Sub-optimality of Achievable Regions in Two Fundamental Network Information Theory Settings

YAZDANPANAH, Mehdi

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Doctor of Philosophy

in

Information Engineering

The Chinese University of Hong Kong

December 2018

Abstract of thesis entitled: Sub-optimality of Achievable Regions in Two Fundamental Network Information Theory Settings

Submitted by YAZDANPANAH, Mehdi

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in December 2018

This thesis establishes the sub-optimality of two specific achievable regions: one for the interference channel and the other for the three-receiver broadcast channel with two degraded message sets. Determining the optimality or the sub-optimality of the two achievable regions were posed as open questions in the book Network Information Theory, [11].

An interference channel models multiple point-to-point communication links over a shared medium. Han and Kobayashi formulated an achievable region for two receiver interference channels in 1981. In particular, open problem 6.4 in [11] asks: Is the Han–Kobayashi region tight in general? We answer this question in the negative by exhibiting specific interference channels where achievable rates are found that lie outside the Han–Kobayashi region.

A broadcast channel models a downlink scenario where a single sender wishes to reliably communicate (possibly) different messages simultaneously to different receivers. A two degraded message sets setting is one where there are two independent messages, and there is a set of receivers that wishes to decode one of

the messages, while the rest of the receivers wish to decode both of the messages. For the case of two receivers, the capacity region for this setting was determined in a seminal work by Korner and Marton in 1977. The extension of this result to three or more receivers has remained open since then. In particular, open problem 8.2 in [11] asks: Is superposition coding optimal for this setting? We again answer this question in the negative by exhibiting specific scenarios where achievable rates are found that lie outside the superposition coding region.

The two results mentioned above have been obtained by using the same (rather well-known) idea: by demonstrating that the multi-letter extensions of the candidate achievable regions outperform the original ones for specific examples. The main contribution of this thesis is in developing techniques and ideas that enable one to make explicit characterizations of the extremizers of the non-convex optimization problems that are needed to evaluate the various achievable regions. This thesis also makes novel contributions related to developing outer bounds to the capacity regions in the specific examples under consideration that improve on the previously best-known ones.

摘要

本文研究是信息中的基通信模型，即干信道和具有降信息集合和三接收者的播信道。於通信模型，作者明了目前已知可到的最大速率非其最信道容量。

干信道是一用模 或更多 接受端通 共享媒介 行通 的模型。1981年，Han和Kobayashi 推 出干 信道 的一 可 到的 速率 域。 自此之後， 速率 域是否是其信道容量的疑 而未 三十年。因此，《 信息 》一中在尚未解 的6.4中提出：Han-Kobayashi 速率域是否就是干信道的最解？本篇文中，作者在特定的干信道中，可到的最大速率超出Han-Kobayashi 速率域。通反例，作者而未的出了否定的答案。

播信道是一用模一端想同(可能)不同的信息定送不同的接收端的模型。特地，如果有互相立的信息，而且一部分接收端只想解接收其中一信息，而其接收端想要解接收信息，我之具有降信息集合的播信道。考只有接收端的情，1977年Korner和Marton性地通加的方法出了情下的信道容量。然而自那之後，果在三或者更多接收端的情下的推是否就是播信道的信道容量的疑一直未能解。因此，《信息》一中在尚未解的8.2中提出：加在三或者更多接收端的情下是否就是播信道的最解？本篇文中，作者同找到了特定的播信道中，可到的最大速率超出加的速率域。藉由反例，作者出了否定的答案。

以上果都是用相同的（同也是相有名的）思路明的：找出特定的例子，使得待定的可到速率域在送多字下的延伸原本的可到速率域。本文的主要是展了

一系列思想和方法，可以助明算定各可到的速率域中所需解的非凸化中的最解。文也在推特定例子下的信道容量的上界上做出了新性的工作，所推出的信道容量的上界之前已知的最佳上界。

Acknowledgement

I would like to acknowledge my adviser Professor Chandra Nair for his exemplary academic guidance and mentorship over the past five years. His structured, critical approach has greatly inspired me, as has his discipline and conduct.

I would also like to acknowledge the contributions of the thesis committee members who read through drafts of this work. Professor Amin Aminzadeh Gohari, Professor Soung-Chang Liew, Professor Raymond Yeung and Professor Angela Yingjun Zhang provided helpful recommendations which enabled me to improve this thesis.

I have been lucky to have the support of great friends and family during my graduate life. Mehrdad Tahernia is an outstanding friend who became like a family member during our graduate studies. Rachel is my daily source of love, motivation and encouragement.

Finally I am grateful to my mum and dad without whom none of this would have been possible. Their love and support has been uncompromising, and their hard work continues to inspire me. To me they are the very best.

This thesis is dedicated to my mother and father.

Contents

List of Figures

Notations

This is a general guideline for the notation used in this thesis. Deviations from this guideline are pointed out as they happen.

- script letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$... denote finite sets and $|\mathcal{X}|$ is the cardinality of set $\mathcal{X}$.
- $\mathbb{R}$ is the real line and $\mathbb{R}^d$ is the d-dimensional Euclidean space.
- $\mathbb{Z}$ and $\mathbb{N}$ respectively denote integer and natural numbers.
- Lowercase letters $x, y, z, ...$ denote constants and values of random variables.
- $\bar{x} = 1 - x$.
- Uppercase letters U, X, Y, Z, Q denote random variables.
- Uppercase letter W denotes channel transition matrices.
- We use $X_i^j = (X_i, X_{i+1}, ..., X_j)$ to denote a $(j - i + 1)$ vector for $1 \leq i \leq j$.

  When $i = 1$ we drop the subscript, i.e., $X^j = (X_1, X_2, ..., X_j)$.
- Script letters $\mathscr{A}, \mathscr{C}, \mathscr{S}$ denote rate regions $\subseteq \mathbb{R}^d$.
- $\otimes$ is the Kronecker product operator and $W^{\otimes n} = \underbrace{W \otimes ... \otimes W}_{n}$.

- $\oplus_M$ is the Minkowski sum of two sets.

- $[i : 2^a] = \{i, i+1, ..., \lfloor 2^a \rfloor\}$, where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.

- Probability of an event $\mathcal{A}$ is denoted by $P(\mathcal{A})$

- $X \sim p(x^n)$ means that $p(x^n)$ is the probability mass function (pmf) of the discrete random vector X^n. The function $p_{X^n}(\tilde{x}^n)$ is equal to $P\{X^n = \tilde{x}^n\}$ for all $\tilde{x}^n \in \mathcal{X}^n$. We drop the subscript when refering to $p_{X^n}(x^n)$.

- $p(y^n|x^n)$ is a collection of conditional pmfs, one for every $x^n \in \mathcal{X}^n$.

- We say that $X \to Y \to Z$ forms a Markov chain if $p(x, y, z) = p(x)p(y|x)p(z|y)$.

- The upper concave envelope of a function $f(x)$ over domain $\mathcal{D}$ is defined as

$$\mathfrak{C}[f(x)](x_0) = \inf\{g(x_0) : g(x) \text{ is concave in } x \in \mathcal{D}, g(x) \geq f(x) \forall x \in \mathcal{D}\}.$$

- $H_b(x)$ denotes the binary entropy function

$$H_b(x) := -x \log x - (1 - x) \log(1 - x).$$

Remark: We extensively use information-theoretic quantities such as entropy, conditional entropy, mutual information, and their properties throughout this thesis. Since these are named quantities (or inequalities) and we use the standard notation (such as the one in [11]), we do not repeat them here.

Chapter 1

Introduction

The tenets of information theory were developed by Shannon in his seminal paper [27]. This work laid the foundations for digital communications and understanding the limits of communication for a point-to-point setting. The principles and ideas have led to many advances in communications and coding; the point-to-point communication setting is considered a rather mature field.

Network information theory, on the other hand, studies the fundamental limits of communication in a multi-user (network) setting where several communication requirements happen simultaneously over a shared medium; or in the case of wireless communication, over a shared electromagnetic spectrum. To develop a theory for understanding these limits it is imperative that we understand the limits of the basic building blocks. Two of the basic building blocks of a network communication setting are the broadcast channel and the interference channel.

Broadcast Channel

A broadcast channel models the simultaneous communication of information from one source to several receivers. The information may be independent or nested. Examples of broadcast channel include digital TV broadcasting or communication of a cellular tower to cell phone users in its coverage area [6].

A 2-receiver discrete memoryless broadcast channel consists of an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, all of finite sizes, and a probability transition matrix $W(y, z|x)$.

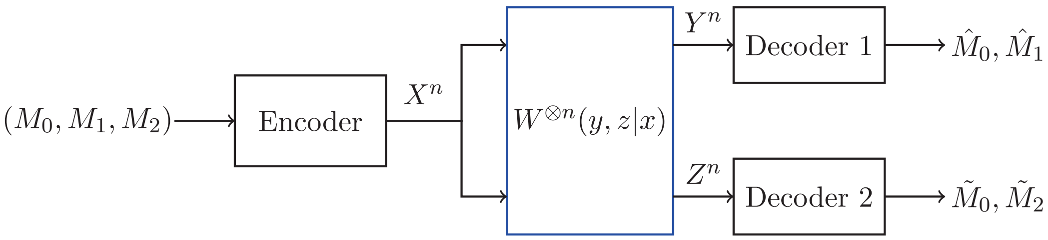


Figure 1.1: A two-receiver memoryless broadcast channel

Figure 1.1 represents a communication model where there is a single transmitter and two receivers. A message M_0 needs to be communicated to both the receivers, while messages M_1 and M_2 needs to be communicated to its intended receivers Y and Z respectively. A $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$ code for this model consists of

- Message sets $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ of sizes $\lfloor 2^{nR_0} \rfloor$, $\lfloor 2^{nR_1} \rfloor$, and $\lfloor 2^{nR_2} \rfloor$, respectively. The messages M_0, M_1, and M_2 are assumed to be independent and uniformly distributed over the message sets.

- An encoder that maps the set of possible message triples to sequences of input symbols, i.e., $x^n = \mathbf{Enc}(m_0, m_1, m_2)$ where $x^n \in \mathcal{X}^n$.

- Two decoders, one at each receiver, that maps its received sequence to an estimated message pair, i.e., $(\hat{m}_0, \hat{m}_1) = \mathbf{Dec_1}(y^n)$ and $(\tilde{m}_0, \tilde{m}_2) = \mathbf{Dec_2}(z^n)$, where $(\hat{m}_0, \hat{m}_1) \in \mathcal{M}_0 \times \mathcal{M}_1$, and $(\tilde{m}_0, \tilde{m}_2) \in \mathcal{M}_0 \times \mathcal{M}_2$.

The probability of error, $P_\epsilon^{(n)}$, is defined as

$$P_\epsilon^{(n)} = \mathrm{P}\left(\left\{(\hat{M}_0, \hat{M}_1) \neq (M_0, M_1)\right\} \cup \left\{(\tilde{M}_0, \tilde{M}_2) \neq (M_0, M_2)\right\}\right).$$

A rate triple (R_0, R_1, R_2) is said to be **achievable** if there is a sequence of $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$ codes for which the probability of error goes to zero as n goes to infinity. The **capacity region** is defined as the closure of the set of all achievable rate triples.

Interference Channel

An interference channel models simultaneous communication of messages between sender receiver pairs over a shared medium. A 2-receiver discrete memoryless interference channel consists of two input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, all of finite sizes, and a probability transition matrix $W(y_1, y_2|x_1, x_2)$.

Figure 1.2 models a communication setting where there are two sender-receiver pairs. One transmitter wishes to communicate a message M_1 to its receiver, while the other transmitter wishes to communicate an independent message M_2 to its receiver. A $(2^{nR_1}, 2^{nR_2}, n)$ code for this model consists of

- Message sets $\mathcal{M}_1$ and $\mathcal{M}_2$ of sizes $\lfloor 2^{nR_1} \rfloor$ and $\lfloor 2^{nR_2} \rfloor$, respectively. The messages M_1 and M_2 are assumed to be independent and uniformly distributed

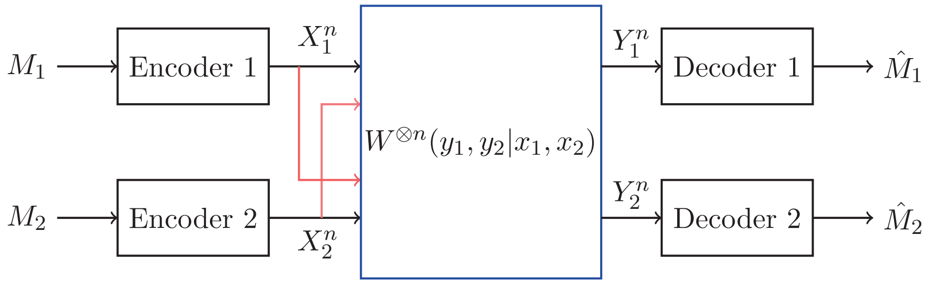


Figure 1.2: A two-receiver memoryless interference channel

over the message sets.

- Two independent encoders that separately map the messages to sequences of input symbols, i.e., $x_1^n = \mathbf{Enc}(m_1)$ and $x_2^n = \mathbf{Enc}(m_2)$ where $x_1^n \in \mathcal{X}_1^n$ and $x_2^n \in \mathcal{X}_2^n$.

- Two decoders, one at each receiver, that map its received sequence to an estimated message, i.e., $\hat{m}_1 = \mathbf{Dec_1}(y_1^n)$ and $\hat{m}_2 = \mathbf{Dec_2}(y_2^n)$, where $\hat{m}_1 \in \mathcal{M}_1$, and $\hat{m}_2 \in \mathcal{M}_2$.

The probability of error, $P_\epsilon^{(n)}$, is defined as

$$P_\epsilon^{(n)} = \mathrm{P}\left(\left\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\right\}\right).$$

A rate pair (R_1, R_2) is said to be **achievable** if there is a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes for which the probability of error goes to zero as n goes to infinity. The **capacity region** is defined as the closure of the set of all achievable rate pairs.

Generic setting

A set of data rate tuples is said to be **achievable** for a given communication setting if there is a sequence, in the number of channel uses, of encoding and decoding strategies such that the probability of decoding error goes to zero as the number of channel uses tends to infinity. The closure of the set of all possible achievable rate tuples is defined to be the **capacity region** for the given communication setting.

On computable characterizations of capacity regions

The holy-grail for such communication problems in network information theory is to obtain a *computable-characterization* of the capacity region. As is wont in the information theory literature, we will adopt a narrower definition of computability wherein we would like to compute the capacity region as an optimization problem of a continuous function over a compact subset of a finite dimensional Euclidean space. In particular optimization problems, that evaluate to the capacity region, involving functionals defined over a single-use of the channel are informally referred to as **single-letter characterizations**.

Computable characterizations of the capacity region of the models depicted in Figures 1.1 and 1.2 are *central open questions* in network information theory.

1.1 Background and Summary of contributions

A vast majority of the major contributions in network information theory came during the 1970's and early 1980's, and several achievable regions for fundamental settings were formulated during this period. The optimality of a few of them had been settled early on; yet the optimality of a large number of them remained undetermined.

Being two fundamental building blocks of the communication networks field, considerable effort has been expended on studying the capacity regions of broadcast and interference channel models. There have been successful characterizations of the capacity region for special classes of broadcast and interference channels [1–3, 5, 8–10, 12, 14, 15, 17–21, 25, 26, 29]. We refer interested readers to Chapters 5, 6, 8, and 9 of the book [11] for an overview of known results and techniques.

For instance, open problems numbered 5.1, 5.2, 6.1, 6.4, 8.2, 8.3, 8.4, 9.3 in [11] concern the capacity regions or the optimality of certain achievable regions for some classes of interference and broadcast channels. For each question above, there is a candidate (natural) achievable region whose optimality or sub-optimality had not been established and the answers to these were considered the natural next steps for improving our understanding of the state-of-the-art coding schemes. The candidate achievable region for open problem 5.2 had been shown to be sub-optimal in [22], while the optimality of the candidate region for problem 9.3 was established in [13]. This thesis demonstrates the sub-optimality of the candidate achievable regions for open questions 6.4 and 8.2. The results in this thesis first appeared in [23] and [24] respectively.

1.1.1 Superposition coding region and the broadcast channel

Cover [6] introduced the idea of superposition coding motivated by channels where one receiver is stronger than the other receiver. The idea is to have the codewords grouped into well-separated clusters with the "cluster-centers" carrying the message that both receivers decode, denoted by an auxiliary codebook $\{U^n\}$, and the codewords $\{X^n\}$ within a cluster carrying different messages for the other (stronger) receiver. An illustration of the superposition coding idea is given in Figure 1.3.

