_{i=1}^{|\mathcal{S}|} \frac{1}{|\mathcal{S}|} \log \frac{\frac{1}{|\mathcal{S}|}}{q_i} \\
&= -\log(|\mathcal{S}|\overline{q}),
\end{aligned}$$

where $\overline{q}$ is the geometric mean of the source probabilities, i.e. $\overline{q} \triangleq \left(\prod_{i=1}^{|\mathcal{S}|} q_i\right)^{\frac{1}{|\mathcal{S}|}} \le \frac{1}{|\mathcal{S}|}$, if $-\log(|\mathcal{S}|\overline{q}) \ge E(\log|\mathcal{S}|, W)$, then $e(R,Q)$ and $E(R,W)$ has an intersection R_o and

$$E_T(Q,W) = e(R_o,Q) = E(R_o,W).$$

Otherwise, according to (4.81),

$$E_T(Q,W) = \max_R \min\{e(R,Q), E(R,W)\} = \max_R e(R,Q) = -\log(|\mathcal{S}|\overline{q})$$

is determined only by the source. If the intersection R_o exists and $R_o \geq R_{cr}$, the tandem exponent can be determined exactly, if $R_o < R_{cr}$, we can bound the tandem exponent by

$$\max_R \min\left\{e(R,Q), E_L(R,W)\right\} \leq E_T(Q,W) \leq \max_R \min\left\{e(R,Q), E_U(R,W)\right\}, \quad (4.82)$$

where $E_L(R,W)$ is a lower bound and $E_U(R,W)$ is an upper bound for the channel reliability function $E(R,W)$. Similarly, if we use $E_r(R,W)$ as $E_L(R,W)$, then we call it the "random-coding lower bound" to $E_T(Q,W)$; if we use $E_{sp}(R,W)$ as $E_U(R,W)$, then we call it the "sphere-packing upper bound" to $E_T(Q,W)$.

In general, we know that $E_J(Q,W) \geq E_T(Q,W)$ since tandem coding is a special case of joint source-channel coding. We are hence interested in determining the condition for which $E_J(Q,W) > E_T(Q,W)$. Note that if $e(R,Q)$ and the random-coding lower bound of $E(R,W)$ have no intersection, i.e., $-\log(|\mathcal{S}|\overline{q}) < E_r(\log|\mathcal{S}|, W)$, then $E_J(Q,W)$ is strictly bigger than $E_T(Q,W)$ since for any $H(Q) \leq R \leq \min\{C, \log|\mathcal{S}|\}$,

$$E_J(Q,W) \geq E_r(R,W) + e(R,Q) \geq E_r(R,W) > -\log(|\mathcal{S}|\overline{q}) = E_T(Q,W).$$

So in the rest of this section, we assume that $-\log(|\mathcal{S}|\overline{q}) \geq E_r(\log|\mathcal{S}|, W)$ and $H(Q) < C$.

Proposition 4.4 *If* $\max\left(H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right), E_o(1) - e(R_{cr},Q)\right) \geq R_{cr}$*, then* $E_J(Q,W) > E_T(Q,W)$*, where* R_{cr} *is the critical rate of the channel.*

Proof: By Proposition 4.1, if the DMS and the DMC satisfy

$$H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R_{cr},$$

the random-coding lower bound of $E_J(Q,W)$ must be attained at an R_m such that $R_{cr} \leq R_m$. Note that since we assumed that $H(Q) < C$, then

$$H(Q) < R_m < \min\{C, \log|\mathcal{S}|\},$$

since $R_m = H(Q)$ or $R_m = C$ if and only if $H(Q) = C$. Accordingly, the joint source-channel exponent is exactly determined and we have

$$\begin{aligned} E_J(Q,W) &= e(R_m,Q) + E_{sp}(R_m,W) \\ &> \max_R \min\{e(R,Q), E_{sp}(R,W)\} \quad \text{(since both terms are positive)} \\ &\geq E_T(Q,W). \quad \text{(by (4.82))} \end{aligned}$$

Similarly, if the intersection rate R_o of $e(R,Q)$ and $E(R,W)$ satisfies $R_o \geq R_{cr}$, then the tandem source-channel exponent is exactly determined and we have

$$\begin{aligned} E_T(Q,W) &= E_r(R_o,W) = e(R_o,Q) \\ &\leq \max(e(R_m,Q), E_r(R_m,W)) \\ &< e(R_m,Q) + E_r(R_m,W) \quad \text{(since both terms are positive)} \\ &= E_J(Q,W). \end{aligned}$$

Finally, we complete the proof by observing that

$$E_r(R_{cr},W) = E_o(1) - R_{cr} \geq e(R_{cr},Q)$$

is an equivalent condition for $R_o \geq R_{cr}$. □

From Figures 4.2, 4.4, and 4.7, we note that if the source and channel pair is located in region B, then the corresponding joint source-channel exponent is bigger than the tandem exponent because $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}$ holds. We next give some numerical examples in order to illustrate the extent to which $E_J(Q,W)$ can be bigger than $E_T(Q,W)$, when the condition $\max\left(H(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}), E_o(1) - e(R_{cr},Q)\right) \geq R_{cr}$ is satisfied.

Example 4.8 *In Figure 4.16, we compare the joint source-channel exponent with the tandem exponent for binary DMS* $\{q, 1-q\}$ *and BSC pairs. When* $q = 0.1$, $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R_{cr}$ *is satisfied for* $\varepsilon \geq 0.0005$, *while* $E_o(1) - e(R_{cr},Q) \geq R_{cr}$ *is* *satisfied for $\varepsilon \geq 0.0012$, which implies that $E_J(Q,W) > E_T(Q,W)$ if $\varepsilon \geq 0.0005$. We only show the two exponents for $\varepsilon \geq 0.0012$, since $E_T(Q,W)$ is determined exactly for $\varepsilon \geq 0.0012$. (Note that for $0.0005 < \varepsilon < 0.0012$, we still can obtain the lower and upper bounds of $E_T(Q,W)$.) When $q = 0.2$, both $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R_{cr}$ and $E_o(1) - e(R_{cr}, Q) > R_{cr}$ are satisfied for $\varepsilon > 0.0001$, so $E_J(Q,W) > E_T(Q,W)$ for $\varepsilon \geq 0.0001$.*

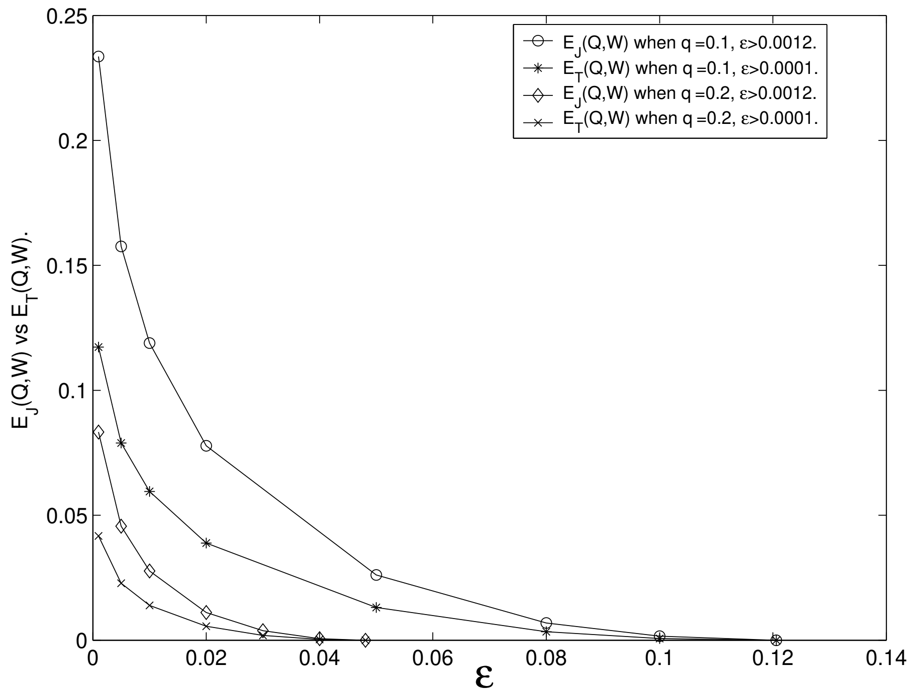

Figure 4.16: Joint source-channel exponent vs Tandem exponent for binary DMS $\{q, 1-q\}$ and BSC with parameter ε.

Example 4.9 *In Figure 4.17, we compare the joint source-channel exponent with the tandem exponent for binary DMS $\{q, 1-q\}$ and BEC pairs. When $q = 0.2$, $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R_{cr}$ is satisfied for $\alpha \geq 0.0426$, while $E_o(1) - e(R_{cr}, Q) \geq R_{cr}$ is satisfied for $\alpha \geq 0.0702$, which implies that $E_J(Q,W) > E_T(Q,W)$ if $\alpha \geq 0.0426$.*

When $q = 0.3$, $H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \geq R_{cr}$ is satisfied for $\alpha \geq 0.0161$, while $E_o(1) -$

$e(R_{cr}, Q) \geq R_{cr}$ is satisfied for $\alpha \geq 0.0272$. Thus $E_J(Q,W) > E_T(Q,W)$ if $\alpha \geq 0.0161$. Note that for $\alpha < 0.0161$, we are not sure if $E_J(Q,W)$ is strictly bigger than $E_T(Q,W)$; in other words, the tandem exponent can be equal to the joint source-channel exponent.

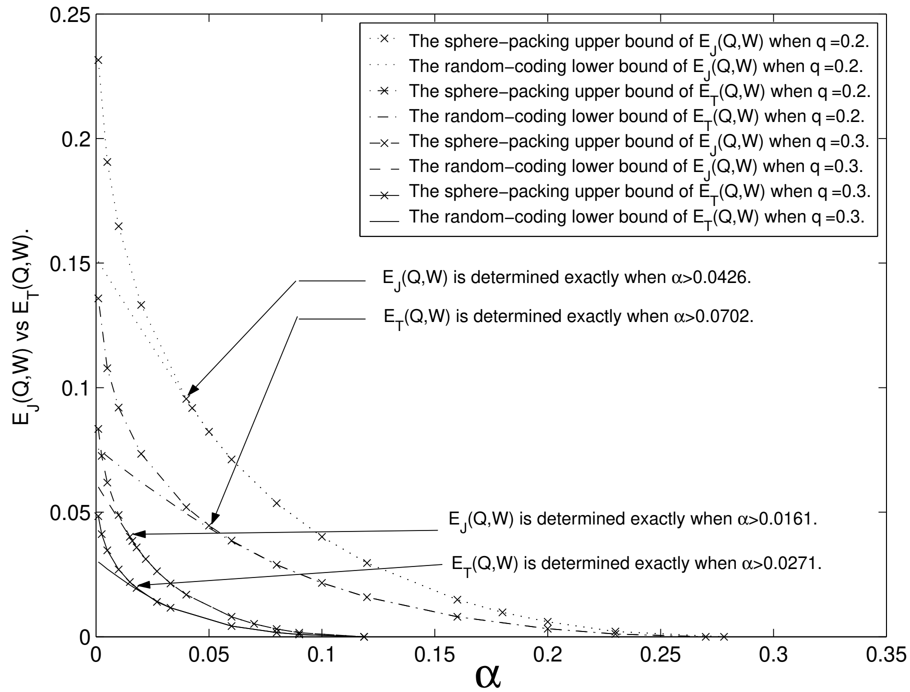

Figure 4.17: Joint source-channel exponent vs tandem exponent for binary DMS $\{q, 1-q\}$ and BEC with parameter α.

From Figures 4.16 and 4.17, we remark that the the joint source-channel exponent can be substantially larger than the tandem exponent.

Now consider a DMS and DMC pair such that

$$\max\left(H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right), E_o(1) - e(R_{cr}, Q)\right) < R_{cr},$$

then the joint source channel exponent is attained at R_m such that $R_m < R_{cr}$, and the intersection of $e(R,Q)$ and $E(R,W)$ occurs at a rate less than R_{cr}. In this case, neither

$E_J(Q,W)$ nor $E_T(Q,W)$ are exactly known. If the lower bound of $E_J(Q,W)$ is strictly bigger than the upper bound of $E_T(Q,W)$, then we must have $E_J(Q,W) > E_T(Q,W)$. Hence we obtain the following proposition.

Proposition 4.5 *If there exists an R_1 such that $e(R_1,Q) = E_{sp}(R_1,W)$, then when $E_o(1) - E_s(1) > e(R_1,Q)$, we have $E_J(Q,W) > E_T(Q,W)$; if such R_1 does not exist, then when $E_o(1) - E_s(1) > -\log(|\mathcal{S}|\overline{q})$, we also have $E_J(Q,W) > E_T(Q,W)$.*

Proof: By Theorem 4.1, it follows that

$$E_J(Q,W) \geq E_o(1) - E_s(1).$$

Now if $e(R,Q)$ and $E_{sp}(R,W)$ has an intersection at R_1, then $E_{sp}(R_1,W) = e(R_1,Q) \geq e(R_o,Q) = E_T(Q,W)$ since R_1 must be no less than R_o. If such R_1 does not exist, which means $E_{sp}(R,W) > e(R,Q)$ for all R, then according to our assumption,

$$E_o(1) - E_s(1) > -\log(|\mathcal{S}|\overline{q}) = \max_R e(R,Q) \geq E_T(Q,W).$$

$\square$

Proposition 4.5 is a sufficient condition to judge whether the joint source-channel coding exponent outperforms the tandem coding exponent if both the $E_J(Q,W)$ and $E_T(Q,W)$ cannot been determined exactly. But as we know, the sphere-packing upper bound for the channel reliability function is not tight for small rates. Thus some DMS and DMC pairs may not satisfy the condition $E_o(1) - E_s(1) > e(R_1,Q)$, even when the corresponding $E_J(Q,W)$ is much bigger than $E_T(Q,W)$. Therefore, we can replace the sphere-packing upper bound by the straight-line bound, which is much tighter when R is small, for the DMC for which $E_{ex}(0,W) < \infty$.

Definition 4.6 *[4, P.160] (Straight-Line exponent $E_{sl}(R,W)$) For any DMC such that $E_{ex}(0,W) < \infty$, $E_{sl}(R,W)$ is the smallest linear function of R being tangent to the curve $E_{sp}(R,W)$ and also satisfies $E_{sl}(0,W) = E_{ex}(0,W)$.*

Theorem 4.5 *[4] (Straight-Line Upper Bound) For a DMC such that $E_{ex}(0, W) < \infty$, let R_s be the R for which $E_{sl}(R, W) = E_{sp}(R, W)$, then*

$$E(R, W) \leq E_{sl}(R, W) \tag{4.83}$$

for $0 < R < R_s$.