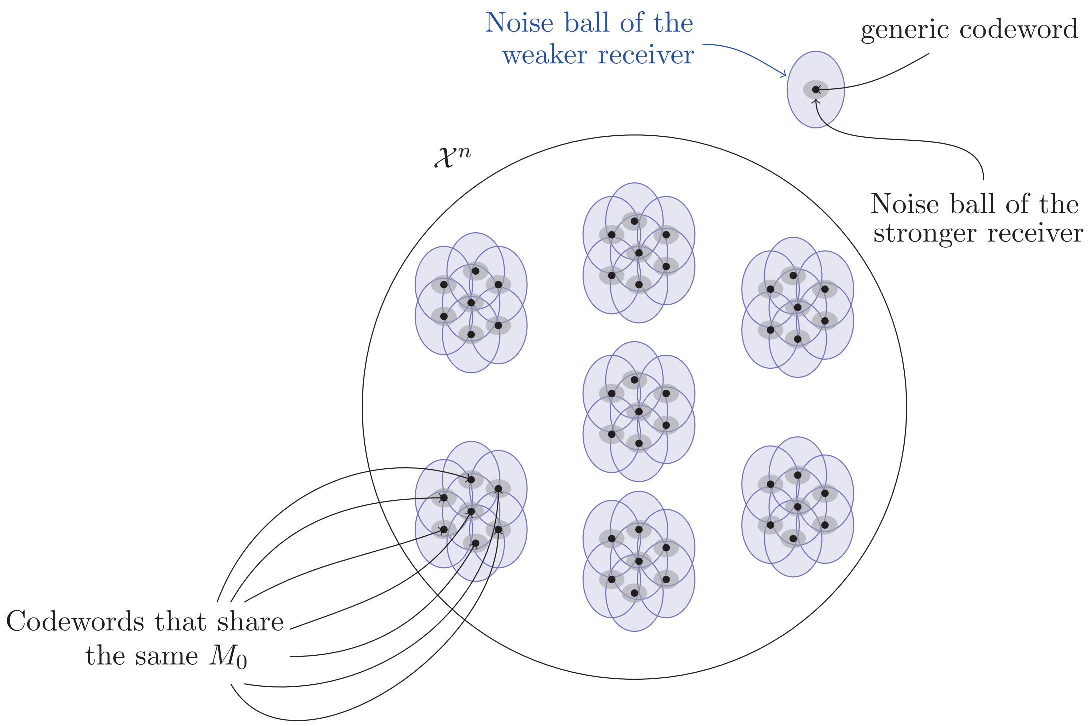


Figure 1.3: Superposition coding strategy

Consider the communication model depicted in Figure 1.1. Superposition coding strategy can be used to obtain the rate-triple $(R_0, R_1, 0)$ stated in Theorem 1.

Theorem 1 ((8.1) & Theorem 8.1 in [11], Superposition coding region)**.** *The set of rate pairs (R_0, R_1) that satisfy*

$$R_0 \leq I(U;Z),$$

$$R_0 + R_1 \leq I(U;Z) + I(X;Y|U),$$

$$R_0 + R_1 \leq I(X;Y),$$

for some $p(u,x)$ with $|\mathcal{U}| \leq |\mathcal{X}| + 1$ is achievable.

In a seminal paper Korner and Marton [15] established that the superposition coding region matches restriction of the capacity region, of the communication model in Figure 1.1, to the plane $R_2 = 0$.

Remark 1. Note that the restriction of the capacity region for the communication model in Figure 1.1 to the plane $R_1 = R_2 = 0$ asks for the maximum rate of the common message that can be simultaneously communicated to both the receivers. In this case this maximum rate is given by $\max_{p(x)} \min\{I(X;Y), I(X;Z)\}$. This result also extends naturally to the case when there are three or more receivers.

A natural follow-up question is the following: does the optimality of the superposition coding region extend to the case when there are three or more receivers? The two simplest extensions along these lines are the following:

- Setting A: In this setting the common message M_0 is to be decoded by all the three receivers while the message M_1 is to be decoded by one of the receivers.
- Setting B: In this setting the common message M_0 is to be decoded by all the three receivers while the message M_1 is to be decoded by two of the receivers.

For the setting A, Nair and El Gamal [19] showed that the extension of the superposition coding region is strictly sub-optimal. The idea was to exploit the channel diversity to the receivers requiring only the common message to improve the rate region. The use of channel diversity was done through an indirect decoding idea which is presented in Section 8.2 of [11].

For the setting B, two receivers need to decode the transmitted codeword while one receiver needs to decode the cluster corresponding to the common message. Since there is only one receiver requiring only the common message, there is no channel diversity to be exploited here. The region obtained using an extension of the indirect decoding approach also collapsed to the superposition coding region for setting B (see Proposition 9 in [19]). Thus open question 8.2 in [11] asked: Is superposition coding optimal for the general 3-receiver discrete memoryless broadcast channel with one message to all three receivers and another message to two receivers?

In answer to this, one of the main results in this thesis is the following theorem.

Theorem 2. *There are channel settings for which the superposition coding region is strictly inside the capacity region for the three-receiver broadcast channel with one message to all three receivers and another message to two receivers.*

The proof of this theorem follows immediately from Theorem 7 in Chapter 2.

1.1.2 Han–Kobayashi region and the interference channel

In interference channel model depicted in Figure 1.2 each receiver receives a noisy version of a combination of all the transmitted signals. Two natural coding strategies in the presence of interference are either to treat the interference as noise or to completely decode the interference and cancel it. Han and Kobayashi devised a coding strategy that incorporated both these strategies. Each receiver decodes a part of the interfering message and treats the undecoded part of the interference as noise. This scheme includes the two natural strategies as the two extreme cases.

Theorem 3 (Theorem 6.4 in [11], Han–Kobayashi achievable region)**.** *A rate-pair* (R_1, R_2) *is achievable for the channel,* W*, described in Figure 1.2 if*

$$\begin{aligned}
R_1 &< I(X_1; Y_1|U_2, Q),\\
R_2 &< I(X_2; Y_2|U_1, Q),\\
R_1 + R_2 &< I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q),\\
R_1 + R_2 &< I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q),\\
R_1 + R_2 &< I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q),\\
2R_1 + R_2 &< I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q),\\
R_1 + 2R_2 &< I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q)
\end{aligned}$$

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$*, where* $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 4$*,* $|\mathcal{U}_2| \leq |\mathcal{X}_2| + 4$*, and* $|\mathcal{Q}| \leq 7$*.*

Let the achievable region given by Theorem 3 be denoted by $\mathscr{A}^{\mathrm{HK}}(W)$.

Remark 2. Auxiliary random variables in Theorem 3 represent the interference that is decoded by the unintended receiver. That is, U_1 is the part of M_1 that Y_2 decodes and U_2 is the part of M_2 that Y_1 decodes.

A simpler achievable region for the interference channel can be obtained by treating interference as noise.

Theorem 4 (Interference-as-noise achievable region)**.** *A rate-pair* (R_1, R_2) *is achievable for the channel,* W*, described in Figure 1.2 if*

$$R_1 < I(X_1; Y_1|Q), \tag{1.1}$$

$$R_2 < I(X_2; Y_2|Q), \tag{1.2}$$

for some pmf $p(q)p(x_1|q)p(x_2|q)$*, where* $|\mathcal{Q}| \leq 2$*.*

Let the achievable region given by Theorem 4 be denoted by $\mathscr{A}^{\text{TIN}}(W)$.

Corollary 1. *$\mathscr{A}^{TIN}(W) \subseteq \mathscr{A}^{HK}(W)$ for all W.*

Proof. This can be seen immediately by setting $U_1 = U_2 = 0$, constant random variables, in the Han–Kobayashi achievable region. □

Remark 3. *Interference cancellation region* is a special case of the Han–Kobayashi region when $U_1 = X_1, U_2 = X_2$. This strategy achieves the set of rate pairs that satisfy

$$R_1 \leq I(X_1; Y_1 | X_2, Q),$$

$$R_2 \leq I(X_2; Y_2 | X_1, Q),$$

$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_1 | Q), I(X_1, X_2; Y_2 | Q)\}$$

for some $p(q)p(x_1|q)p(x_2|q)$.

Open question 6.4 in [11], naturally, asked: Is Han–Kobayashi bound tight in general?

In answer to this, the second main result of this thesis is the following.

Theorem 5. *Han–Kobayashi region is not optimal for the interference channel with two sender-receiver pairs.*

The proof follows immediately from Corollary 7 in Chapter 3.

1.2 A generic approach for testing the optimality of achievable schemes

The idea that we explain in this section (Lemma 1) is not new. The main contribution of this thesis however, is to make the idea in Lemma 1 work in the two instances mentioned previously.

To describe the idea, we first define the multi-letter extension of an achievable region for a communication setting S. Let us fix a generic communication setting S and an achievable strategy for this setting. For a channel W the achievable strategy induces an achievable region, denoted by $\mathscr{A}(W) \subseteq \mathbb{R}_+^d$ for some finite dimension d. In general we can assume that: (i) $\mathscr{A}(W)$ is closed, (ii) $\alpha\mathscr{A}(W) \subseteq \mathscr{A}(W)$, $\forall\alpha \in [0,1]$, and (iii) using a time-sharing argument that $\mathscr{A}(W)$ is convex. By viewing the k consecutive time-slots we obtain the channel $W^{\otimes k}$, and the same achievable strategy induces a region $\mathscr{A}(W^{\otimes k})$ for this *k-letter extension* of the original channel. Clearly the region $\frac{1}{k}\mathscr{A}(W^{\otimes k})$ is achievable region for the original channel by treating k consecutive time-slots as a single large time-slot.

Definition 1. An achievable strategy defined for a generic communication setting S is said to be **asymptotically capacity achieving** if the sequence of regions $\frac{1}{n}\mathscr{A}(W^{\otimes n})$ converges to the capacity region $\mathscr{C}(W)$ for every W.

Lemma 1. *An asymptotically capacity achieving achievable strategy defined for a communication setting S is optimal if and only if*

$$\mathscr{A}(W^{\otimes 2}) = \mathscr{A}(W) \oplus_M \mathscr{A}(W) \quad \forall W.$$

Proof. First we show that $\mathscr{A}(W^{\otimes 2}) = \mathscr{A}(W) \oplus_M \mathscr{A}(W)$ $\forall W$, implies optimality. From the convexity of $\mathscr{A}(W)$ for any W, it is immediate that

$$\mathscr{A}(W) = \frac{1}{2}\left(\mathscr{A}(W) \oplus_M \mathscr{A}(W)\right) = \frac{1}{2}\mathscr{A}(W^{\otimes 2}).$$

Hence by induction, for all $k \geq 1$ we have

$$\mathscr{A}(W) = \frac{1}{2^k}\mathscr{A}(W^{\otimes 2^k}) \quad \forall W.$$

Since $\mathscr{A}$ is asymptotically capacity achieving, $\forall \epsilon > 0,\ \exists N_\epsilon$ such that

$$n(1-\epsilon)\mathscr{C}(W) \subseteq \mathscr{A}(W^{\otimes n}) \quad \forall n \geq N_\epsilon.$$

Considering k such that $2^k > N_\epsilon$, we obtain

$$2^k(1-\epsilon)\mathscr{C}(W) \subseteq \mathscr{A}(W^{\otimes 2^k}) = 2^k \mathscr{A}(W).$$

This implies that $(1-\epsilon)\mathscr{C}(W) \subseteq \mathscr{A}(W)$. Since $\mathscr{A}(W)$ is assumed to be closed, by taking $\epsilon \to 0$ we obtain $\mathscr{C}(W) \subseteq \mathscr{A}(W)$, the non-trivial direction. This concludes the "if" direction.

To show the other direction, if there is a W such that $\mathscr{A}(W^{\otimes 2}) \supsetneq \mathscr{A}(W) \oplus_M \mathscr{A}(W)$, then as $\mathscr{C}(W) \supseteq \frac{1}{2}\mathscr{A}(W^{\otimes 2})$ (argued earlier), we see that $\mathscr{C}(W) \supsetneq \mathscr{A}(W)$.

□

Definition 2. An achievable strategy defined for a generic communication setting S is said to be **super-additive** if $\mathscr{A}(W^{\otimes(m+n)}) \supseteq \mathscr{A}(W^{\otimes m}) \oplus_M \mathscr{A}(W^{\otimes n})$ where $\oplus_M$ denotes the Minkowski sum of the two regions.

Remark 4. Most natural achievable strategies are super-additive for memoryless channels as one can concatenate (small-probability of error) codes of length m for $W^{\otimes m}$ and codes of length n for $W^{\otimes n}$ to obtain a (small-probability of error) code of length $(m+n)$ for $W^{\otimes(m+n)}$.

1.2.1 Remarks on employing the strategy for testing optimality

There are several instances in network information theory, see open problems 5.1, 5.2, 6.1, 6.4, 8.2, 8.3, 8.4, 9.3 in [11] for a sub-collection, where the optimality of a given achievable scheme is not known. Given infinite computational power and time, the generic method described above should enable one to answer each of these questions. In particular, to show sub-optimality one just needs to exhibit a W for which $\mathscr{A}(W^{\otimes 2}) \supsetneq 2\mathscr{A}(W)$. The main difficulty in carrying out this program easily is the difficulty in the evaluation of $\mathscr{A}(W)$. Even for small problem instances, optimization problems that come up in the evaluation of $\mathscr{A}(W)$ are usually non-convex and the dimensionality of the space makes it practically infeasible to test the containment. Hence, to employ this generic strategy, one needs to come up with new tools and ideas to tackle the non-convex problems that arise as well as develop strong intuitions for the classes of channels for which $\mathscr{A}(W^{\otimes 2}) \supsetneq 2\mathscr{A}(W)$. The thesis does the above for the two problems mentioned previously. There are a few other instances (open problems 5.1, 6.1, 8.3, 8.4 in [11]) where the same idea has not yet yielded concrete results.

In Chapter 2, we will exhibit a class of channels that demonstrate the sub-optimality of the superposition coding region for the three receiver broadcast channel with two degraded message set and in Chapter 3 we present a class of binary interference channels that show the sub-optimality of the Han–Kobayashi region for the interference channel.

□ **End of chapter.**

Chapter 2

3-receiver broadcast channel with two degraded message sets

A sender X, who has access to two independent messages (M_0, M_1), wishes to communicate M_0 reliably to three receivers, denoted by $Y, \hat{Y}$, and Z, and M_1 reliably to a subset of the receivers $Y, \hat{Y}$. The sender encodes the messages, uniformly distributed over sets $[1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$, into a sequence X^n, which is then transmitted over a discrete memoryless broadcast channel, $W^{\otimes n}(y, \hat{y}, z|x)$. Three receivers who receive sequences $Y^n, \hat{Y}^n, Z^n$, respectively, wish to decode messages as depicted in Figure 2.1.

The following achievable region is the straightforward extension of the superposition coding region for two receivers, Theorem 1, to the described scenario.

Theorem 6 (Superposition coding achievable region, Corollary 1 [19]). *The union of the set of rate pairs* (R_0, R_1) *satisfying*

$$R_0 \leq I(U; Z)$$

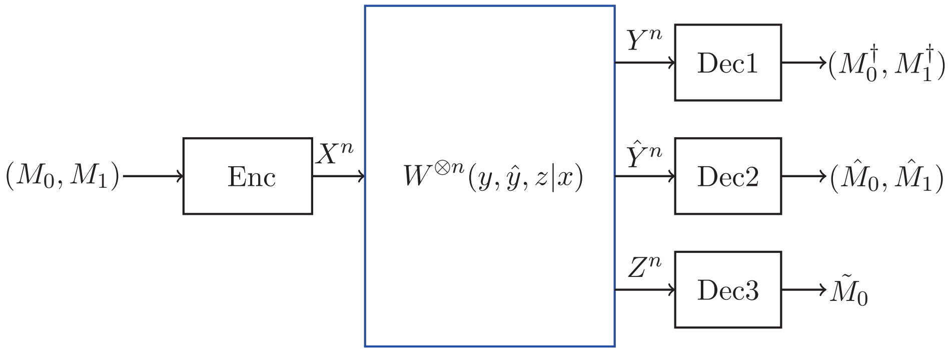


Figure 2.1: Three receiver broadcast channel with two degraded message sets

$$R_0 + R_1 \leq I(U;Z) + \min\{I(X;Y|U), I(X;\hat{Y}|U)\}$$

$$R_0 + R_1 \leq \min\{I(X;Y), I(X;\hat{Y})\}$$

where the union is taken over all pairs of random variables (U, X) *such that* $|\mathcal{U}| \leq |\mathcal{X}| + 2$ *and* $U \to X \to (Y, \hat{Y}, Z)$ *forms a Markov chain is achievable.*

Let $\mathscr{A}^{SC}(W)$ denote the above region for the channel W. It is immediate that $\mathscr{A}^{SC}(W)$ is convex and hence it is uniquely characterized by its supporting hyper-planes. In particular the set of $\mathscr{A}_\lambda^{SC}(W) := \max(\lambda R_0 + R_1)$, $(R_0, R_1) \in \mathscr{A}^{SC}(W)$, $\forall \lambda \geq 1$, uniquely determine $\mathscr{A}_\lambda^{SC}(W)$. We use the term *sum-rate* to refer to the above when $\lambda = 1$ and the term *weighted sum-rate* when talking about a generic λ.

Lemma 2. *Superposition coding region stated in Theorem 6 is asymptotically capacity achieving.*

Proof. This is a standard argument and the proof is presented only for completeness. By Fano's inequality and the data-processing inequality any sequence of coding strategies, such that the probability of error goes to zero, for the described setting satisfies

$$\begin{aligned} R_0 &\leq \frac{1}{n} I(M_0; Z^n) + \epsilon_n, \\ R_0 + R_1 &\leq \frac{1}{n}(I(M_0; Z^n) + I(X^n; Y^n|M_0)) + \epsilon_n, \\ R_0 + R_1 &\leq \frac{1}{n}(I(M_0; Z^n) + I(X^n; \hat{Y}^n|M_0)) + \epsilon_n, \\ R_0 + R_1 &\leq \frac{1}{n} I(X^n; Y^n) + \epsilon_n \\ R_0 + R_1 &\leq \frac{1}{n} I(X^n; \hat{Y}^n) + \epsilon_n \end{aligned}$$

for some ϵ_n that tends to zero as $n \to \infty$. Note that the codebook as well as the channel induces the distribution on $(M_0, M_1, X^n, Y^n, \hat{Y}^n, Z^n)$. Setting $U = M_0$ in (see Theorem 6 and its n-letter extension) shows that $\frac{1}{n}\mathscr{A}^{SC}(W^{\otimes n}) \to \mathscr{C}$ which completes the proof. □

In the next section we show that Theorem 6 can be strictly smaller than the capacity region.

2.1 Strict sub-optimality of the superposition coding achievable region

The example that shows the strict sub-optimality is a reversely degraded multi-level broadcast erasure channel, belonging to the class depicted in Figure 2.2. Each sub-channel is a binary erasure channel (BEC) with erasure probability

$$X_a \to Y_a : BEC(e_a),\ X_b \to Y_b : BEC(e_b)$$

$$X_a \to \hat{Y}_a : BEC(\hat{e}_a),\ X_b \to \hat{Y}_b : BEC(\hat{e}_b)$$

$$X_a \to Z_a : BEC(f_a),\ X_b \to Z_b : BEC(f_b).$$

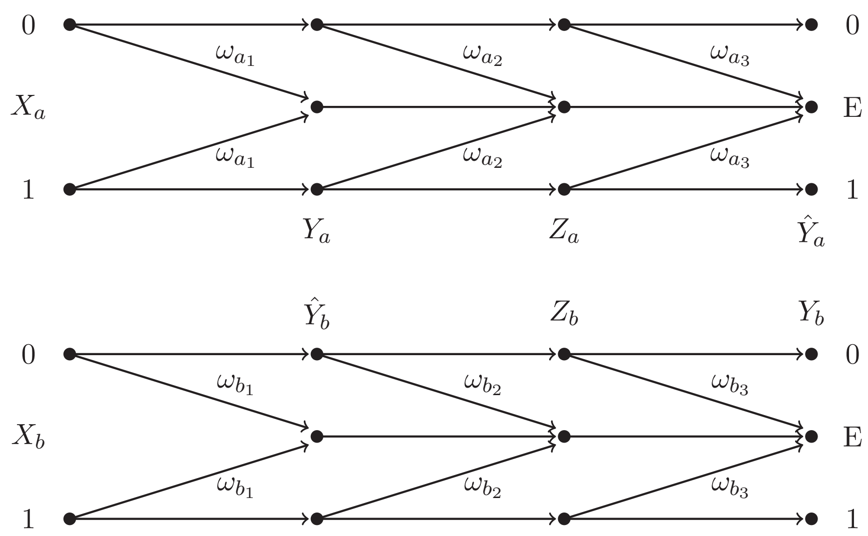


Figure 2.2: Product broadcast erasure channel

The order of channels in Figure 2.2 implies that $\hat{e}_a \geq f_a \geq e_a$ and $e_b \geq f_b \geq \hat{e}_b$. In particular the erasure probabilities of sub-channels are given by

$$e_a = \omega_{a_1},\ 1 - f_a = (1 - \omega_{a_1})(1 - \omega_{a_2}),\ 1 - \hat{e}_a = (1 - \omega_{a_1})(1 - \omega_{a_2})(1 - \omega_{a_3})$$

$$\hat{e}_b = \omega_{b_1},\ 1 - f_b = (1 - \omega_{b_1})(1 - \omega_{b_2}),\ 1 - e_b = (1 - \omega_{b_1})(1 - \omega_{b_2})(1 - \omega_{b_3}).$$

Proposition 1. *For the broadcast channel in Figure 2.2, it suffices to consider uniform distribution on $X = (X_a, X_b)$ to obtain the superposition coding region (Theorem 6).*

Proof. This is a symmetrization argument similar to the one in [18]. The argument is presented in a slightly more general fashion that is required for the proof of the proposition, in that the number of product components can be larger than 2.

Let π be either of the two permutations of $\{0, 1\}$. By an abuse of notation, let π also denote the induced permutation of $\{0, E, 1\}$ by mapping E to E. Then note for any generic symmetric erasure channel as in Figure 2.2, $W(Y = \pi(y)|X = \pi(x)) = W(Y = y|X = x)$. Now consider a product erasure channel structure where the inputs are $x_1, ..., x_k$ and let the corresponding outputs be $y_1, .., y_k$. Given any probability distribution $p(x_1, ..., x_k)$, let $p(y_1, .., y_k)$ denote the induced output distribution. Let $\pi_1, ..., \pi_k$ be any set of permutations of $\{0, 1\}$; and let $r(x_1, .., x_k) = p(\pi_1(x_1), ..., \pi_k(x_k))$ denote an induced input distribution. Then note that the induced output distribution is given by

$$\begin{aligned} r(y_1, ..., y_k) &= \sum_{x_1,..,x_k} r(x_1, .., x_k) \prod_{i=1}^{k} W_i(y_i|x_i) \\ &\overset{(a)}{=} \sum_{x_1,..,x_k} r(x_1, .., x_k) \prod_{i=1}^{k} W_i(\pi_i(y_i)|\pi_i(x_i)) \\ &= \sum_{x_1,..,x_k} p(\pi_1(x_1), .., \pi_k(x_k)) \prod_{i=1}^{k} W_i(\pi_i(y_i)|\pi_i(x_i)) \\ &= p(\pi_1(y_1), ..., \pi_1(y_k)). \end{aligned}$$

In the above, (a) follows from the symmetry of the component channels. Thus the output probability vector $r(\mathbf{y})$ is just a permutation of the original probability vector $p(\mathbf{y})$, and hence entropy of $Y_1, .., Y_k$ remains unchanged.

Given a joint distribution $p(u, x_1, .., x_k)$ (or on (U, X_a, X_b) as is this case), let Q denote a uniform random variable distributed over $[1 : 2^k]$. Identify with each Q a unique collection of permutations $\pi_1^q, .., \pi_k^q$. (for instance, using the binary representation). Define $\tilde{U} = (Q, U)$ and consider a joint distribution defined as follows:

$$r((q, u), x_1, .., x_k) = \frac{1}{2^k} p(u, \pi_1^q(x_1), ..., \pi_k^q(x_k)).$$

Note that the induced distributions on $(X_1, .., X_k)$ is uniform and that

$$r(x_1, .., x_k|(q, u)) = p(\pi_1^q(x_1), ..., \pi_k^q(x_k)|u)$$

$$r(y_1, .., y_k|(q, u)) = p(\pi_1^q(y_1), ..., \pi_k^q(y_k)|u)$$

where the second equality follows the argument presented earlier.

Hence note the following inequalities for any collection of outputs of symmetric channels

$$H_p(Y_1, .., Y_k) \overset{(a)}{\leq} H_r(Y_1, .., Y_k)$$

$$H_p(Y_1, .., Y_k|U) \overset{(b)}{=} H_r(Y_1, .., Y_k|U, Q)$$

$$H_p(Y_1, .., Y_k|X_1, .., X_k, U) \overset{(b)}{=} H_r(Y_1, .., Y_k|X_1, ..., X_k, U, Q),$$

where (a) follows since uniform input distribution maximizes entropy for symmetric erasure channels, and equalities denoted by (b) is due to the fact that permutations of probability vectors do not change their entropies. Thus every term occurring in the superposition coding region is non-decreasing by virtue of this symmetrization using Q, which induces a uniform distribution on X. □

The following corollary is immediate.

Corollary 2. *Superposition coding region for the product broadcast erasure channel in Fig 2.2 is the intersection of* $\{(R_0, R_1)|R_0 + R_1 < \min(c_Y, c_{\hat{Y}})\}$ *and the region* $\mathscr{S}$ *defined as the union of the set of rate pairs* (R_0, R_1) *satisfying*

$$R_0 < I(U; Z)$$

$$R_0 + R_1 < I(U; Z) + \min\{I(X; Y|U), I(X; \hat{Y}|U)\}$$

where the union is taken over all pairs of random variables (U, X) *such that* $|\mathcal{U}| \leq |\mathcal{X}| + 2$, $U \to X \to (Y, \hat{Y}, Z)$ *forms a Markov chain, and* $X = (X_a, X_b)$ *is* *uniformly distributed.* $c_Y = (1-e_a)+(1-e_b)$ *and* $c_{\hat{Y}} = (1-\hat{e}_a)+(1-\hat{e}_b)$ *are the capacities for channels* $W(y|x)$ *and* $W(\hat{y}|x)$*.*

Thus the key difficulty in computation of the superposition coding region is reduced to computation of region $\mathscr{S}$.

Proposition 2. For any $\lambda > 1$*,* $\mathscr{S}_\lambda := \max(\lambda R_0 + R_1),\ (R_0, R_1) \in \mathscr{S}$ *is given by*

$$\mathscr{S}_\lambda = \lambda c_Z + \min_{\alpha\in[0,1]} \max_{p(x)} \big(\alpha I(X;Y) + \bar{\alpha} I(X;\hat{Y}) - \lambda I(X;Z)\big),$$

where c_Z *is the capacity for channel* $W(z|x)$*.*

Proof. We know that it suffices to consider X to be uniformly distributed. Thus $\max_{\mathscr{S}}(\lambda R_0 + R_1)$ is given by

$$\max_{p(u|x)} \left(\lambda I(U;Z) + \min\{I(X;Y|U), I(X;\hat{Y}|U)\}\right),$$

where X is uniform.

An immediate application of min-max result, Corollary 2 in [12], yields that

$$\begin{aligned}&\max_{p(u|x)} \left(\lambda I(U;Z) + \min\left(I(X;Y|U), I(X;\hat{Y}|U)\right)\right) \\ &= \min_{\alpha\in[0,1]} \max_{p(u|x)} \left(\lambda I(U;Z) + \alpha I(X;Y|U) + (1-\alpha) I(X;\hat{Y}|U)\right).\end{aligned}$$

Noting that $I(U;Z) = I(X;Z) - I(X;Z|U) = c_Z - I(X;Z|U)$ (since uniform X achieves c_Z), we re-write the above as

$$\begin{aligned}&\min_{\alpha\in[0,1]} \max_{p(u|x)} \left(\lambda I(U;Z) + \alpha I(X;Y|U) + (1-\alpha) I(X;\hat{Y}|U)\right) \\ =&\min_{\alpha\in[0,1]} \max_{p(u|x)} \left(\lambda c_Z + \alpha I(X;Y|U) + (1-\alpha) I(X;\hat{Y}|U) - \lambda I(X;Z|U)\right) \\ =&\lambda c_Z + \min_{\alpha\in[0,1]} \max_{p(x)} \left(\alpha I(X;Y) + \bar{\alpha} I(X;\hat{Y}) - \lambda I(X;Z)\right).\end{aligned}$$

The non-trivial part of the last equality follows by applying the symmetrization argument to $p(x)$, the distribution that maximizes the quantity $\alpha I(X;Y) + \bar{\alpha} I(X;\hat{Y}) - \lambda I(X;Z)$. □

Remark 5. Note that the above two propositions regarding computation of superposition coding region for product broadcast channels apply to all symmetric (appropriately defined) channels and do not depend on the fact that the symmetric channel under consideration is an erasure channel. The next lemma on the other hand uses the erasure nature of the component channels.

Lemma 3. *Consider a product erasure channel mapping* $X_1, .., X_k$ *to* $Y_1, .., Y_k$ *with erasure probabilities* $\epsilon_1, ..., \epsilon_k$. *Then*

$$I(X_1, ..., X_k; Y_1, ..., Y_k) = \sum_{S \subseteq [1:k]} \left(\prod_{i \in S} (1 - \epsilon_i) \prod_{j \notin S} \epsilon_j \right) H(X_S),$$

where $X_S = (X_i : i \in S)$.

Proof. This is proved by induction on k. Observe that $k = 1$ is immediate. Note that $I(X_1, ..., X_k; Y_1, ..., Y_k) = I(X_1, ..., X_{k-1}; Y_1, ..., Y_{k-1}) + I(X_k; Y_k | Y_1, , , Y_{k-1})$. A simple calculation yields that

$$\begin{aligned} &I(X_k; Y_k | Y_1, ..., Y_{k-1}) \\ &\quad = \sum_{S_1 \subseteq [1:k-1]} \left(\prod_{i \in S_1} (1 - \epsilon_i) \prod_{j \notin S_1} \epsilon_j \right) (1 - \epsilon_k) H(X_k | X_{S_1}). \end{aligned}$$

Combining this term with induction hypothesis completes the proof. □

Remark 6. Combining Proposition 2 with Lemma 3 shows that computation of the superposition coding region for a product erasure broadcast channel reduces to computation of the maximum of a linear combination of entropic-vectors, a subset of $\mathbb{R}^{2^k - 1}$ generated by subsets of k binary random variables. When $k = 2$, for every $\alpha \in [0, 1]$, we wish to maximize a linear combination of the vector $[H(X_a), H(X_b), H(X_a, X_b)]$, where the coefficients are determined using Proposition 2 and Lemma 3. Note that X_a and X_b are binary random variables.

2.1.1 A specific example

There are many examples where two-letter superposition coding region beats the single-letter superposition coding region. However, below we produce a concrete example where using the machinery developed above we are able to explicitly demonstrate the gap between single-letter and 2-letter regions.

Theorem 7. For the reversely degraded three receiver product broadcast erasure channel as shown in Figure 2.2 with parameters

$$e_a = 1/2 \qquad \hat{e}_a = 1 \qquad f_a = 17/22$$

$$e_b = 1/2 \qquad \hat{e}_b = 0 \qquad f_b = 9/34$$

the following holds:

(i) *the non-trivial boundary (i.e. excluding the axes) of the superposition coding region is determined by the two lines:*

$$R_0 + R_1 = 1 \ \ \textit{and} \ \ \frac{11}{10}R_0 + R_1 = \frac{18}{17}.$$

(ii) *the non-trivial boundary of the 2-letter superposition coding region is determined by the two lines:*

$$R_0 + R_1 = 1 \ \ \textit{and} \ \ \frac{484}{435}R_0 + R_1 = \frac{528}{493}.$$

Proof. From Corollary 2, and $c_Y = c_{\hat{Y}} = 1$, the line $R_0 + R_1 = 1$ is immediate. To compute the superposition coding region (single-letter or 2-letter) it remains to compute the region $\mathscr{S}$.

Proof of (i), i.e. *Computation of the single-letter region.*

For the single-letter superposition coding region, we first show that any $(R_0, R_1) \in \mathscr{S}$ satisfies

$$\frac{11}{10}R_0 + R_1 \leq \frac{18}{17}.$$

Since $c_Z = (1 - f_a) + (1 - f_b) = \frac{180}{187}$, from Proposition 2 (taking $\alpha = \frac{1}{2}$), the inequality above will follow if we show that

$$\frac{1}{2}I(X;Y) + \frac{1}{2}I(X;\hat{Y}) - \frac{11}{10}I(X;Z) \leq 0 \ \ \forall p(x).$$

Here $X = (X_a, X_b), Y = (Y_a, Y_b), \hat{Y} = (\hat{Y}_a, \hat{Y}_b)$ and $Z = (Z_a, Z_b)$. Expanding the left hand side using Lemma 3 and substituting our choices of erasures yields

$$\frac{1}{2}I(X;Y) + \frac{1}{2}I(X;\hat{Y}) - \frac{11}{10}I(X;Z) = -\frac{1}{17}H(X_b|X_a),$$

implying the upper bound.

Next we show that the intersection of the two lines $R_0 + R_1 = 1$ and $\frac{11}{10}R_0 + R_1 = \frac{18}{17}$ belongs to the superposition coding region (completing the characterization).