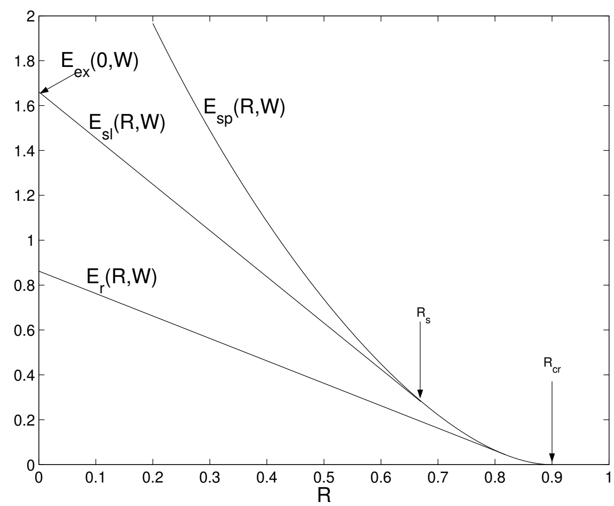

Figure 4.18: Straight-line upper bound.

Proposition 4.6 *For any DMC such that $E_{ex}(0, W) < \infty$, if there exists an intersection R_1 such that $e(R_1, Q) = E_{sl}(R_1, W)$, then when $E_o(1) - E_s(1) > e(R_1, Q)$, we have $E_J(Q, W) > E_T(Q, W)$.*

Remark: To verify whether $E_J(Q, W) > E_T(Q, W)$ for a DMS and DMC pair, we normally employ Proposition 4.4 first. If the condition is not satisfied, we use Proposition 4.6. If there is no intersection between $E_{sl}(R, W)$ and $e(R, Q)$ or $E_{ex}(0, W) = \infty$, we then consider Proposition 4.5.

In practice, for a DMS and an arbitrary DMC (not symmetric), it is not easy to determine the intersection of $e(R, Q)$ and $E_{sp}(R, W)$ or $E_{sl}(R, W)$, because the computation of $E_{sp}(R, W)$ and R_s is hard. But we can compute $E_o(\rho)$ and $E_r(R, W)$ easily with Arimoto's algorithm [10]. We next introduce some simpler but weaker sufficient conditions.

Proposition 4.7 *If $E_o(1) - E_s(1) > e(R_{cr}, Q)$, then $E_J(Q,W) > E_T(Q,W)$.*

Proof: If $R_o \geq R_{cr}$, then Proposition 4.4 is satisfied. If $R_o < R_{cr}$, then $e(R_{cr}, Q) > e(R_1, Q)$, and Proposition 4.5 is satisfied, where R_1 is the intersection of $e(R,Q)$ and $E_{sp}(R,W)$. $\square$

For a DMC such that $E_{ex}(0,W) < \infty$, we may denote the upper bound of the channel reliability function as

$$E_s(R,W) = \begin{cases} E_{sl}(R,W), & 0 \leq R \leq R_s, \\ E_{sp}(R,W), & R_s \leq R \leq C. \end{cases}$$

It is easily seen that $E_s(R,W)$ is also convex in R. Now connect $(0, E_{ex}(0,W))$ and $(R_{cr}, E_{ex}(R_{cr}, W))$ with a line, denoted by l_1, and connect $(R_m, e(R_m, Q))$ and $(\log|\mathcal{S}|, e(\log|\mathcal{S}|, Q))$ with a line, denoted by l_2. Suppose the intersection of the two curves $E_s(R,W)$ and $e(R,Q)$ is $(R_1, e(R_1,Q))$, and the intersection of l_1 and l_2 is (R_l, E_{R_l}). Because of the convexity of $e(R,Q)$ and $E_s(R,W)$, E_{R_l} must be bigger than $e(R_1,Q)$. Thus the condition $E_o(1) - E_s(1) > E_{R_l}$ ensures that Proposition 4.5 and 4.6 hold.

Proposition 4.8 *For a DMS and a DMC for which $E_{ex}(0,W) < \infty$, if*

$$E_o(1) - E_s(1) \geq E_{R_l} \triangleq \frac{k_2 \log|\mathcal{S}| \left(k_1 + \frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|} \log q_i\right) + k_1 E_{ex}(0,W)}{k_1 - k_2},$$

then $E_J(Q,W) > E_T(Q,W)$, where

$$k_1 = \frac{D\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}} \parallel q_i\right) + \log\left(|\mathcal{S}|\left(\prod_{i=1}^{|\mathcal{S}|} q_i\right)^{\frac{1}{|\mathcal{S}|}}\right)}{H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) - \log|\mathcal{S}|}$$

and

$$k_2 = \frac{E_o(1) - R_{cr} - E_{ex}(0,W)}{R_{cr}}.$$

Remark: We note that even though Proposition 4.8 is weaker than Propositions 4.5 and 4.6, the computation is much simpler since we only need to know $E_o(1)$, $E_{ex}(0, W)$ and R_{cr}.

Example 4.10 *Consider the binary DMS* $\{q, 1-q\}$ *and the BEC with parameter* α, *then*

$$E_o(1) - E_s(1) = 1 - \log(1+\alpha) - 2\log(\sqrt{q} + \sqrt{1-q}),$$

and

$$E_{ex}(0, W) = -\frac{\log(\alpha)}{2}.$$

Note that $\log|\mathcal{S}| = 1 \geq C$, *which guarantees that* $E_s(R, W)$ *and* $e(R, Q)$ *have an intersection. Thus if any one of the conditions*

$$\begin{array}{ll} (1) & H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}, \\ (2) & E_o(1) - e(R_{cr}, Q) > R_{cr}, \\ (3) & E_o(1) - E_s(1) > e(R_{cr}, Q), \\ (4) & E_o(1) - E_s(1) \geq E_{R_l} = \dfrac{k_2 \log|\mathcal{S}|\left(k_1 + \frac{1}{|\mathcal{S}|}\sum_{i=1}^{|\mathcal{S}|}\log q_i\right) + k_1 E_{ex}(0, W)}{k_1 - k_2}, \end{array}$$

is satisfied, then the joint source-channel exponent strictly outperforms the tandem exponent. The above conditions are illustrated in Figure 4.19, where

$$\begin{aligned} A &\triangleq \left\{(\alpha, q) : H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) > R_{cr}\right\}, \\ B &\triangleq \{(\alpha, q) : E_o(1) - e(R_{cr}, Q) > R_{cr}\} \\ C &\triangleq \{(\alpha, q) : E_o(1) - E_s(1) > e(R_{cr}, Q)\} \\ D &\triangleq \{(\alpha, q) : E_o(1) - E_s(1) > E_{R_l}\}, \\ E &\triangleq \{(\alpha, q) : H(Q) < C\} \\ F &\triangleq (A \cup B \cup C \cup D) \cap E = \{(\alpha, q) : E_J(Q, W) > E_T(Q, W)\} \\ G &\triangleq \overline{E} = \{(\alpha, q) : E_J(Q, W) = E_T(Q, W) = 0\} \end{aligned}$$

We observe that $E_J(Q,W) > E_T(Q,W)$ for a large (α, q) pairs, see Region F. Note that when (α, q) are very small (such as in Region H), we are not sure whether the joint source-channel exponent is still strictly bigger than the tandem exponent. Some numerical examples in Figure 4.20 show that the sphere-packing upper bound of $E_T(Q,W)$ is bigger than the random-coding lower bound of $E_J(Q,W)$ for $q = 0.1$ and $q = 0.2$. In fact, as we will argue later, in Region H there exist some (α, q) pairs, for which the random-coding lower bound of $E_J(Q,W)$ is smaller than the upper bound of $E_T(Q,W)$. (Actually we show a stronger result: the tandem exponent $E_T(Q,W)$ can be bigger than the lower bound of $E_J(Q,W)$.)