Let U be a ternary random variable such that $\mathrm{P}(U = 0) = 13/34$, $\mathrm{P}(U = 1) = 7/34$, $\mathrm{P}(U = 2) = 14/34$. Conditionals are given by:

$$(X_a, X_b)|(U = 0) = (0, 0)$$

$$(X_a, X_b)|(U = 1) = (M, 0)$$

$$(X_a, X_b)|(U = 2) = (M, M),$$

where M is an unbiased binary random variable. Let Q be a random variable that symmetrizes the distribution of X (in the sense of the proof of Proposition 1) and let $\tilde{U} = (U, Q)$. Substituting $(\tilde{U}, X)$ into Theorem 6 yields:

$$R_0 \leq I(\tilde{U}; Z_a, Z_b) = \frac{10}{17}$$

$$R_0 + R_1 \leq \min\{I(X_a, X_b; Y|\tilde{U}), I(X_a, X_b; \hat{Y}_a, \hat{Y}_b|\tilde{U})\} + I(\tilde{U}; Z_a, Z_b) = 1$$

$$R_0 + R_1 \leq \min\{I(X_a, X_b; Y), I(X_a, X_b; \hat{Y}_a, \hat{Y}_b)\} = 1.$$

Thus $(R_0, R_1) = (\frac{10}{17}, \frac{7}{17})$ lying at the intersection of the two lines $R_0 + R_1 = 1$ and $\frac{11}{10}R_0 + R_1 = \frac{18}{17}$ belongs to the superposition coding region. This establishes the superposition coding region.

Proof of (ii), i.e. *Computation of the 2-letter region.*

The proof mimics the single-letter case. We show that any (R_0, R_1) belonging to the region $\mathscr{S}$ for the 2-letter channel satisfies

$$\frac{484}{435}R_0 + R_1 \leq \frac{528}{493}.$$

Since $c_Z = (1-f_a)+(1-f_b) = \frac{180}{187} = \frac{528\times435}{493\times484}$, from Proposition 2 (taking $\alpha = \frac{88}{174}$), the inequality above will follow if we show that

$$\frac{88}{174}I(X;Y) + \frac{86}{174}I(X;\hat{Y}) - \frac{484}{435}I(X;Z) \leq 0 \quad \forall p(x).$$

In the above, $X = (X_{a1}, X_{b1}, X_{a2}, X_{b2})$ and similarly for others. Expanding the left hand side using Lemma 3 and substituting our choices of erasures yields

$$\begin{aligned}
&-\frac{17}{174}I(X_{b1}; X_{b2}) - \frac{19}{2958}\big(I(X_{b1}; X_{b2}|X_{a1}) + I(X_{b1}; X_{b2}|X_{a2})\big)\\
&-\frac{2}{29}I(X_{a1}; X_{a2}|X_{b1}X_{b2}) - \frac{2543}{50286}I(X_{b1}; X_{b2}|X_{a1}X_{a2})\\
&-\frac{35}{493}\Big(H(X_{b1}|X_{a_1}X_{a2}X_{b2}) + H(X_{b2}|X_{a_1}X_{b1}X_{a2})\Big)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{174}\Big(I(X_{a1};X_{b1}|X_{a2})+I(X_{a1};X_{b2}|X_{a2})\\
&\qquad\qquad+I(X_{a2};X_{b1}|X_{a1})+I(X_{a2};X_{b2}|X_{a1})\Big)\\
&-\frac{1}{174}\Big(I(X_{a1};X_{b1}|X_{a2}X_{b2})+I(X_{a1};X_{b2}|X_{a2}X_{b1})\\
&\qquad\qquad+I(X_{a2};X_{b1}|X_{a1}X_{b2})+I(X_{a2};X_{b2}|X_{a1}X_{b1})\Big),
\end{aligned}
$$

which is term-by-term upper bounded by zero, implying the bound $\frac{484}{435}R_0+R_1\leq\frac{528}{493}$.

Let U be a ternary random variable such that $\mathrm{P}(U=0)=20/119$, $\mathrm{P}(U=1)=88/119$, $\mathrm{P}(U=2)=11/119$. Conditionals are given by:

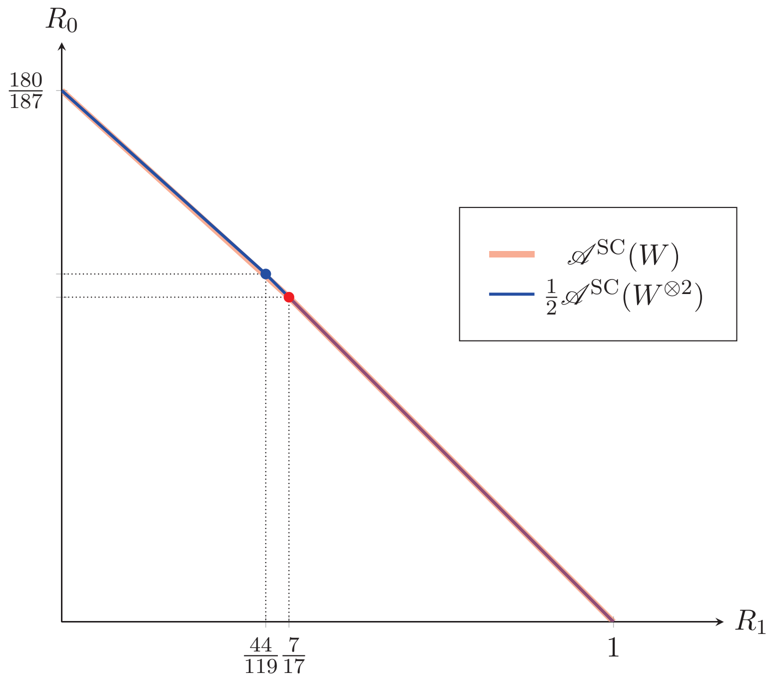
$$
\begin{aligned}
(X_{a1},X_{b1},X_{a2},X_{b2})|(U=0)&=(0,0,0,0)\\
(X_{a1},X_{b1},X_{a2},X_{b2})|(U=1)&=(M_1,M_1,M_1,0)\\
(X_{a1},X_{b1},X_{a2},X_{b2})|(U=2)&=(M_1,0,M_2,0),
\end{aligned}
$$

where M_1 and M_2 are two independent unbiased binary random variables. Let Q be a random variable that symmetrizes the distribution of X (in the sense of the proof of Proposition 1) and let $\tilde{U}=(U,Q)$. Substituting $(\tilde{U},X)$ into the normalized two-letter version of Theorem 6 yields:

$$
\begin{aligned}
R_0&\leq\frac{1}{2}I(\tilde{U};Z)=\frac{75}{119}\\
R_0+R_1&\leq\frac{1}{2}\Big(I(\tilde{U};Z)+\min\{I(X;Y|\tilde{U}),I(X;\hat{Y}|\tilde{U})\}\Big)=1\\
R_0+R_1&\leq\frac{1}{2}\min\{I(X;Y),I(X;\hat{Y}\}=1,
\end{aligned}
$$

where $X=(X_{a1},X_{b1},X_{a2},X_{b2})$ and similarly for others. Thus $(R_0,R_1)=(\frac{75}{119},\frac{44}{119})$ lying at the intersection of the two lines $R_0+R_1=1$ and $\frac{484}{435}R_0+R_1=\frac{528}{493}$ belongs to the two-letter superposition coding region. This establishes Theorem 7. $\square$

Figure 2.3 shows the single-letter and 2-letter superposition coding regions for this channel. The blue line-segments indicate the 2-letter superposition coding region and the red line-segments indicate the one-letter superposition coding region. The bottom plot zooms on the part where the gap between the two region is more visible, i.e., the intersection points of the two lines that constitute the single-letter and 2-letter superposition coding regions.



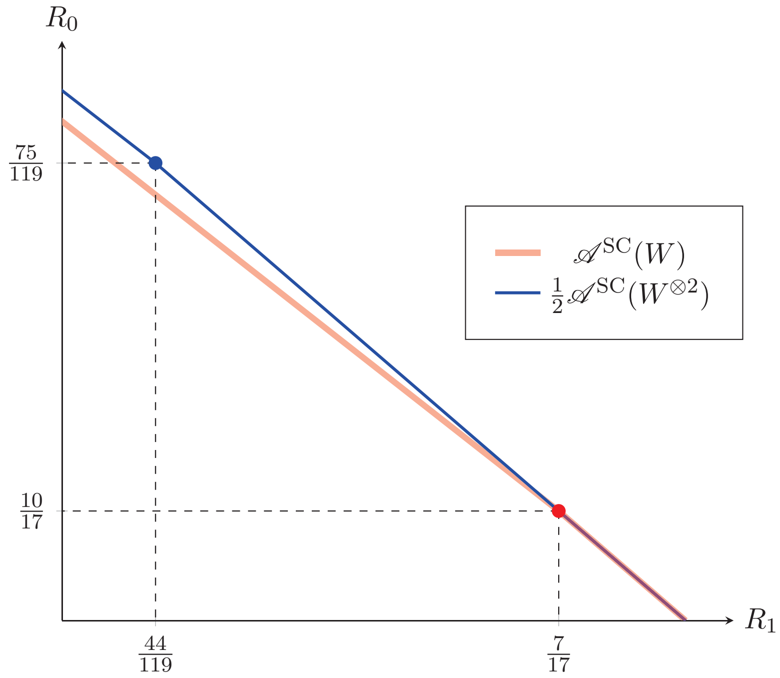


Figure 2.3: Plots (full and zoomed) of the single-letter and 2-letter superposition coding regions

2.2 Outer bound via concentration of mutual information over memoryless erasure channels

For any distribution $p(x_a^n, x_b^n)$ define

$$\mathcal{H}_n(k,l) = \frac{1}{\binom{n}{k}\binom{n}{l}} \sum_{S,T\subseteq[n]:|S|=k,|T|=l} H(X_{aS}, X_{bT}).$$

Lemma 4. The following inequalities hold:

(*i*) $\mathcal{H}_n(k,l) \leq \mathcal{H}_n(k+1,l) \leq \frac{k+1}{k}\mathcal{H}_n(k,l), \quad 1 \leq k \leq n-1.$

(*ii*) $\mathcal{H}_n(k,l) \leq \mathcal{H}_n(k,l+1) \leq \frac{l+1}{l}\mathcal{H}_n(k,l), \quad 1 \leq l \leq n-1.$

(*iii*) *(Concavity)* $\mathcal{H}_n(k-1,l) + \mathcal{H}_n(k+1,l) \leq 2\mathcal{H}_n(k,l).$

Proof. Clearly the (*ii*) can be obtained in a similar fashion as (*i*) by exchanging the coordinates. Hence it suffices to establish (*i*). Observe that for any $\mathcal{S}', T \subseteq$

$[n] : |\mathcal{S}'| = k+1, |\mathcal{T}| = l$, we have

$$(k+1)H(X_{a\mathcal{S}'}, X_{b\mathcal{T}}) = \sum_{i\in\mathcal{S}'} \left(H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) + H(X_{ai}|X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) \right),$$

where $\mathcal{S}'_i = \mathcal{S}' \setminus \{i\}$. Further,

$$\begin{aligned}\sum_{i\in\mathcal{S}'} H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) &\le \sum_{i\in\mathcal{S}'} \left(H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) + H(X_{ai}|X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) \right) \\ &\le H(X_{a\mathcal{S}'}, X_{b\mathcal{T}}) + \sum_{i\in\mathcal{S}'} H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}).\end{aligned}$$

Hence

$$\sum_{i\in\mathcal{S}'} H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}) \le (k+1)H(X_{a\mathcal{S}'}, X_{b\mathcal{T}}) \le \frac{k+1}{k} \sum_{i\in\mathcal{S}'} H(X_{a\mathcal{S}'_i}, X_{b\mathcal{T}}).$$

Summing over all $\mathcal{S}', \mathcal{T} \subseteq [n] : |\mathcal{S}'| = k+1, |\mathcal{T}| = l$ we obtain

$$\begin{aligned}(n-k) \sum_{\substack{\mathcal{S},\mathcal{T}\subseteq[n]: \\ |\mathcal{S}|=k \\ |\mathcal{T}|=l}} H(X_{a\mathcal{S}}, X_{b\mathcal{T}}) &\le (k+1) \sum_{\mathcal{S}',\mathcal{T}\subseteq[n]:|\mathcal{S}'|=k+1,|\mathcal{T}|=l} H(X_{a\mathcal{S}'}, X_{b\mathcal{T}}) \\ &\le \frac{k+1}{k}(n-k) \sum_{\substack{\mathcal{S},\mathcal{T}\subseteq[n]: \\ |\mathcal{S}|=k \\ |\mathcal{T}|=l}} H(X_{a\mathcal{S}}, X_{b\mathcal{T}}).\end{aligned}$$

Substituting we obtain,

$$\begin{aligned}(n-k)\binom{n}{k}\binom{n}{l}\mathcal{H}_n(k,l) &\le (k+1)\binom{n}{k+1}\binom{n}{l}\mathcal{H}_n(k,l) \\ &\le \frac{k+1}{k}(n-k)\binom{n}{k}\binom{n}{l}\mathcal{H}_n(k,l).\end{aligned}$$

This is equivalent to (i) after canceling the binomial coefficients.

To establish (iii) we start with the following (immediate) inequality, For any $\mathcal{S} \subset [n]$ and $i, j \in [n] \setminus \mathcal{S}$ we have

$$H(X_{a\mathcal{S}}, X_{b\mathcal{T}}) + H(X_{a\mathcal{S}^{i,j}}, X_{b\mathcal{T}}) \le H(X_{a\mathcal{S}^i}, X_{b\mathcal{T}}) + H(X_{a\mathcal{S}^j}, X_{b\mathcal{T}}),$$

where $\mathcal{S}^{i,j} = \mathcal{S} \cup \{i,j\}, \mathcal{S}^i = \mathcal{S} \cup \{i\}$, and $\mathcal{S}^j = \mathcal{S} \cup \{j\}$. Summing the above inequality over all $\mathcal{S}$ with $|\mathcal{S}| = k-1$, $\mathcal{T}$ with $|\mathcal{T}| = l$, and over all pairs $\{i,j\} \notin \mathcal{S}$ we obtain

$$\binom{n-k+1}{2}\binom{n}{k-1}\binom{n}{l}\mathcal{H}_n(k-1,l) + \binom{k+1}{2}\binom{n}{k+1}\binom{n}{l}\mathcal{H}_n(k+1,l)$$
$$\leq 2k\binom{n}{k}\binom{n}{l}\mathcal{H}_n(k,l).$$

This is equivalent to (iii) after canceling the binomial coefficients. □

Remark 7. The above Lemma can be considered as a minor generalization of the well-known Han's inequality or Shearer's lemma.

The following corollary is immediate from Lemma 4 by an induction argument, and hence its proof is omitted.

Corollary 3. *The following inequalities hold:*

(i) $\mathcal{H}_n(k,l) \leq \mathcal{H}_n(k+k_0, l+l_0) \leq \frac{k+k_0}{k}\frac{l+l_0}{l}\mathcal{H}_n(k,l)$

for $0 \leq k_0 \leq n-k, 0 \leq l_0 \leq n-l$.

(ii) $\frac{k-k_0}{k}\frac{l-l_0}{l}\mathcal{H}_n(k,l) \leq \mathcal{H}_n(k-k_0, l-l_0) \leq \mathcal{H}_n(k,l)$ *for* $0 \leq k_0 \leq k, 0 \leq l_0 \leq l$.

(iii) *(Concavity)* $\frac{m}{n}\mathcal{H}_n(k_1,l) + \frac{n-m}{n}\mathcal{H}_n(k_2,l) \leq \mathcal{H}_n\left(\frac{mk_1+(n-m)k_2}{n}, l\right)$

for $0 \leq m, k_1, k_2, l \leq n$.

Proposition 3 (Concentration of mutual information over memoryless product erasure channel). *Consider a product erasure channel,* $W_a(y_a|x_a) \otimes W_b(y_b|x_b)$, *mapping* X_a, X_b *to* Y_a, Y_b *with erasure probabilities* ϵ_a, ϵ_b, *respectively.*
Let the channel from (X_a^n, X_b^n) *to* (Y_a^n, Y_b^n) *be defined according to* $\prod_{i=1}^n W_a(y_{ai}|x_{ai})W_b(y_{bi}|x_{bi})$. *Then*

$$I(X_a^n, X_b^n; Y_a^n, Y_b^n) = \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor) + O\left(\sqrt{n\log n}\right).$$

Proof. In the proof, we will assume that $0 < \epsilon_a, \epsilon_b < 1$. The boundary cases are easier and the reader can check that the proofs follow similarly. Using Lemma 3 we have

$$\begin{aligned}
I(X_a^n, X_b^n; Y_a^n, Y_b^n) &= \sum_{\mathcal{S},\mathcal{T}\subseteq[n]} \left((1-\epsilon_a)^{|\mathcal{S}|}\epsilon_a^{n-|\mathcal{S}|}(1-\epsilon_b)^{|\mathcal{T}|}\epsilon_b^{n-|\mathcal{T}|}\right) H(X_{a\mathcal{S}}, X_{b\mathcal{T}}) \\
&= \sum_{0\le k,l\le n} \left((1-\epsilon_a)^{k}\epsilon_a^{n-k}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\right) \sum_{\substack{\mathcal{S},\mathcal{T}\subseteq[n]: \\ |\mathcal{S}|=k \\ |\mathcal{T}|=l}} H(X_{a\mathcal{S}}, X_{b\mathcal{T}}) \\
&= \sum_{0\le k,l\le n} \binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\mathcal{H}(k,l).
\end{aligned}$$

For any $|\mathcal{S}| = k, |\mathcal{T}| = l$, $0 \le H(X_{a\mathcal{S}}, X_{b\mathcal{T}}) \le \sum_{i\in\mathcal{S}} H(X_{ai}) + \sum_{j\in\mathcal{T}} H(X_{bj})$, implies

$$0 \le \mathcal{H}(k,l) \le k\log|\mathcal{X}_a| + l\log|\mathcal{X}_b| \le n\log|\mathcal{X}_a||\mathcal{X}_b|.$$

Hoeffding's inequality says that

$$\sum_{k\notin[n(1-p-\delta),n(1-p+\delta)]} \binom{n}{k}(1-p)^k p^{n-k} \le 2e^{-2\delta^2 n}.$$

Define $\mathcal{K} = \{k : n(1-\epsilon_a-\delta) \le k \le n(1-\epsilon_a+\delta), k \in \mathbb{N}\}$ and $\mathcal{L} = \{l : n(1-\epsilon_b-\delta) \le k \le n(1-\epsilon_b+\delta), l \in \mathbb{N}\}$.

Observe that

$$\begin{aligned}
&\left(\sum_{0\le k,l\le n}\binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\mathcal{H}(k,l)\right) - \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor) \\
&\le \sum_{0\le k,l\le n}\binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\left|\mathcal{H}(k,l) - \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor)\right| \\
&= \sum_{k\notin\mathcal{K}}\binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\left(\sum_{0\le l\le n}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\left|\mathcal{H}(k,l) - \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor)\right|\right) \\
&\quad + \sum_{l\notin\mathcal{L}}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\left(\sum_{k\in\mathcal{K}}\binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\left|\mathcal{H}(k,l) - \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor)\right|\right) \\
&\quad + \sum_{l\in\mathcal{L},k\in\mathcal{K}}\binom{n}{k}(1-\epsilon_a)^{k}\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^{l}\epsilon_b^{n-l}\left|\mathcal{H}(k,l) - \mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor, \lfloor n(1-\epsilon_b)\rfloor)\right|
\end{aligned}$$

$$\leq 2ne^{-2\delta^2 n}\log|\mathcal{X}_a||\mathcal{X}_b| + 2ne^{-2\delta^2 n}\log|\mathcal{X}_a||\mathcal{X}_b|$$
$$+\sum_{l\in\mathcal{L},k\in\mathcal{K}}\binom{n}{k}(1-\epsilon_a)^k\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^l\epsilon_b^{n-l}\left|\mathcal{H}(k,l)-\mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor,\lfloor n(1-\epsilon_b)\rfloor)\right|.$$

The last inequality follows from Hoeffding's inequality and from the uniform bound on $\mathcal{H}(k,l)$.

From Corollary 3 it follows that for $k\in\mathcal{K}, l\in\mathcal{L}$

$$\begin{aligned}
&|\mathcal{H}(k,l)-\mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor,\lfloor n(1-\epsilon_b)\rfloor)| \\
\leq&\mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor,\lfloor n(1-\epsilon_b)\rfloor)\left(\frac{\delta}{1-\epsilon_a}+\frac{\delta}{1-\epsilon_b}+\frac{\delta^2}{(1-\epsilon_a)(1-\epsilon_b)}\right) \\
\leq&2n\left(\frac{\delta}{1-\epsilon_a}+\frac{\delta}{1-\epsilon_b}+\frac{\delta^2}{(1-\epsilon_a)(1-\epsilon_b)}\right)\log|\mathcal{X}_a||\mathcal{X}_b|.
\end{aligned}$$

Therefore we obtain

$$\begin{aligned}
&\sum_{0\leq k,l\leq n}\binom{n}{k}(1-\epsilon_a)^k\epsilon_a^{n-k}\binom{n}{l}(1-\epsilon_b)^l\epsilon_b^{n-l}\mathcal{H}(k,l) \\
&\leq 2ne^{-2\delta^2 n}\log|\mathcal{X}_a||\mathcal{X}_b| + 2ne^{-2\delta^2 n}\log|\mathcal{X}_a||\mathcal{X}_b| \\
&\quad+2n\left(\frac{\delta}{1-\epsilon_a}+\frac{\delta}{1-\epsilon_b}+\frac{\delta^2}{(1-\epsilon_a)(1-\epsilon_b)}\right)\log|\mathcal{X}_a||\mathcal{X}_b| \\
&\quad+\mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor,\lfloor n(1-\epsilon_b)\rfloor).
\end{aligned}$$

Taking $\delta=\sqrt{\frac{\log n}{n}}$ and putting things together, we obtain

$$I(X_a^n,X_b^n;Y_a^n,Y_b^n)=\mathcal{H}(\lfloor n(1-\epsilon_a)\rfloor,\lfloor n(1-\epsilon_b)\rfloor)+O\left(\sqrt{n\log n}\right)$$

thus establishing the proposition. □

Theorem 8 (Outer bound). *Any achievable rate pair (R_0,R_1) for the channel depicted in Figure 2.2 with parameters*

$$e_a=1/2 \qquad \hat{e}_a=1 \qquad f_a=17/22$$

$$e_b = 1/2 \qquad \hat{e}_b = 0 \qquad f_b = 9/34,$$

must satisfy the constraints

$$R_0 + R_1 \leq 1 \quad and \quad \frac{187}{160}R_0 + R_1 \leq \frac{18}{16}.$$

Proof. Since $c_Y = c_{\hat{Y}} = 1$, the line $R_0 + R_1 \leq 1$ is immediate. The non-trivial and the interesting part of the proof is the fact that by analyzing the limiting n-letter superposition coding region which is asymptotically capacity achieving we are able to derive the constraint

$$\frac{187}{160}R_0 + R_1 \leq \frac{18}{16}.$$

Similar to the computation of the single-letter and 2-letter superposition coding regions, from Proposition 2 (taking $\alpha = \frac{85}{160}$), the inequality above will follow if we show that

$$\limsup_n \max_{p(x_a^n, x_b^n)} \frac{1}{n}\Bigg(\frac{85}{160} I(X_a^n, X_b^n; Y_a^n, Y_b^n) + \frac{75}{160} I(X_a^n, X_b^n; \hat{Y}_a^n, \hat{Y}_b^n) - \frac{187}{160} I(X_a^n, X_b^n; Z_a^n, Z_b^n) \Bigg) \leq 0.$$

From Proposition 3 it suffices to show that:

$$\limsup_n \max_{p(x_a^n, x_b^n)} \frac{1}{n}\Bigg(\frac{85}{160}\mathcal{H}(n(1-e_a), n(1-\epsilon_b)) + \frac{75}{160}\mathcal{H}(n(1-\hat{e}_a), n(1-\hat{e}_b)) - \frac{187}{160}\mathcal{H}(n(1-f_a), n(1-f_b)) \Bigg) \leq 0.$$

Substituting for the parameters, we wish to show that

$$\limsup_n \max_{p(x_a^n, x_b^n)} \frac{1}{n}\left(\frac{85}{160}\mathcal{H}(\frac{n}{2}, \frac{n}{2}) + \frac{75}{160}\mathcal{H}(0, n) - \frac{187}{160}\mathcal{H}(\frac{5n}{22}, \frac{25n}{34}) \right) \leq 0.$$

From Corollary 3 we have the following inequalities:

$$
\begin{aligned}
&\mathcal{H}_n(\frac{n}{2},\frac{n}{2}) \leq \mathcal{H}_n(\frac{n}{2},\frac{25n}{34}),\\
&\frac{5}{11}\mathcal{H}_n(\frac{n}{2},\frac{25n}{34}) + \frac{6}{11}\mathcal{H}_n(0,\frac{25n}{34}) \leq \mathcal{H}(\frac{5n}{22},\frac{25n}{34}),\\
&\frac{8}{17}\mathcal{H}_n(0,n) + \frac{9}{17}\mathcal{H}_n(0,\frac{n}{2}) \leq \mathcal{H}_n(0,\frac{25n}{34}),\\
&\frac{17}{25}\mathcal{H}_n(0,\frac{25n}{34}) \leq \mathcal{H}_n(0,\frac{n}{2}).
\end{aligned}
$$

Multiplying the first inequality by 85, the second by 187, the third by $\frac{75\times 17}{8}$, and the fourth by $\frac{27\times 25}{8}$ and adding together we obtain

$$85\mathcal{H}_n(\frac{n}{2},\frac{n}{2}) + 75\mathcal{H}_n(0,n) \leq 187\mathcal{H}(\frac{5n}{22},\frac{25n}{34}),$$

establishing the upper bound.

□

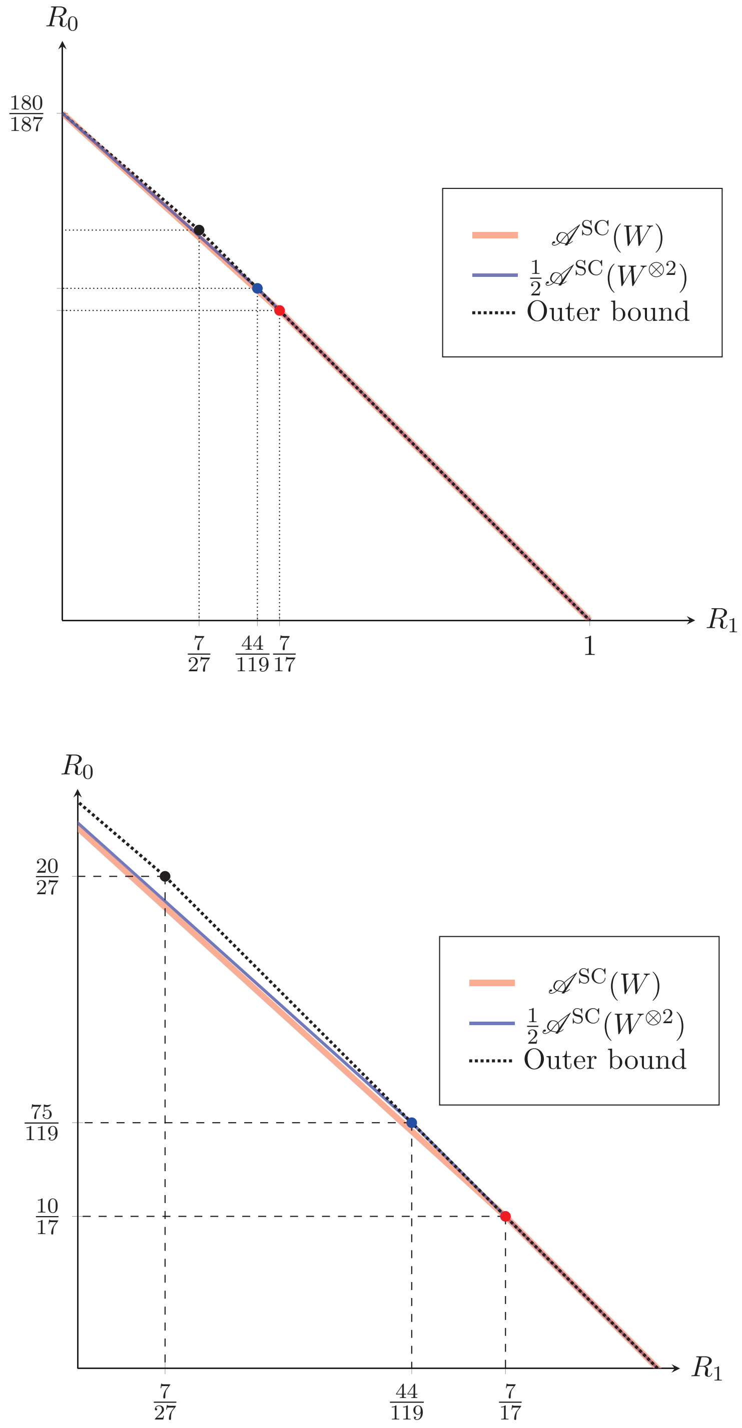


Figure 2.4: Plots (full and zoomed) of the single-letter, 2-letter superposition coding regions and the (new) outer bound

2.3 Summary and Discussion

The strict sub-optimality of the superposition coding region for the message setting in Figure 2.1 is shown by demonstrating a channel for which the 2-letter superposition coding region strictly outperforms the single-letter region. For the specific example, we are able to completely characterize the 2-letter superposition coding region. Notice that the optimizing distributions (that yield the non-trivial corner point along the $R_0 + R_1 = 1$ line in the counterexample) have optimizers where the same X is transmitted across the parallel channels. Further, in the 2-letter scheme, the same X is even transmitted across two consecutive-time slots, for some choices of U. This shows that superposition coding does not fully exploit the spatial and temporal diversity provided by the different channels to Y and $\hat{Y}$. Hidden in the optimizers are some hints as to how best to exploit the missing diversity gains.

The computation of higher-letter superposition coding regions boils down to finding choices of λ and α that make a certain linear combination of entropies of subsets of binary random variables negative for all probability distributions. Secondly, using the sub-modularity of entropy and the idea behind testing Shannon-type inequalities, one can get upper bounds on the critical λ, the slope of the capacity region around $(R_0, R_1) = (c_Z, 0)$. Even in the two-letter case (where the four variables $X_{a1}, X_{b1}, X_{a2}, X_{b2}$ involved are binary) restricting oneself to Shannon-type inequalities and maximizing the linear combination could have resulted in an non-entropic extreme point [28]. This would have led to an outer bound to the 2-letter superposition coding region. Luckily for us, the extreme point in the region calculations using sub-modularity (Shannon-type) constraints turns out to be achievable; thus yielding a precise characterization of the 2-letter superposition coding region. One interesting question that is worth pursuing is whether this phenomenon continues to hold for higher-letter computations as well.

A very interesting observation is that the optimizing code for the 2-letter region is a (simple) linear code on block of length 2. This demonstrates that linear block codes outperform memoryless codes in this setting. It is worth investigating the performance limits of linear codes for this setting.

Finally, another important contribution of our analysis that opens up potential avenues for research is that we obtained an explicit outer bound for the capacity region, not from a single-letter expression but analyzing limiting multi-letter achievable regions. This opens up potentially different ways of obtaining computable outer bounds.

□ **End of chapter.**

Chapter 3

Interference Channel

3.1 Strict sub-optimality of the Han–Kobayashi achievable region

In this section we show the sub-optimality of the Han–Kobayashi achievable region for the interference channel. For completeness we recall the Han–Kobayashi achievable region.

[Theorem 6.4 in [11], Han–Kobayashi achievable region] *A rate-pair* (R_1, R_2) *is achievable for the channel,* W*, described in Figure 1.2 if*

$$R_1 < I(X_1; Y_1|U_2, Q), \tag{3.1}$$

$$R_2 < I(X_2; Y_2|U_1, Q), \tag{3.2}$$

$$R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q), \tag{3.3}$$

$$R_1 + R_2 < I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q), \tag{3.4}$$

$$R_1 + R_2 < I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q), \tag{3.5}$$

$$2R_1 + R_2 < I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q), \tag{3.6}$$

$$R_1 + 2R_2 < I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q) \tag{3.7}$$

for some pmf $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$, *where* $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 4$, $|\mathcal{U}_2| \leq |\mathcal{X}_2| + 4$, *and* $|\mathcal{Q}| \leq 7$.

Lemma 5. *The Han–Kobayashi achievable region is asymptotically capacity achieving.*

Proof. Since $\frac{1}{n}\mathscr{A}^{\text{HK}}(W^{\otimes n}) \subseteq \mathscr{C}(W)\ \forall n$, it suffices to show the other direction. We will in particular show that $\frac{1}{n}\mathscr{A}^{\text{TIN}}(W^{\otimes n}) \to \mathscr{C}(W)$ and this suffices due to Corollary 1. By Fano's inequality and data-processing inequality any sequence of good codebooks for an interference channel satisfies,

$$R_1 \leq \frac{1}{n}I(M_1; Y_1^n) + \epsilon_n \leq \frac{1}{n}I(X_1^n; Y_1^n) + \epsilon_n,\ R_2\frac{1}{n}I(M_2; Y_12^n) + \epsilon_n \leq \frac{1}{n}I(X_2^n; Y_2^n) + \epsilon_n$$

for some ϵ_n that tends to zero $n \to \infty$. This shows that $\frac{1}{n}\mathscr{A}^{\text{TIN}}(W^{\otimes n}) \to \mathscr{C}(W)$, completing the proof. □

This justifies the use of Lemma 1 to test the optimality of the Han–Kobayashi region. As discussed in Chapter 1, to conduct the optimality test, one needs to evaluate the region for some non-trivial channels. But even numerical computation of the Han–Kobayashi region is computationally infeasible for random channels with binary inputs. This is essentially due to the non-convexity of the underlying optimization problem. One strategy is to consider interference channels where the computation becomes tractable. Along these lines, the following class was defined in [16].