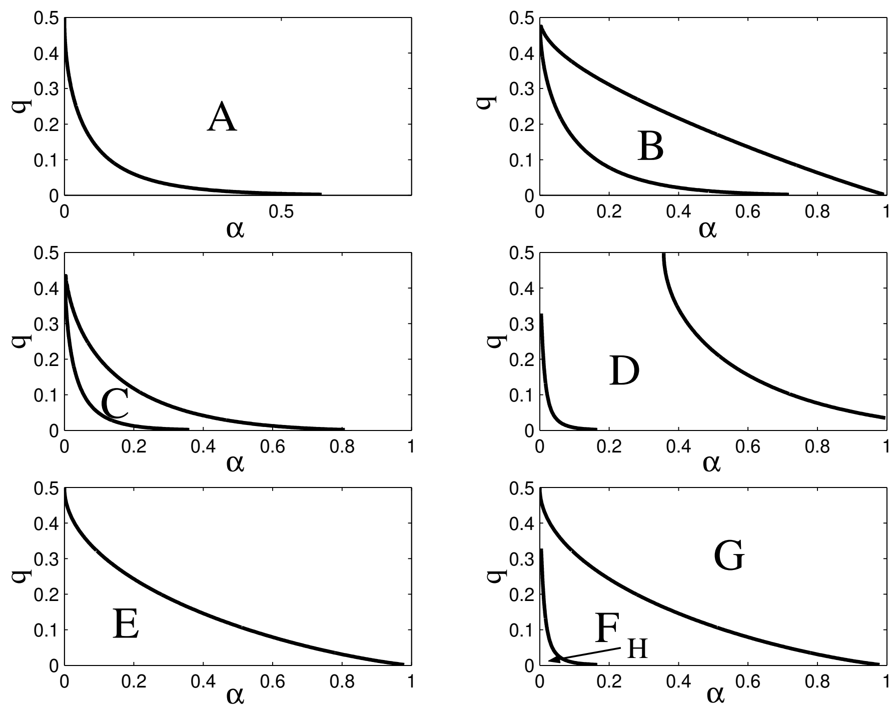

Figure 4.19: The regions for the (α,q) pairs in the binary DMS and BEC systems.

In Figure 4.19 we note that $E_o(1) - E_s(1) \geq E_{R_l}$ is satisfied in a fairly large region (see Region D), and we also know that the condition $E_o(1) - E_s(1) \geq e(R_1, Q)$ would involve more (α, q) pairs, where R_1 is the intersection of $E_s(R,W)$ and $e(R,Q)$. One

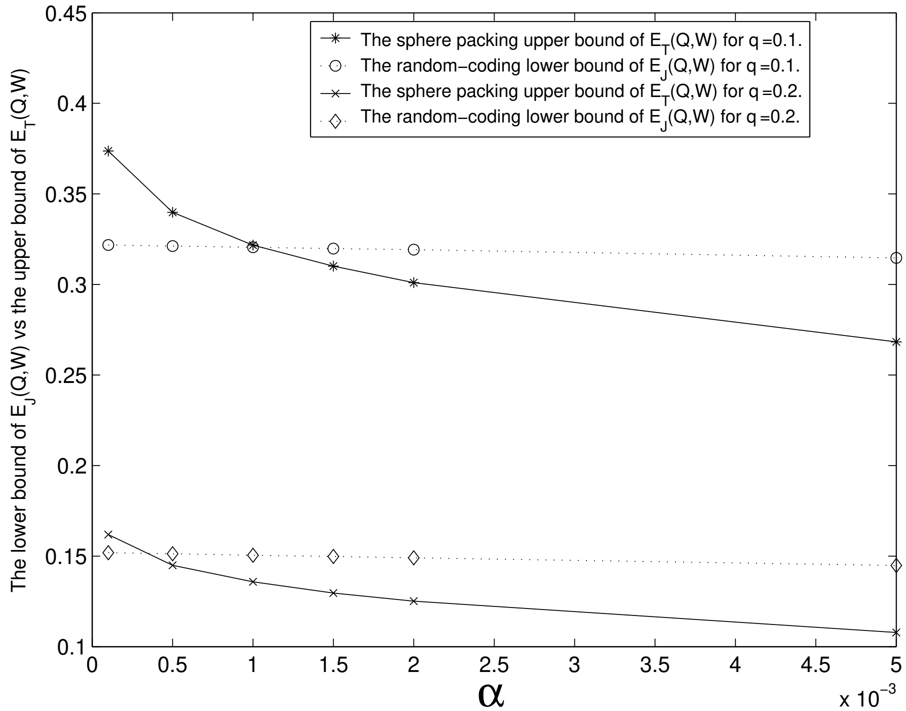

Figure 4.20: The sphere-packing upper bound of tandem exponent can be bigger than the random-coding lower bound of joint source-channel exponent.

may ask whether the lower bound of $E_J(Q,W)$ is strictly bigger than the upper bound of $E_T(Q,W)$ for all source and channel pairs, where the upper bound can be the sphere-packing upper bound, straight-line upper bound or any other known upper bounds. By Theorem 3.1, $E_J(Q,W)$ can be bounded by

$$\min_R[e(R,Q)+E_r(R,W)] \leq E_J(Q,W) \leq \min_R[e(R,Q)+E(R,W)], \tag{4.84}$$

where $H(Q) \leq R \leq \min(\log|\mathcal{S}|, C)$ and $E(R,W)$ is the channel reliability function. Unfortunately, according to (4.82) and (4.84), we cannot show that $E_J(Q,W)$ is strictly bigger than $E_T(Q,W)$ for all source and channel pairs. Conversely, we note that there exist some source-channel pairs such that the corresponding tandem exponent can be bigger than the lower bound of the joint source-channel exponent, only if $E(R,W) \neq$

$E_r(R, W)$. To show this, we can choose a sequence of distributions $Q = \{q_1, ..., q_{|\mathcal{S}|}\}$ such that $D\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}} \parallel Q\right) \to 0^+$, and $-\log(|\mathcal{S}|\overline{q}) \to \infty$, and hence the random-coding lower bound of $E_J(Q, W)$ is attained at $R_m = H\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}}\right) \to R_o^-$, where R_o is the intersection of $e(R, Q)$ and $E(R, W)$. It follows that

$$E(R_o, W) - E_r\left(R_m, W\right) > e(R_m, Q) = D\left(\frac{\sqrt{q_i}}{\sum_{i=1}^{|\mathcal{S}|}\sqrt{q_i}} \parallel Q\right) \to 0^+,$$

which implies,

$$E_T(Q, W) = E(R_o, W) > E_r\left(R_m, W\right) + e\left(R_m, Q\right),$$

which is the random-coding lower bound to $E_J(Q, W)$. One possible choice for the distribution Q is

$$q_1 \to 0, \quad \text{and} \quad q_i \to \frac{1}{|\mathcal{S}| - 1}, i \neq 1.$$

4.5 Lossy Joint Source-Channel Exponent with Hamming Distortion Measure

In Section 2.1.2 we noted that in general the source reliability function $F(R, Q, \Delta)$ under a distortion threshold Δ is not continuous in R and little is known about the function $F(R, Q, \Delta)$. Thus it is very hard to analytically compute the joint source-channel exponent by using $F(R, Q, \Delta)$. Here we only consider the simplest case: the joint source-channel exponent for a binary DMS and a DMC under the Hamming distortion measure $d_H(\cdot, \cdot)$, given by

$$d_H(s, \widetilde{s}) = \begin{cases} 1, & \text{if } s \neq \widetilde{s}, \\ 0, & \text{if } s = \widetilde{s}, \end{cases} \tag{4.85}$$

Lemma 4.7 *[7, P.342] The rate-distortion function $R(Q,\Delta)$ for a binary DMS $Q = \{q, 1-q\}$ under the Hamming distortion measure is given by*

$$R(Q,\Delta) = \begin{cases} h_b(q) - h_b(\Delta), & 0 \le \Delta \le \min(q, 1-q), \\ 0, & \Delta > \min(q, 1-q). \end{cases} \tag{4.86}$$

Remark: From Lemma 4.7, we see that if $\Delta > \frac{1}{2}$, then $R(Q,\Delta) = 0$.