Definition 3. An interference channel $W(y_1, y_2|x_1, x_2)$ is said to have *very weak* *interference* if

$$I(U_1; Y_2|X_2) \leq I(U_1; Y_1),$$

$$I(U_2; Y_1|X_1) \leq I(U_2; Y_2).$$

for all auxiliaries (U_1, U_2) such that the joint probability distribution satisfies $p(u_1, u_2, x_1, x_2, y_1, y_2) = p_1(u_1, x_1)p_2(u_2, x_2)p(y_1, y_2|x_1, x_2)$.

By considering (3.5) and the very weak interference conditions, it can be readily shown that for this class the sum-rate of the Han–Kobayashi region reduces to the treating interference as noise,

$$\max(R_1 + R_2) = \max_{\substack{p(x_1)\\p(x_2)}} I(X_1; Y_1) + I(X_2; Y_2), \quad (R_1, R_2) \in \text{Han–Kobayashi region.}$$

This simplification allowed the authors of [16] to test whether the the Han–Kobayashi region is sum-rate optimal or not for this class. Numerical simulations suggested that it is optimal and the authors could establish it for some classes within very weak interference channels. On the contrary the weighted sum rate of the Han–Kobayashi region does not simplify readily under the very weak interference assumptions and its computation remains infeasible.

Here we define a further sub-class of the very weak interference channels where the computation of the entire Han–Kobayashi region becomes viable.

Definition 4. Let clean Z interference channel (CZIC) be the class of interference channels where one of the sender–receiver pairs has a clean communication link. In this thesis it is assumed that the second sender-receiver pair has this property, i.e. $Y_2 = X_2$. This setting is depicted in Figure 3.1.

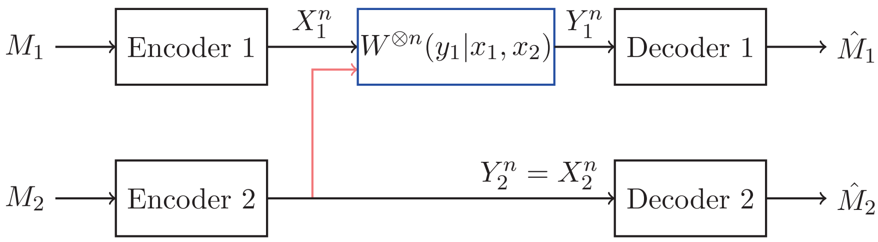


Figure 3.1: CZIC - Clean Z Interference Channel

The simplicity of this class allows us to simplify the characterization of the Han–Kobayashi region for CZIC. Proposition 3.12 shows an equivalent characterization of the Han–Kobayashi region for this class.

Proposition 4. *The Han–Kobayashi region for CZIC is identical to the set of rate pairs* (R_1, R_2) *that satisfy*

$$R_1 < I(X_1; Y_1|U_2, Q), \tag{3.8}$$

$$R_2 < H(X_2|Q), \tag{3.9}$$

$$R_1 + R_2 < I(X_1, U_2; Y_1|Q) + H(X_2|U_2, Q) \tag{3.10}$$

for some pmf $p(q)p(u_2|q)p(x_2|u_2)p(x_1|q)$, *where* $|\mathcal{U}_2| \leq |\mathcal{X}_2|$ *and* $|\mathcal{Q}| \leq 2$.

Proof. On one hand, it is a simple exercise to note that the Han–Kobayashi region for CZIC reduces to the three constraints above by setting $U_1 = \phi$. Hence, the above region is a subset of the Han–Kobayashi region.

Conversely, (3.8) is identical to (3.1) of the Han–Kobayashi region. (3.9) and (3.10) are respectively looser constraints than (3.2) and (3.3) of the Han–Kobayashi region, which makes the above region at least as large as the original Han–Kobayashi region, thus proving their equivalence.

Note that the changes in cardinality of U_2 and Q follow from standard applications of cardinality reduction techniques all while the underlying region remains the same, as shown in Appendix C of [11]. Therefore, we do not have to take these changes into account when talking about the two regions' equivalence. □

The time sharing random variable Q in the Han–Kobayashi region makes sure that this region is convex. Hence to investigate the Han–Kobayashi region in relation to the capacity region we consider the regions' supporting hyper-planes, that is

$$\begin{aligned} \mathscr{A}_\lambda^{\mathrm{HK}} &:= \max(\lambda R_1 + R_2), \quad \forall (R_1, R_2) \in \text{Han–Kobayashi achievable region} \\ \mathscr{C}_\lambda &:= \max(\lambda R_1 + R_2), \quad \forall (R_1, R_2) \in \text{capacity region} \end{aligned}$$

We consider two regimes for λ, $\lambda \in [0, 1]$ and $\lambda \in (1, \infty)$. First we show that the Han–Kobayashi region is optimal in the first regime.

Proposition 5. *For CZIC,* $\mathscr{A}_\lambda^{HK} = \mathscr{C}_\lambda$ *for* $\lambda \in [0, 1]$.

Proof. The proof is a standard converse argument. By Fano's inequality, any achievable $(2^{nR_1}, 2^{nR_2}, n)$ code must satisfy

$$n(\lambda R_1 + R_2) - n\epsilon \leq H(X_2^n) + \lambda I(X_1^n; Y_1^n)$$

The right hand side of the above inequality, by chain rule, equals to

$$\begin{aligned} &\sum_{i=1}^{n} H(X_{2i}|X_2^{i-1}) + \lambda I(X_1^n; Y_{1i}|Y_1^{i-1}) \\ \leq &\sum_{i=1}^{n} H(X_{2i}|X_2^{i-1}) + \lambda I(X_1^n, Y_1^{i-1}; Y_{1i}) \\ \overset{(a)}{\leq} &\sum_{i=1}^{n} H(X_{2i}|X_2^{i-1}) + \lambda I(X_{1i}, X_2^{i-1}; Y_{1i}) \\ \leq &\sum_{i=1}^{n} H(X_{2i}) - I(X_{2i}; X_2^{i-1}) + \lambda I(X_{1i}; Y_{1i}) + \lambda I(X_2^{i-1}; Y_{1i}|X_{1i}) \end{aligned}$$

$$\overset{(b)}{\leq} \sum_{i=1}^{n} H(X_{2i}) - I(X_{2i}; X_2^{i-1}) + \lambda I(X_{1i}; Y_{1i}) + \lambda I(X_2^{i-1}; X_{2i}) \tag{3.11}$$

where (a) follows from the Markov chain formed by $(Y_1^{i-1}, X_1^{n \setminus i}) \to \left(X_{1i}, X_2^{i-1}\right) \to Y_{1i}$ and (b) from the Markov chain formed by $X_2^{i-1} \to (X_{1i}, X_{2i}) \to Y_{1i}$ and the independence of X_1^n and X_2^n.

For $0 \leq \lambda \leq 1$, (3.11) is less than or equal to $n\,(\max(H(X_2) + \lambda I(X_1; Y_1))$ which shows that, as ϵ tends to zero, any achievable rate pair (R_1, R_2) must satisfy,

$$\lambda R_1 + R_2 \leq \max_{p_1(x_1)p_2(x_2)} H(X_2) + \lambda I(X_1; Y_1).$$

The proof is complete by observing that the above rate can be achieved by setting $U_2 = \phi$ in the Han–Kobayashi region. □

In the second regime, $\lambda \in (1, \infty)$, we show that $\mathscr{C}_\lambda$ can be strictly larger than $\mathscr{A}_\lambda^{\mathrm{HK}}$ which proves the sub-optimality of the Han–Kobayashi region for IC. The proof involves evaluating the Han–Kobayashi region for a particular CZIC. The following Proposition helps us evaluate $\mathscr{A}_\lambda^{\mathrm{HK}}$ for $\lambda > 1$.

Proposition 6. *For CZIC, for all $\lambda > 1$, $\mathscr{A}_\lambda^{HK}$ equals to*

$$\max_{\substack{p_1(x_1)\\ p_2(x_2)}} \left(I(X_1, X_2; Y_1) + \mathop{\mathfrak{C}}_{p_2(x_2)} \left[H(X_2) - I(X_2; Y_1 | X_1) + (\lambda - 1) I(X_1; Y_1) \right] \right). \tag{3.12}$$

Proof. Consider the constraints on $R_1 + R_2$ and R_1 as stated in Proposition 4. and observe that any (R_1, R_2) in the Han–Kobayashi region must satisfy the following inequality for some $p(q)p_2(u_2, x_2|q)p_1(x_1|q)$.

$$\mathscr{A}_\lambda^{\mathrm{HK}} \leq (\lambda - 1) I(X_1; Y_1 | U_2, Q) + I(X_1, U_2; Y_1 | Q) + H(X_2 | U_2, Q) \tag{3.13}$$

Rewrite the right hand side of the above as

$$\begin{aligned}
&I(X_1, X_2; Y_1|Q) + H(X_2|U_2, Q) - I(X_2; Y_1|U_2, X_1, Q) + (\lambda - 1)I(X_1; Y_1|U_2, Q) \\
&\overset{(a)}{=} I(X_1, X_2; Y_1|Q) + \underset{p_2(x_2|q)}{\mathfrak{C}}\big[H(X_2|Q) - I(X_2; Y_1|X_1, Q) + (\lambda - 1)I(X_1; Y_1|Q)\big] \\
&\overset{(b)}{\le} \max_{\substack{p_1(x_1)\\ p_2(x_2)}} \left(I(X_1, X_2; Y_1) + \underset{p_2(x_2)}{\mathfrak{C}}\big[H(X_2) - I(X_2; Y_1|X_1) + (\lambda - 1)I(X_1; Y_1)\big] \right),
\end{aligned}$$

where (a) follows directly from the definition of the upper concave envelope and (b) from the fact that Q computes an average, and the average is less than the maximum.

On the other hand, for any $p_2(u_2, x_2)p_1(x_1)$, the following rate pair

$$(R_1, R_2) = (I(X_1; Y_1|U_2),\ H(X_2|U_2) + I(U_2; Y_1))$$

belongs to the Han–Kobayashi region as it satisfies the constraints. Thus, $\mathscr{A}_\lambda^{\mathrm{HK}}$ is larger than or equal to

$$\begin{aligned}
&\max_{\substack{p_2(u_2,x_2)\\ p_1(x_1)}} \lambda I(X_1; Y_1|U_2) + H(X_2|U_2) + I(U_2; Y_1) \\
&= \max_{\substack{p_2(u_2,x_2)\\ p_1(x_1)}} I(X_1, U_2; Y_1) + H(X_2|U_2) + (\lambda - 1)I(X_1; Y_1|U_2) \\
&= \max_{\substack{p_2(u_2,x_2)\\ p_1(x_1)}} I(X_1, X_2; Y_1) + H(X_2|U_2) - I(X_2; Y_1|U_2, X_1) + (\lambda - 1)I(X_1; Y_1|U_2) \\
&\overset{(c)}{=} \max_{\substack{p_2(x_2)\\ p_1(x_1)}} \left(I(X_1, X_2; Y_1) + \underset{p_2(x_2)}{\mathfrak{C}}\big[H(X_2) - I(X_2; Y_1|X_1) + (\lambda - 1)I(X_1; Y_1)\big] \right),
\end{aligned}$$

where (c) also follows directly from the definition of the upper concave envelope, see [4]. This establishes the converse and completes the proof of the proposition.

□

The two following lemmas help us simplify the expression in Proposition 6.

Lemma 6. *The following holds for any $f(x)$ and $g(x)$.*

$$\underset{x}{\mathfrak{C}}[f(x)+g(x)] \leq \underset{x}{\mathfrak{C}}[f(x)] + \underset{x}{\mathfrak{C}}[g(x)].$$

Proof. $\underset{x}{\mathfrak{C}}[f(x)] + \underset{x}{\mathfrak{C}}[g(x)]$ is concave as it is the sum of two concave envelopes. On the other hand $\underset{x}{\mathfrak{C}}[f(x)]$ and $\underset{x}{\mathfrak{C}}[g(x)]$ are by definition respectively larger than or equal to $f(x)$ and $g(x)$. The proof is complete by noting that $\underset{x}{\mathfrak{C}}[f(x)+g(x)]$ is the smallest concave function that dominates $f(x)+g(x)$. □

Lemma 7. *Let $l(x)$ be an affine function of x. For any $f(x)$ we have*

$$\underset{x}{\mathfrak{C}}[f(x)+l(x)] = \underset{x}{\mathfrak{C}}[f(x)] + l(x).$$

Proof. Consider the Lemma 6 for the two functions $(f(x)+l(x))$ and $(-l(x))$.

$$\underset{x}{\mathfrak{C}}[(f(x)+l(x)-l(x)] \leq \underset{x}{\mathfrak{C}}[f(x)+l(x)] + \underset{x}{\mathfrak{C}}[-l(x)] = \underset{x}{\mathfrak{C}}[f(x)+l(x)] - l(x).$$

This establishes the non-trivial direction to prove the lemma. □

Corollary 4. *For CZIC, for all $\lambda > 1$, $\mathscr{A}_\lambda^{HK}$ equals to*

$$\mathscr{A}_\lambda^{HK} = \max_{\substack{p_1(x_1)\\ p_2(x_2)}} \left(H(Y_1) + \underset{p_2(x_2)}{\mathfrak{C}} \left[H(X_2) + (\lambda-1)H(Y_1) - \lambda H(Y_1|X_1)\right] \right). \quad (3.14)$$

Proof. From Proposition 6, $\mathscr{A}_\lambda^{\mathrm{HK}}$ is equal to

$$\max_{\substack{p_1(x_1)\\ p_2(x_2)}} \left(I(X_1,X_2;Y_1) + \underset{p_2(x_2)}{\mathfrak{C}} \left[H(X_2) - I(X_2;Y_1|X_1) + (\lambda-1)I(X_1;Y_1)\right] \right).$$

Note that $H(Y_1|X_1,X_2)$ is linear in $P_2(X_2)$, hence by Lemma 7

$$\begin{aligned} &\underset{p_2(x_2)}{\mathfrak{C}} \left[H(X_2) - I(X_2;Y_1|X_1) + (\lambda-1)I(X_1;Y_1)\right] \\ &= \underset{p_2(x_2)}{\mathfrak{C}} \left[H(X_2) + (\lambda-1)H(Y_1) - \lambda H(Y_1|X_1)\right] + H(Y_1|X_1,X_2), \end{aligned}$$

which establishes the Corollary. □

The main obstacle of computing $\mathscr{A}_\lambda^{\mathrm{HK}}$ lies in the evaluation of the concave envelope in (3.14). In general computing the concave envelope of multi-variate functions is a difficult task, however, for single-variate functions it's rather easy to be carried out with extremely high precision. Here we are mainly concerned with the function $\underset{p_2(x_2)}{\mathfrak{C}}[H(X_2)+(\lambda-1)H(Y_1)-\lambda H(Y_1|X_1)]$. It is easy to verify that this function has at most two two inflection points when X_2 is binary. Figure 3.2 shows one instance of the above function when the there are two inflection points.