Lemma 4.8 *For binary DMS $\{q, 1-q\}$ $(q < \frac{1}{2})$ under the Hamming distortion measure and distortion threshold Δ such that $\Delta \le \frac{1}{2}$,*

1) If $0 \le \Delta \le q$, the source reliability function

$$F(R,Q,\Delta) = \begin{cases} \min_{p:h_b(p)=R+h_b(\Delta)} \widetilde{D}(p \parallel q), & h_b(q) - h_b(\Delta) \le R \le 1 - h_b(\Delta), \\ 0, & 0 < R \le h_b(q) - h_b(\Delta), \end{cases} \tag{4.87}$$

is continuous and convex in R;

2) If $q < \Delta$, the source reliability function

$$F(R,Q,\Delta) = \min_{p:h_b(p)=R+h_b(\Delta)} \widetilde{D}(p \parallel q), \quad 0 < R \le 1 - h_b(\Delta) \tag{4.88}$$

is continuous and convex in R, where $\widetilde{D}(p \parallel q) \triangleq p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$.

Proof: The proof of continuity and convexity of $F(R,Q,\Delta)$ is the same as for Theorem 2.2. Since $F(R,Q,\Delta)$ is increasing in R (by Corollary 2.3), it follows that $F(R,Q,\Delta)$ is strictly increasing and convex in $[h_b(q) - h_b(\Delta), 1 - h_b(\Delta)]$ if $0 \le \Delta \le q$, and $(0, 1 - h_b(\Delta)]$ if $q < \Delta$. Thus

$$F(R,Q,\Delta) = \min_{P:R(Q,\Delta) \ge R} D(P \parallel Q) = \min_{P:R(Q,\Delta)=R} D(P \parallel Q).$$

Substituting (4.86) into the right hand side, yields (4.87) and (4.88). $\square$

Since we have obtained the expression of $F(R,Q,\Delta)$, we can apply Theorem 3.3

to compute the joint source-channel exponent lower and upper bounds for the binary DMS $\{q, 1-q\}$ and any DMC systems, under the Hamming distortion measure. We remark that the source reliability function $F(R, Q, \Delta)$ is very similar to $e(R, Q)$. The only difference is that the constraint is $h_b(p) = R + c$ here ($h_b(\Delta)$ can be seen as a constant c), instead of $h_b(p) = R$. We observe that

$$\begin{aligned}
\inf_R [E(R, W) + F(R, Q, \Delta)] &= \min_R [E(R, W) + \min_{h_b(p)=R+c} D(p \parallel q)] \qquad (4.89)\\
&= \min_R [E(R, W) + e(R + c, Q)]\\
&= \min_{R'} [E(R' - c, W) + e(R', Q)]\\
&= \min_{R'} \left[E(R', W) + c\rho' + e(R', Q)\right],
\end{aligned}$$

where $E(R, W)$ might be $E_r(R, W)$ or $E_{sp}(R, W)$ and ρ' achieves $E_r(R, W)$ or $E_{sp}(R, W)$, depending on R'. Since function $F(R, Q, \Delta)$ is continuous in $R > 0$, we change the notation 'inf' to 'min' in (4.89). By introducing Lagrange multipliers to find out the stationary point (R', p), it can be shown that $\min_{R=h_b(p)-c}[E(R, W) + D(p \parallel q)]$ will have the parametric expression $[E_o(\rho_c) - E_s(\rho_c) + c\rho_c]$, where ρ_c depends on the constraint $R = h_b(p) - c$. Since the computation is very similar to that of zero distortion case, we skip it and give the following results.

Theorem 4.6 *Given a binary DMS* $\{q, 1-q\}$ ($q \leq \frac{1}{2}$) *and a DMC* $\{W : \mathcal{X} \to \mathcal{Y}\}$ *under the Hamming distortion measure and distortion threshold* Δ ($\Delta \leq \frac{1}{2}$)*, and provided*

$$h_b(q) - h_b(\Delta) \leq C \quad \text{and} \quad R_\infty = \lim_{\rho\to\infty} \frac{E_o(\rho)}{\rho} < 1 - h_b(\Delta),$$

where $C = \lim_{\rho\to 0} \frac{E_o(\rho)}{\rho}$ *is the channel capacity , then the joint source-channel exponent satisfies*

1) If $0 \leq \Delta < \frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}$*, then*

$$E_J^\Delta(Q, W) \geq \max_{0\leq\rho\leq 1} \left[E_o(\rho) - (1+\rho)\log\left(q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}\right) + \rho h_b(\Delta)\right]. \qquad (4.90)$$

Otherwise, if $\Delta \geq \frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}$, *then*

$$E_J^{\Delta}(Q,W) > E_o(1) + \widetilde{D}(\Delta \parallel q). \tag{4.91}$$

2)

$$E_J^{\Delta}(Q,W) \leq \max_{\rho \geq 0} \left[E_o(\rho) - (1+\rho)\log\left(q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}\right) + \rho h_b(\Delta) \right]. \tag{4.92}$$

Remark: If $h_b(q) - h_b(\Delta) \geq C$, then $E_J^{\Delta}(Q,W) = 0$. If $R_\infty \geq 1 - h_b(\Delta)$, then $E_J^{\Delta}(Q,W) = +\infty$.

Proposition 4.9 $E_\Delta(\rho) \triangleq E_o(\rho) - (1+\rho)\log\left(q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}\right) + \rho h_b(\Delta)$ *is strictly concave in* ρ ($\rho \geq 0$) *and has a global maximum at a finite* ρ *iff* $h_b(q) - h_b(\Delta) < C$ *and* $R_\infty < 1 - h_b(\Delta)$. *Denoting* $\rho^* \triangleq \arg\max_{\rho \geq 0} E_\Delta(\rho)$, *then the sphere-packing upper bound to* $E_J^{\Delta}(Q,W)$ *is attained at* $R_m^* = H\left(\frac{q_i^{\frac{1}{1+\rho^*}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho^*}}}\right)$. *Letting* $\rho' = \min(\rho^*, 1)$, *when* $\Delta \leq \frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}$, *the lower bound is attained at rate* $R_m' = H\left(\frac{q_i^{\frac{1}{1+\rho'}}}{\sum_{i=1}^{|\mathcal{S}|} q_i^{\frac{1}{1+\rho'}}}\right)$. *When* $\Delta \geq \frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}$, *the lower bound is attained at* $R \to 0$, *so* $E_J^{\Delta}(Q,W)$ *is strictly bigger than the bound (4.91).*

Example 4.11 (*Binary DMS over BSC with Hamming distortion*) *For a binary DMS* $\{q, 1-q\}$ *and a BSC with error transition probability* ε, *the joint source-channel exponent under the Hamming distortion measure with distortion threshold* Δ ($\Delta < \frac{1}{2}$) *satisfies*