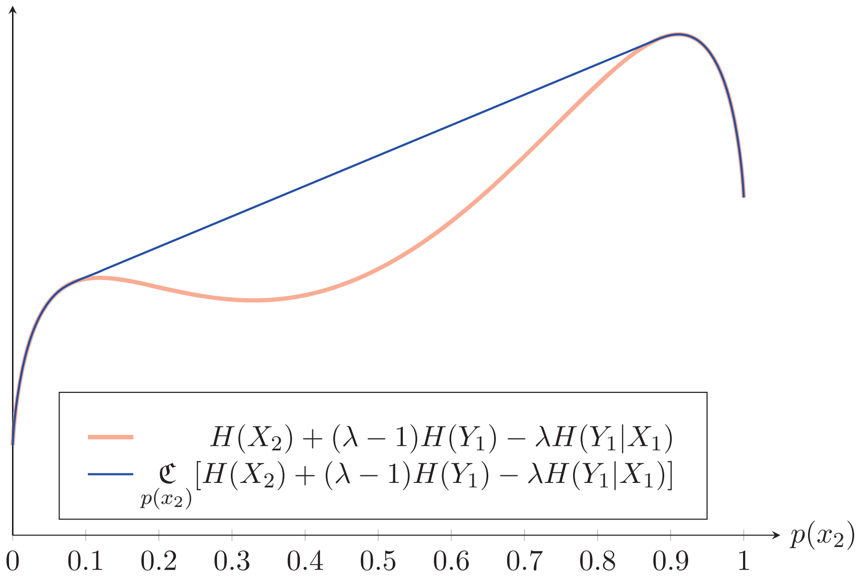


Figure 3.2: The shape of the concave envelope for binary CZIC

We consider $W(Y_1|X_1,X_2) \in$ CZIC with binary X_1, X_2 and Y_1 where we are able to numerically evaluate $\mathscr{A}_\lambda^{\mathrm{HK}}(W)$ with extremely high precision. Within this class of channels, it is rather easy to find Ws where

$$\mathscr{A}_\lambda^{\mathrm{HK}}(W) < \frac{1}{2}\mathscr{A}_\lambda^{\mathrm{TIN}}(W^{\otimes 2}), \quad \text{for some fixed } \lambda \in (1,\infty). \tag{3.15}$$

This implies that the Han–Kobayashi region is strictly sub-optimal for the interference channel, thus establishing Theorem 5. □

A number of examples where the capacity region is larger than the Han–Kobayashi region, i.e. inequality 3.15 holds, are listed in Table 3.1. The channels in Table 3.1 are represented using matrices of the form

$$\begin{bmatrix} P(Y_1 = 0|(X_1, X_2) = (0,0)) & P(Y_1 = 0|(X_1, X_2) = (0,1)) \\ P(Y_1 = 0|(X_1, X_2) = (1,0)) & P(Y_1 = 0|(X_1, X_2) = (1,1)) \end{bmatrix},$$

and the values of $\mathscr{A}_\lambda^{\mathrm{HK}}(W)$ and $\frac{1}{2}\mathscr{A}_\lambda^{\mathrm{TIN}}(W^{\otimes 2})$ are truncated to six decimal places.

λ	$W(Y_1\|X_1, X_2)$	$\mathscr{A}_\lambda^{\mathrm{HK}}(W)$	$\frac{1}{2}\mathscr{A}_\lambda^{\mathrm{TIN}}(W^{\otimes 2})$
2	$\begin{bmatrix} 1 & 0.5 \\ 1 & 0 \end{bmatrix}$	1.107516	1.108141
9	$\begin{bmatrix} 0.12 & 0.89 \\ 0.21 & 0.62 \end{bmatrix}$	1.074484	1.075544
12	$\begin{bmatrix} 0.01 & 0.58 \\ 0.20 & 0.74 \end{bmatrix}$	1.289830	1.293760
14	$\begin{bmatrix} 0.78 & 0.07 \\ 0.46 & 0.05 \end{bmatrix}$	1.426526	1.432419
15	$\begin{bmatrix} 0.91 & 0.22 \\ 0.66 & 0.15 \end{bmatrix}$	1.323766	1.339065
16	$\begin{bmatrix} 0.91 & 0.13 \\ 0.62 & 0.06 \end{bmatrix}$	1.515421	1.534724
18	$\begin{bmatrix} 0.38 & 0.87 \\ 0.12 & 0.79 \end{bmatrix}$	1.449959	1.468577

Table 3.1: Examples

Remark 8. Given the importance of this sub-optimality result to the field, it is important that we need to make sure that the gains are not the result of numerical inaccuracies (of computing the concave envelope and the global maximizers). To this end, we first identify a sub-class where the concave envelope is computed explicitly and since the resultant expressions are elementary functions we employ interval arithmetic[1] to give formal bounds on the global maximizers.

3.1.1 A specific sub-class

Definition 5. $S(c)$ is the class of CZIC with binary inputs and outputs where

$$\begin{bmatrix} P(Y_1 = 0|(X_1, X_2) = (0,0)) & P(Y_1 = 0|(X_1, X_2) = (0,1)) \\ P(Y_1 = 0|(X_1, X_2) = (1,0)) & P(Y_1 = 0|(X_1, X_2) = (1,1)) \end{bmatrix} = \begin{bmatrix} 1 & c \\ 1 & 0 \end{bmatrix},$$

This channel is depicted in Figure 3.3 as two point-to-point channels $X_1 \to Y_1$ for different choices of X_2.

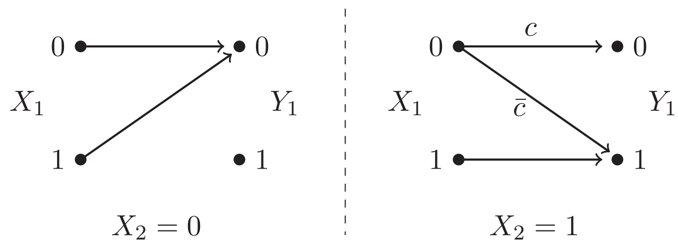


Figure 3.3: A special class of binary CZIC - $S(c)$

Consider the the function under the concave envelope operator in (3.14),

$$H(X_2) + (\lambda - 1)H(Y_1) - \lambda H(Y_1|X_1)$$

[1]See https://en.wikipedia.org/wiki/Interval_arithmetic for the use of interval arithmetic as a numerical tool.

For the channel $S(c)$ the above is equal to

$$f_{(\lambda,c)}(p,q) := (1-\lambda\bar{p})H_b(q) + (\lambda-1)H_b(q+pc\bar{q}) - \lambda p H_b(q+c\bar{q}). \tag{3.16}$$

where $p := P(X_1 = 0)$ and $q := P(X_2 = 0)$.

The following Lemma characterizes the concave envelope of $f_{(\lambda,c)}(p,q)$.

Lemma 8. Consider the bivarite function $f_{(\lambda,c)}(p,q)$, over $(p,q) \in [0,1]\times[0,1]$.

$$\underset{q}{\mathfrak{C}}[f_{(\lambda,c)}(p,q)] = \begin{cases} f_{(\lambda,c)}(p,q) & q \geq \max(0,\hat{q}) \\ \dfrac{f_{(\lambda,c)}(p,\hat{q}) - f_{(\lambda,c)}(p,0)}{\hat{q}}q + f_{(\lambda,c)}(p,0) & \textit{Otherwise} \end{cases},$$

where $\hat{q}$ is the solution of $\dfrac{f_{(\lambda,c)}(p,\hat{q}) - f_{(\lambda,c)}(p,0)}{\hat{q}} = \dfrac{\partial f_{(\lambda,c)}(p,q)}{\partial q}\Big|_{\hat{q}}$.

Proof. Consider the first and second partial derivatives of $f(p,q)$ w.r.t. q,

$$\frac{\partial f_{(\lambda,c)}(p,q)}{\partial q} = (1-\lambda\bar{p})J(q) + (\lambda-1)(1-pc)J(q+pc\bar{q}) - \lambda p\bar{c}J(q+c\bar{q}) \tag{3.17}$$

$$\frac{\partial^2 f_{(\lambda,c)}(p,q)}{\partial q^2} = \frac{pc\,(c(\lambda\bar{p}-1) - (c(\lambda\bar{p}-1)+1)q)}{\ln 2q\bar{q}(cp+(1-cp)q)(c+\bar{c}q)} \tag{3.18}$$

where $J(x) = \log(\dfrac{1-x}{x})$ is the derivative of the binary entropy function. Observe that

1. If $p \in (\frac{\lambda-1}{\lambda}, 1)$, then (3.18) is negative for $q \in (0,1)$, i.e., if $p > \frac{\lambda-1}{\lambda}$, $f_{(\lambda,c)}(p,q)$ is concave in q and $\underset{q}{\mathfrak{C}}[f_{(\lambda,c)}(p,q)] = f_{(\lambda,c)}(p,q)$.

2. If $p \in (0, \frac{\lambda-1}{\lambda})$, then (3.18) has one solution, $q^* \in (0,1)$.

$$q^* = \frac{c(\lambda\bar{p}-1)}{c(\lambda\bar{p}-1)+1}.$$

   In fact, $f_{(\lambda,c)}(p,q)$ is convex for $q \in (0,q^*)$ and concave for $q \in (q^*,1)$. Thus $\underset{q}{\mathfrak{C}}[f_{(\lambda,c)}(p,q)]$ consists of two parts. First part is a tangent line from the point $f_{(\lambda,c)}(p,0)$ to the function $f_{(\lambda,c)}(p,\hat{q})$ and the second part is equal to $f_{(\lambda,c)}(p,q)$. Figure 3.4 depicts this function for $(p,\lambda,c) = (\frac{2}{10}, 2, \frac{1}{2})$.

To find the point where the tangent line meets the function, $(\hat{q})$, we need to solve the following equation

$$\frac{f_{(\lambda,c)}(p,\hat{q}) - f_{(\lambda,c)}(p,0)}{\hat{q}} = \frac{\partial f_{(\lambda,c)}(p,q)}{\partial q}\Big|_{\hat{q}}. \tag{3.19}$$

Because the function is initially convex and then concave, the above equation will have at most one solution $\hat{q} \neq 0$. Furthermore the first derivative (3.17) at $q = 1$ and $q = 0$ equals $-\infty$, which implies that for all $\lambda > 1$, there is a solution in $(0,1)$ and this completes the proof.

□

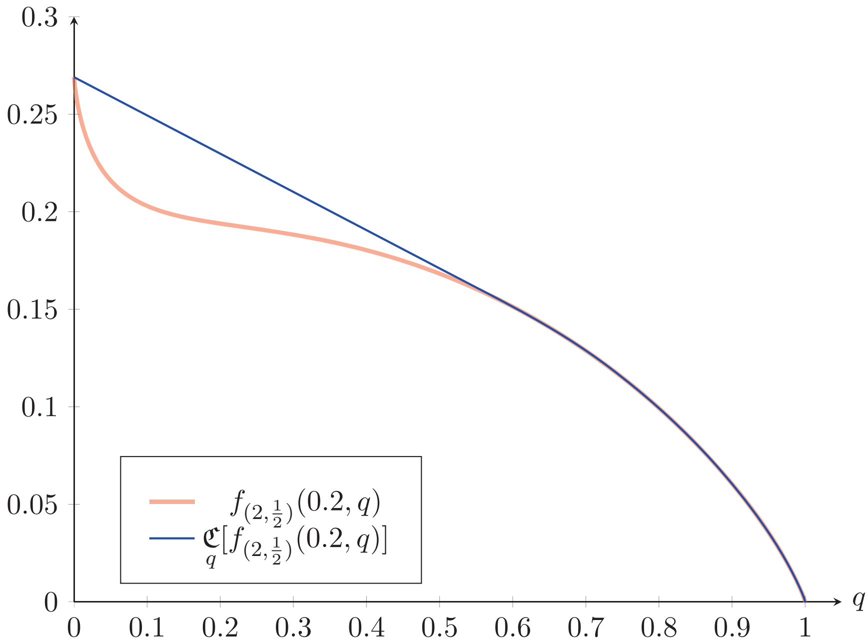


Figure 3.4: The shape of the concave envelope for $S(c)$

Remark 9. It is easy to verify that at $\lambda = 2$ the solution of the equation (3.19) is equal to $\hat{q} = \dfrac{c(1-2p)}{c(1-2p)+p}$. In particular when $c = \frac{1}{2}$

$$\mathop{\mathfrak{C}}_{q}[f_{(2,\frac{1}{2})}(p,q)] = \begin{cases} f_{(2,\frac{1}{2})}(p,q) & q \geq \max(0, 1-2p) \\ \frac{f_{(2,\frac{1}{2})}(p,1-2p) - f_{(2,\frac{1}{2})}(p,0)}{1-2p} q + f_{(2,\frac{1}{2})}(p,0) & \text{Otherwise,} \end{cases}$$

The following Proposition helps us demonstrate that the capacity region of the channel $S(\dfrac{1}{2})$ is strictly larger than the Han–Kobayashi achievable region. To this end we compare the following quantities:

$$\mathscr{A}_2^{\mathrm{HK}}(S(\frac{1}{2})) := \max_{(R_1,R_2)\in\mathscr{A}^{\mathrm{HK}}(S(\frac{1}{2}))} 2R_1 + R_2$$

$$\mathscr{A}_2^{\mathrm{TIN}}(S(\frac{1}{2})^{\otimes 2}) := \max_{(R_1,R_2)\in\mathscr{A}^{\mathrm{TIN}}(S(\frac{1}{2})^{\otimes 2})} 2R_1 + R_2$$

$$\mathscr{C}_2(S(\frac{1}{2})) := \max_{(R_1,R_2)\in\mathscr{C}} 2R_1 + R_2.$$

Proposition 7. $\mathscr{A}_2^{HK}(S(\frac{1}{2})) \leq \ 1.10769 < 1.1081 < \frac{1}{2}\mathscr{A}_2^{TIN}(S(\frac{1}{2})^{\otimes 2}) \leq \mathscr{C}_2(S(\frac{1}{2}))$.

Proof. We first show that $\mathscr{A}_2^{\mathrm{HK}}(S(\frac{1}{2})) \leq \ 1.10769$. By substituting the upper concave envelope from Remark 9 into (3.14) we get $\mathscr{A}_2^{\mathrm{HK}}(S(\frac{1}{2})) = \max\limits_{p,q\in[0,1]^2} g(p,q)$ where

$$g(p,q) = \begin{cases} H_b(q + \frac{p}{2}\bar{q}) + f_{(2,\frac{1}{2})}(p,q) & q \geq \max(0, 1-2p) \\ H_b(q + \frac{p}{2}\bar{q}) + \frac{f_{(2,\frac{1}{2})}(p,1-2p) - f_{(2,\frac{1}{2})}(p,0)}{1-2p} q + f_{(2,\frac{1}{2})}(p,0) & \text{Otherwise,} \end{cases}$$

and as defined in (3.16)

$$f_{(2,\frac{1}{2})}(p,q) = (2p-1)H_b(q) + H_b(q + \frac{p\bar{q}}{2}) - 2pH_b(q + \frac{\bar{q}}{2}).$$

Note that $g(p,q)$ has a closed-form expression that includes basic arithmetic functions and logarithm. The interval arithmetic method is a technique which yields formal bounds for functions consisting of basic arithmetic functions and commonly used functions such as logarithms and trigonometric functions. We use the Julia based implementation of this formal method and it yielded that $\max g(p,q) \in [1.10751, 1.10769]$. Refer to the appendix for the code that yields this bound.

Remark 10. Using the numerical minimization tool of Scilab and by plotting the function the following is a very good approximation of the maximum of $g(p,q)$.

$$\max_{p,q}(g(p,q)) = 1.107577, \ (p^*, q^*) \arg\max = (0.507829413, 0.436538150)$$

.

Figure 3.5 depicts the function $g(p,q)$ with various resolutions.

On the other hand, it is easy to verify that the 2-letter treating interference as noise region achieves the rate pair

$$(R_1, R_2) = (0.064029, 0.980083), \quad \text{where } 2R_1 + R_2 = 1.108141,$$

The distribution that achieves the above rate pair for the treating interference as noise region is as follows,

$$P((X_{11}, X_{12}) = (0,0)) = p^* \qquad P((X_{11}, X_{12}) = (1,1)) = 1 - p^*$$

$$P((X_{21}, X_{22}) = (0,0)) = 0.36q^* \qquad P((X_{21}, X_{22}) = (1,1)) = 1 - 1.64q^*$$

$$P((X_{21}, X_{22}) = (0,1)) = 0.64q^* \qquad P((X_{21}, X_{22}) = (1,0)) = 0.64q^*$$

This demonstrates that $1.1081 < \frac{1}{2}\mathscr{A}_2^{\text{TIN}}(S(\frac{1}{2})^{\otimes 2})$ and hence completes the proof of the proposition as the last inequality follows by definition of the capacity region.