1) If $h_b(\varepsilon) + h_b(q) \leq 1 + h_b(\Delta)$ *and* $h_b\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right) + h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right) \geq 1 + h_b(\Delta)$, *then*

$$E_J^{\Delta}(Q,W) = \rho^* - (1+\rho^*)\log\left[\left(\varepsilon^{\frac{1}{1+\rho^*}} + (1-\varepsilon)^{\frac{1}{1+\rho^*}}\right)\left(q^{\frac{1}{1+\rho^*}} + (1-q)^{\frac{1}{1+\rho^*}}\right)\right] + \rho^* h_b(\Delta), \tag{4.93}$$

where ρ^* *is the root of equation*

$$h_b\left(\frac{\varepsilon^{\frac{1}{1+\rho}}}{\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}}\right) + h_b\left(\frac{q^{\frac{1}{1+\rho}}}{q^{\frac{1}{1+\rho}} + (1-q)^{\frac{1}{1+\rho}}}\right) = 1 + h_b(\Delta).$$

2) If $h_b\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right)+h_b\left(\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\right)<1+h_b(\Delta)$ *and* $\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}>\Delta$, *then*

$$E_J^{\Delta}(Q,W)\geq 1-2\log\left[(\sqrt{\varepsilon}+\sqrt{1-\varepsilon})(\sqrt{q}+\sqrt{1-q})\right]+h_b(\Delta). \tag{4.94}$$

3) If $\frac{\sqrt{q}}{\sqrt{q}+\sqrt{1-q}}\leq\Delta$, *then*

$$E_J^{\Delta}(Q,W)>1-2\log\left(\sqrt{\varepsilon}+\sqrt{1-\varepsilon}\right)+\widetilde{D}(\Delta\parallel q). \tag{4.95}$$

In Figure 4.21, we observe that if the pair (ε,q) *is located in region B, then the*

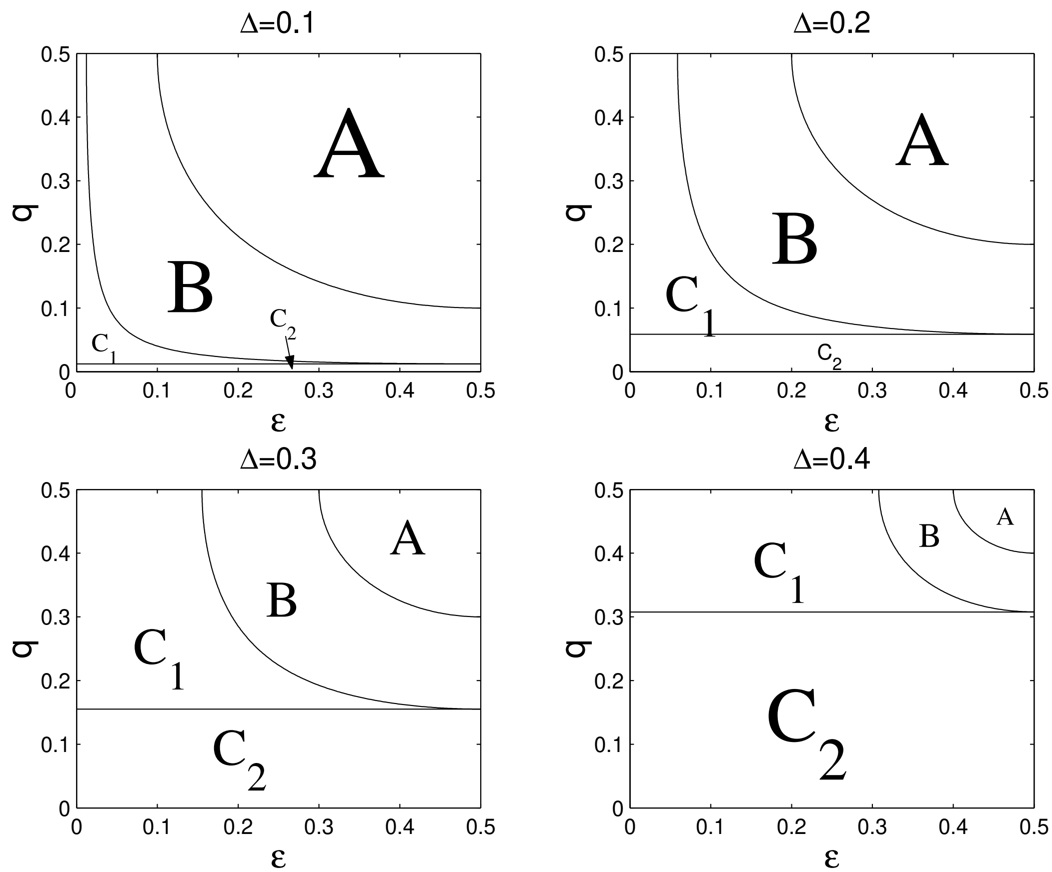

Figure 4.21: The regions for the (ε,q) pairs in the binary DMS and BSC system with Hamming distortion for different values of the distortion threshold Δ.

corresponding joint source-channel exponent can be determined exactly by (4.93). If

(ε, q) is located in region C_1, then (4.94) holds. If (ε, q) is located in region C_2, then (4.95) holds. When $(\varepsilon, q) \in A$, $E_J(Q,W)$ is zero, and the error probability of this communication system will converge to 1 for n sufficiently large. So we are only interested in the cases when $(\varepsilon, q) \in B \cup C_1 \cup C_2$. Figure 4.22 shows the joint source-channel

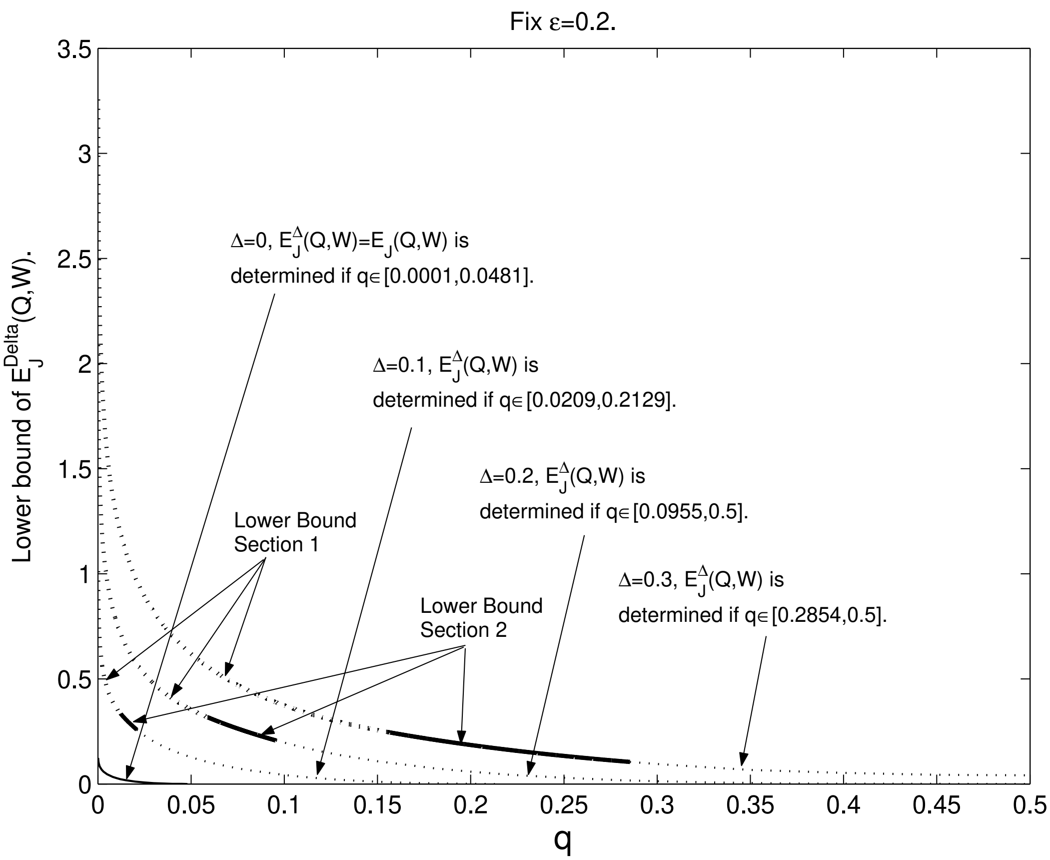

Figure 4.22: The joint source-channel exponent lower bound of the binary DMS $\{q, 1-q\}$ $(q \leq 1)$ and BSC pairs under Hamming distortion.

exponent lower bound of the binary DMS $\{q, 1-q\}$ $(q \leq \frac{1}{2})$ and BSC pairs under different distortion thresholds. We fix the BSC parameter $\varepsilon = 0.2$, and vary q from 0 to 0.5. In Figure 4.22, Lower bound section 1 is determined by (4.95), and Lower bound section 2 is determined by (4.94).

Chapter 5

Summary

Our main results in this project are to establish simple parametric expressions for Csiszár's lower and upper bounds of the joint source-channel exponent:

$$\max_{0\le\rho\le 1} E(\rho) \le E_J(Q,W) \le \max_{0\le\rho} E(\rho), \tag{5.1}$$

where $E(\rho) \triangleq E_o(\rho) - E_s(\rho)$. $E_o(\rho)$ reflects the characteristics of the DMC and $E_s(\rho)$ is determined by the DMS. The practical significance of (5.1) is that the computation of the joint source-channel exponent is actually transformed as the computation of $E_o(\rho)$, since $E_s(\rho)$ is easily obtained. We note that for quasi-symmetric DMC's, $E_o(\rho)$ has a simple form, and thus (5.1) can be expressed via single letter formulas. For arbitrary DMC's (like non-symmetric channels), $E_o(\rho)$ can be obtained indirectly by Arimoto's [10] method. Here we should stress that according to Arimoto's algorithm, $E_o(\rho)$ can be computed if ρ is given. Hence when ρ is limited in an interval, say $[0,1]$, we can search the suitable ρ to maximize $E(\rho)$ with a given fidelity, and thus the lower bound of $E_J(Q,W)$ can be obtained. But for the upper bound, ρ ranges from 0 to infinity. One way is to increase ρ gradually until we find the global maximum point of $E(\rho)$ ($E(\rho)$ is concave and has a global maximum at a finite ρ), but it might take a long time; so it may not be always efficient to compute the upper bound of $E_J(Q,W)$

by Arimoto's method.

Another question we addressed is: for given DMS Q and DMC W, what is the rate R_m for which the joint source-channel lower (upper) bound is attained? In the analytical computation of (5.1), we related R_m and ρ^* by $R_m = H\left(\frac{q_i^{\frac{1}{1+\rho^*}}}{\sum_{i=1}^N q_i^{\frac{1}{1+\rho^*}}}\right)$, where ρ^* achieves the lower (upper) bound of $E_J(Q,W)$ in (5.1). Proposition 4.1 gives a sufficient and necessary condition to verify whether the joint source-channel exponent is determined exactly or not. From (5.1) once again, we also reveal that Gallager's lower bound coincides with Csiszár's result. With a small modification, we can also compute the improved expurgated lower bound if some conditions are satisfied, and the condition when the lower bound can be improved is provided.

Next we employ our results to compare the joint source-channel exponent $E_J(Q,W)$ with the tandem source-channel exponent $E_T(Q,W)$. It is obviously seen that, if any of $E_J(Q,W)$ or $E_T(Q,W)$ is exactly determined (see Proposition 4.4), the joint source-channel coding strictly outperforms the tandem source-channel coding, and numerical examples show that $E_J(Q,W)$ substantially improves $E_T(Q,W)$ for the same source and channel pair. If both of $E_J(Q,W)$ and $E_T(Q,W)$ are not exactly known, we have to use the lower and upper bounds to approximate them. However, we cannot show that $E_J(Q,W) > E_T(Q,W)$ for all source and channel pairs due to the uncertainty of the channel reliability function for $R < R_{cr}$. But we give some sufficient conditions for a DMS Q and DMC W pair for which $E_J(Q,W)$ can be strictly bigger than $E_T(Q,W)$. Fortunately, numerical examples show that our conditions hold for a large class of source-channel pairs.

In the last part of this project, we investigate whether we can generalize our computation results for the lossy source-channel exponent under the Hamming distortion. When the DMS is a binary source, we observe that the source reliability function $F(R,Q,\Delta)$ under the distortion threshold Δ can be actually seen as the $e(R,Q)$ shifted to the left on the abscissa by a constant $c = h_b(\Delta)$. Due to the fact that

$$E(R, W) + e(R + c, Q) = E(R' - c, W) + e(R', Q),$$

where $E(R, W)$ might be $E_r(R, W)$ or $E_{sp}(R, W)$, we can therefore compute the joint source-channel exponent in the same way.

Appendix A

Proof of Lemma 4.4

Proof:

Lemma A.1 *Let P_X denote the channel input distribution, the necessary and sufficient conditions for P_X to maximize $E_o(\rho)$ (given in (2.16)) are*

$$\sum_y P_{Y|X}(y \mid x)^{\frac{1}{1+\rho}} \alpha_y(P_X)^\rho \geq \sum_y \alpha_y(P_X)^{1+\rho} \qquad \forall \quad x \tag{A.1}$$

with equality for all x such that $P_X(x) > 0$. The function $\alpha_y(P_X)$ is given by

$$\alpha_y(P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}}$$

Proof: see Gallager [4, p. 144]. □

Next we show that, for any symmetric DMC, (A.1) is satisfied by uniform distribution. Given a symmetric channel $P_{Y|X}$ with input alphabet size $|\mathcal{X}| = N$ and output alphabet size $|\mathcal{Y}| = M$, consider the n^{th} $(1 \leq n \leq N)$ row and the m^{th} $(1 \leq m \leq M)$

column in the transition matrix as shown below

$$P_{Y|X} = \begin{bmatrix} \cdots & \cdots & \cdots & P_{1,m} & \cdots & \cdots \\ \cdots & \cdots & \cdots & P_{2,m} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{n,1} & P_{n,2} & \cdots & P_{n,m} & \cdots & P_{n,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & P_{N,m} & \cdots & \cdots \end{bmatrix}.$$