□

Without providing further details, we present our numerical computations of the single-letter and 2-letter Han–Kobayashi regions for the channel $S(\frac{1}{2})$. This is

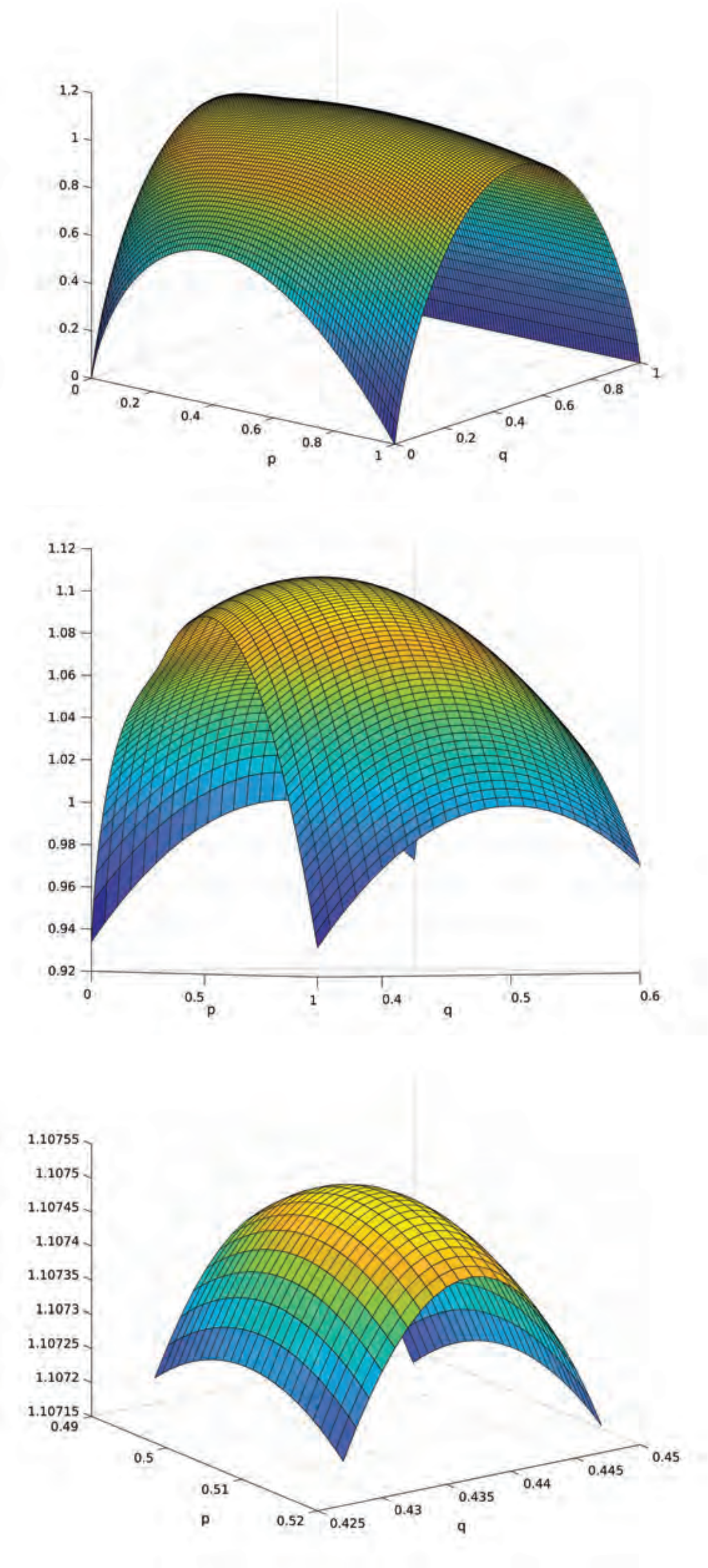


Figure 3.5: $g(p, q)$

basically done by conducting a numerically accurate approximation to the concave envelope and taking a fine grid.

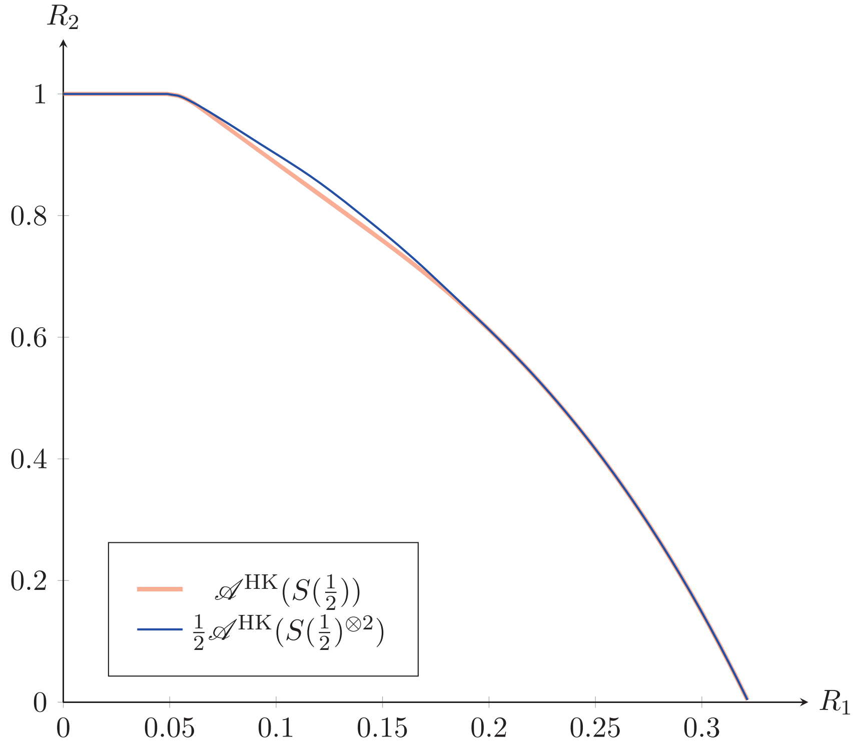


Figure 3.6: Single-letter and 2-letter Han–Kobayshi regions

3.2 Deterministic Binary Interference channel

This section concerns the set of achievable rate pairs for a deterministic interference channel where the inputs are binary and one receiver (Y_1) receives the (logical) AND of the two inputs while the other receiver (Y_2) receives the (logical) OR of the two inputs. This is the only deterministic interference channel setting (up to isomorphism) with binary inputs and outputs whose capacity region has not been established. Etkin and Ordentlich in section (V) of [7] conjectured that the capacity region of this setting (AND–OR channel) coincides with the time-division region ($R_1+R_2 \leq 1$). We disprove this conjecture by demonstrating that the sum-rate of the Han–Kobayashi achievable region is at least 1.0157.

Proposition 8. *Any non-negative $R_1 + R_2$ satisfying*

$$R_1 + R_2 \leq I(X_1; Y_1|U_2, Q) + I(X_2; Y_2|U_1, Q) \tag{3.20}$$

$$R_1 + R_2 \leq I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q) \tag{3.21}$$

$$R_1 + R_2 \leq I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q) \tag{3.22}$$

$$R_1 + R_2 \leq I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q) \tag{3.23}$$

for some $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$ is achievable.

Proof. This follows by applying Fourier-Motzkin elimination on the Han–Kobayashi region in Theorem 3.

□

The following Proposition helps find a simpler characterization of the sum-rate of the Han–Kobayashi achievable region for the AND–OR channel.

Proposition 9. *For the AND–OR channel, any non-negative $R_1 + R_2$ satisfying*

$$R_1 + R_2 \leq I(X_1; Y_1|U_2, Q) + I(X_2; Y_2|U_1, Q) \tag{3.24}$$

$$R_1 + R_2 \leq \frac{1}{2}\Big(I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_2, U_1; Y_2|Q) + I(X_1; Y_1|U_1, U_2, Q)\Big) \tag{3.25}$$

$$R_1 + R_2 \leq I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q) \tag{3.26}$$

for some $p(q)p(u_1, x_1|q)p(u_2, x_2|q)$ is achievable.

Proof. Given a joint distribution $p(q)p_1(u_1, x_1|q)p_2(u_2, x_2|q)$, let the constraints (3.24) $= a$, (3.25) $= b$ and (3.26) $= c$. We now define an induced symmetrized distribution. Let $S \in \{0, 1\}$ be a binary random variable. Define $\tilde{Q} = (S, Q)$ and consider a joint distribution $r(\tilde{Q}, \tilde{U}_1, \tilde{X}_1, \tilde{U}_2, \tilde{X}_2)$ defined as

$$r\left(\tilde{Q} = (0, q), (\tilde{U}_1, \tilde{X}_1) = (u_1, x_1), (\tilde{U}_2, \tilde{X}_2) = (u_2, x_2)\right) = \frac{1}{2}p(q)p_1(u_1, x_1|q)p_2(u_2, x_2|q),$$

$$r\left(\tilde{Q} = (1, q), (\tilde{U}_1, \tilde{X}_1) = (u_2, x_2), (\tilde{U}_2, \tilde{X}_2) = (u_1, x_1)\right) = \frac{1}{2}p(q)p_1(u_1, \bar{x}_1|q)p_2(u_2, \bar{x}_2|q).$$

Observe that substituting the symmetrized distribution r into Proposition 8 for the AND–OR channel yields (3.20) $= a$, (3.21) $= b$, (3.22) $= b$ and (3.23) $= c$. This shows that any non-negative $R_1 + R_2$ satisfying the constraints of this proposition also satisfies the constraints of Proposition 8 for the symmetrized distribution. Therefore the sum-rate given by this proposition is equal to the sum-rate of the Han–Kobayashi achievable region.

□

Proposition 10. *The maximum sum–rate of the Han–Kobayashi region for the AND–OR channel is larger than 1.015.*

Proof. This is done by providing a joint distribution $p(q)p_1(u_1, x_1|q)p_2(u_2, x_2|q)$ on (Q, U_1, U_2, X_1, X_2) for which all of the constraints in Proposition 9 are larger than 1.015. Let $Q \in \{1, 2\}$ be a binary random variable and $p_1(u_1, x_1|q = 1)p_2(u_2, x_2|q = 1)$ be such that

$$U_1 = X_1, \quad U_2 = X_2, \quad P(X_1 = 0) = 0.3331, \quad P(X_2 = 1) = P(X_1 = 0)$$

Let $p_1(u_1, x_1|q = 2)p_2(u_2, x_2|q = 2)$ be such that

$$U_1 = 1, \quad U_2 = X_2, \quad P(X_1 = 0) = 0.4838, \quad P(X_2 = 0) = 0.0792$$

By setting $P(Q = 1) = 0.0773$, the three constraints (3.24), (3.25) and (3.26) on sum-rate respectively evaluate to 1.0395, 1.0157 and 1.0157 which finishes the proof of the proposition.

□

3.3 Summary and Discussion

In this chapter we presented two results. First we showed that there are channels for which Han–Kobayashi achievable region is strictly contained inside the capacity region. Towards this end, we identified a class of channels where the weighted sum-rate expression simplifies and can be computed numerically to a high degree of precision. We identified a further sub-class where we could use interval arithmetic to exhibit a gap. All of these involved working with extremizers of non-convex optimization problems and identifying optimal auxiliary variables.

Our 2-letter distribution that outperforms the single-letter distribution indicates that even repetition coding improved on the memoryless coding. This again, similar to previous chapter, points to the fact that the current schemes do not fully utilize the temporal diversity gains in a memoryless setting.

Finally using the tools we developed to compute the Han–Kobayashi region, we were able to disprove a conjecture (belief) by Etkin and Ordentlich on the optimality of the time-division strategy for the AND-OR interference channel.

□ **End of chapter.**

Appendix

```
In [1]: using IntervalArithmetic

using IntervalOptimisation

In [2]: function binH(r)

if r*(1-r)<10^(-8)

h=0

else

h = -(r*log(2,r) + (1-r)*log(2,1-r))

end

return h

end

Out[2]: binH (generic function with 1 method)

In [3]: function f(p,q)

x = (2*p-1)*binH(q) - 2*p*binH((1+q)/2) + binH(q + p*(1-q)/2)

return x

end

Out[3]: f (generic function with 1 method)

In [4]: function G(p,q)

if q+2*p < 1
```

```
x = binH(q+p*(1-q)/2) + (f(p,1-2*p)-f(p,0))*q/(1-2*p)

+ f(p,0)

else

x = binH(q+p*(1-q)/2) + f(p,q)

end

return x

end
```

```
Out[4]: G (generic function with 1 method)
```

```
In [5]: @time global_max, maximisers = maximise(X ->((x,y) = X; G(x,y)),

(0.01..0.99) x (0.01..0.99),0.2*1e-4);
```

```
9848.472772 seconds (699.59 k allocations: 420.560 MiB, 0.00% gc time)
```

```
In [6]: @interval(global_max)
```

```
Out[6]: [1.10751, 1.10769]
```

Bibliography

[1] V. S. Annapureddy and V. V. Veeravalli, *Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region*, IEEE Transactions on Information Theory **55** (2009), no. 7, 3032–3050.

[2] P. Bergmans, *Random coding theorem for broadcast channels with degraded components*, IEEE Transactions on Information Theory **19** (1973), no. 2, 197–207.

[3] A. Carleial, *A case where interference does not reduce capacity (corresp.)*, IEEE Transactions on Information Theory **21** (1975), no. 5, 569–570.

[4] C.Nair, *Upper concave envelopes and auxiliary random variables*, International Journal of Advances in Engineering Sciences and Applied Mathematics **5** (2013), no. 1, 12–20 (English).

[5] M. Costa and A. E. Gamal, *The capacity region of the discrete memoryless interference channel with strong interference (corresp.)*, IEEE Transactions on Information Theory **33** (1987), no. 5, 710–711.

[6] T. Cover, *Broadcast channels*, IEEE Transactions on Information Theory **18** (1972), no. 1, 2–14.

[7] R. H. Etkin and E. Ordentlich, *Analysis of deterministic binary interference channels via a general outer bound*, IEEE Transactions on Information Theory **57** (2011), no. 5, 2597–2604.

[8] R. G. Gallager, *Capacity and coding for degraded broadcast channels*, Probl. Peredac. Inform. **10(3)** (1974), 3–14.

[9] A. E. Gamal, *The capacity of a class of broadcast channels*, IEEE Transactions on Information Theory **25** (1979), no. 2, 166–169.

[10] A. E. Gamal and M. Costa, *The capacity region of a class of deterministic interference channels (corresp.)*, IEEE Transactions on Information Theory **28** (1982), no. 2, 343–346.

[11] A. E. Gamal and Young-Han Kim, *Network information theory*, Cambridge University Press, 2012.

[12] Y. Geng, A. Gohari, C. Nair, and Y. Yu, *On marton's inner bound and its optimality for classes of product broadcast channels*, IEEE Transactions on Information Theory **60** (2014), no. 1, 22–41.

[13] Y. Geng and C. Nair, *The capacity region of the two-receiver gaussian vector broadcast channel with private and common messages*, IEEE Transactions on Information Theory **60** (2014), no. 4, 2087–2104.

[14] J. Korner and K. Marton, *General broadcast channels with degraded message sets*, IEEE Transactions on Information Theory **23** (1977), no. 1, 60–64.

[15] ———, *Images of a set via two channels and their role in multi-user communication*, IEEE Transactions on Information Theory **23** (1977), no. 6, 751–761.

[16] S. Liu, C. Nair, and L. Xia, *Interference channels with very weak interference*, 2014 IEEE International Symposium on Information Theory, June 2014, pp. 1031–1035.

[17] A. S. Motahari and A. K. Khandani, *Capacity bounds for the gaussian interference channel*, IEEE Transactions on Information Theory **55** (2009), no. 2, 620–643.

[18] C. Nair, *Capacity regions of two new classes of two-receiver broadcast channels*, IEEE Transactions on Information Theory **56** (2010), no. 9, 4207–4214.

[19] C. Nair and A. E. Gamal, *The capacity region of a class of three-receiver broadcast channels with degraded message sets*, IEEE Transactions on Information Theory **55** (2009), no. 10, 4479–4493.

[20] C. Nair and Z. V. Wang, *On 3-receiver broadcast channels with 2-degraded message sets*, 2009 IEEE International Symposium on Information Theory, June 2009, pp. 1844–1848.

[21] ———, *The capacity region of the three receiver less noisy broadcast channel*, IEEE Transactions on Information Theory **57** (2011), no. 7, 4058–4062.

[22] C. Nair and L. Xia, *On three-receiver more capable channels*, 2012 IEEE International Symposium on Information Theory Proceedings, July 2012, pp. 378–382.

[23] C. Nair, L. Xia, and M. Yazdanpanah, *Sub-optimality of han-kobayashi achievable region for interference channels*, 2015 IEEE International Symposium on Information Theory (ISIT), June 2015, pp. 2416–2420.

[24] C. Nair and M. Yazdanpanah, *Sub-optimality of superposition coding region for three receiver broadcast channel with two degraded message sets*, 2017 IEEE International Symposium on Information Theory (ISIT), June 2017, pp. 1038–1042.

[25] H. Sato, *On the capacity region of a discrete two-user channel for strong interference (corresp.)*, IEEE Transactions on Information Theory **24** (1978), no. 3, 377–379.

[26] X. Shang, G. Kramer, and B. Chen, *A new outer bound and the noisy-interference sum–rate capacity for gaussian interference channels*, IEEE Transactions on Information Theory **55** (2009), no. 2, 689–699.

[27] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), no. 3, 379–423.

[28] J. M. Walsh and S. Weber, *Relationships among bounds for the region of entropic vectors in four variables*, 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Sept 2010, pp. 1319–1326.

[29] H. Weingarten, Y. Steinberg, and S. S. Shamai, *The capacity region of the gaussian multiple-input multiple-output broadcast channel*, IEEE Transactions on Information Theory **52** (2006), no. 9, 3936–3964.