Denote set $\mathcal{A} = \{P_{1,m}, P_{2,m}, ..., P_{N,m}\}$, and set $\mathcal{B} = \{P_{n,1}, P_{n,2}, ..., P_{n,M}\}$. When we compare the two sets, identical elements in $\mathcal{A}$ and $\mathcal{B}$ are listed once. For example: set $\{0.2, 0.3, 0.2\}$ is equal to set $\{0.2, 0.3\}$. By definition of a symmetric DMC, the columns of the transition matrix $P_{Y|X}$ are permutations of each other, so each element in the n^{th} row satisfies $P_{n,j} \in \mathcal{A}$ $(j = 1, 2..., M)$, which implies that $\mathcal{B} \subseteq \mathcal{A}$. Similarly, it is easy to see that each element in the m^{th} column must also be in $\mathcal{B}$ and $\mathcal{A} \subseteq \mathcal{B}$. Thus $\mathcal{A} = \mathcal{B}$. Therefore, each row and each column has the same elements and may differ in the frequency of each element. For example, the symmetric channels

$$P_{Y|X} = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.2 & 0.5 & 0.3 \\ 0.5 & 0.3 & 0.2 \end{bmatrix} \quad and \quad P_{Y|X} = \begin{bmatrix} 0.3 & 0.2 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.3 \end{bmatrix}$$

have the same set of elements in each row and column. Given the channel transition matrix W, each column is a permutation of the same set, denoted by

$$\{\underbrace{p_1, ..., p_1}_{a_1}, \underbrace{p_2, ..., p_2}_{a_2}, \quad ..., \quad \underbrace{p_s, ..., p_s}_{a_s}\},$$

where a_i is the frequency of element p_i $(i = 1, ..., s)$ $(p_i \neq p_j$ if $i \neq j)$ such that $\sum_{i=1}^{s} a_i = N$, and each row of the channel transition matrix W is a permutation of the set given by

$$\{\underbrace{p_1, ..., p_1}_{b_1}, \underbrace{p_2, ..., p_2}_{b_2}, \quad ..., \quad \underbrace{p_s, ..., p_s}_{b_s}\},$$

such that $\sum_{i=1}^{s} b_i = M$. Furthermore,

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \cdots = \frac{a_s}{b_s} = \frac{N}{M}$$

since $a_i M = b_i N =$ the total number of element p_i in the matrix W. When the channel input distribution $P_X(x) = \frac{1}{N}$, substituting

$$\alpha_y(P_X) = \frac{1}{N}\sum_{i=1}^{s}(a_i p_i^{\frac{1}{1+\rho}})$$

into the left side of (A.1), yields

$$\begin{aligned}
\sum_y P_{Y|X}(y \mid x)^{\frac{1}{1+\rho}} \alpha_y(P_X)^\rho &= \left(\frac{1}{N}\sum_{i=1}^{s}(a_i p_i^{\frac{1}{1+\rho}})\right)^\rho M \frac{1}{M}\sum_{i=1}^{s}(b_i p_i^{\frac{1}{1+\rho}}) \\
&= \left(\frac{1}{N}\sum_{i=1}^{s}(a_i p_i^{\frac{1}{1+\rho}})\right)^\rho M \frac{1}{N}\sum_{i=1}^{s}(a_i p_i^{\frac{1}{1+\rho}}) \\
&= M\left(\frac{1}{N}\sum_{i=1}^{s}(a_i p_i^{\frac{1}{1+\rho}})\right)^{1+\rho} = \sum_y \alpha_y(P_X)^{1+\rho} \qquad \forall \quad x.
\end{aligned}$$

Now consider a quasi-symmetric DMC $\{W : \mathcal{X} \to \mathcal{Y}\}$. Suppose W can be partitioned along its columns into symmetric sub-matrices $W_1, W_2, ..., W_t$, each W_i having size $|\mathcal{X}| \times |\mathcal{Y}_i|$. It follows that

$$\begin{aligned}
\sum_{y\in\mathcal{Y}} P_{Y|X}(y \mid x)^{\frac{1}{1+\rho}} \alpha_y(P_X)^\rho &= \sum_{i=1}^{t}\sum_{y\in\mathcal{Y}_i} P_{Y|X}(y \mid x)^{\frac{1}{1+\rho}} \alpha_y(P_X)^\rho \\
&= \sum_{i=1}^{t}\sum_{y\in\mathcal{Y}_i} \alpha_y(P_X)^{1+\rho} \\
&= \sum_{y\in\mathcal{Y}} \alpha_y(P_X)^{1+\rho} \qquad \forall \quad x. \qquad \text{(A.2)}
\end{aligned}$$

This completes the proof. $\square$

Appendix B

Arimoto's Iteration Algorithm

The key to compute

$$E(\rho) = -(1+\rho)\log\left[\sum_{i=1}^{|\mathcal{S}|} Q_i^{\frac{1}{1+\rho}}\right] + \rho\left[\max_{P_X}\max_{\Phi_{X|Y}} G_\rho(P_X, \Phi_{X|Y})\right],$$

for a given ρ is to compute

$$C_\rho \triangleq \max_{P_X}\max_{\Phi_{X|Y}} G_\rho(P_X, \Phi_{X|Y}).$$

In [10], Arimoto developed the following iteration algorithm to compute C_ρ.

1) Initialize $t = 0$, and let $P_X^{(t)}(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$;

2) Then iterate the following steps for $t = 1, 2, 3, \cdots$,

$$\Phi_{X|Y}^{(t)}(x|y) = \frac{P_X^{(t)}(x) P_{Y|X}^{1/(1+\rho)}(y|x)}{\sum_{\widetilde{x}\in\mathcal{X}} P_X^{(t)}(\widetilde{x}) P_{Y|X}^{1/(1+\rho)}(y|\widetilde{x})},$$

$$P_X^{(t+1)}(x) = \frac{\left[\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)\left\{\Phi_{X|Y}^{(t)}(x|y)\right\}^{-\rho}\right]^{-1/\rho}}{\sum_{\widetilde{x}\in\mathcal{X}}\left[\sum_{y\in\mathcal{Y}} P_{Y|X}(y|\widetilde{x})\left\{\Phi_{X|Y}^{(t)}(\widetilde{x}|y)\right\}^{-\rho}\right]^{-1/\rho}}.$$

The following results show how this iteration converges to C_ρ.

Lemma B.1 *[10]*

$$\begin{aligned} G_\rho(P_X^{(0)}, \Phi_{X|Y}^{(0)}) &\leq G_\rho(P_X^{(1)}, \Phi_{X|Y}^{(0)}) \leq \cdots \leq G_\rho(P_X^{(t)}, \Phi_{X|Y}^{(t)}) \\ &\leq G_\rho(P_X^{(t+1)}, \Phi_{X|Y}^{(t)}) \leq G_\rho(P_X^{(t+1)}, \Phi_{X|Y}^{(t+1)}) \leq \cdots \leq C_\rho. \end{aligned} \tag{B.1}$$

Theorem B.1 *[10] The sequence* $\left\{G_\rho(P_X^{(t)}, \Phi_{X|Y}^{(t)})\right\}$ *or* $\left\{G_\rho(P_X^{(t+1)}, \Phi_{X|Y}^{(t)})\right\}$ *converges monotonically from below to* C_ρ *as* $t \to \infty$*, for any* $\rho \in (-1, \infty)$*.*

Bibliography

[1] I. Csiszár, "Joint Source-Channel Error Exponent," *Probl. Contr. Inform. Theory*, vol. 9, pp. 315–328, 1980.

[2] I. Csiszár, "On the Error Exponent of Source-Channel Transmission with a Distortion Threshold," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 823–828, Nov. 1982.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[4] R. G. Gallager, *Information Theory and Reliability Communication*. New York: Wiley, 1968.

[5] Brian D. Bunday, *Basic Optimisation Methods*. London: Arnold, 1984.

[6] F. Jelinek, *Probabilistic Information Theory*. New York, McGraw Hill, 1968.

[7] T. M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991

[8] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, Jul. 1974.

[9] K. Marton, "Error Exponent for Source Coding with a Fidelity Criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

[10] S. Arimoto, "Computation of Random Coding Exponent Functions," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 665–671, Nov. 1976.

[11] R. Ahlswede, "Extremal Properties of Rate-Distortion Functions," *IEEE Trans. Inform. Theory*, vol. 36, pp. 166–171, Jan. 1990.

[12] I. Csiszár, "The Method of Types," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2505–2523, Oct. 1998.