Network Effects in Small Networks:
A Study of Cooperation

Thesis by
Parham Noorzad

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended May 22, 2017

Parham Noorzad
ORCID: 0000-0002-0201-3791

To my family

ACKNOWLEDGMENTS

I feel very lucky to have had Prof. Michelle Effros as my advisor. During the last five years of working with her, not only have I been able to amplify my strengths, but also develop abilities and skills that I had been completely ignorant of. Clearly, in working with an experienced information theorist, one obtains a fair amount of technical knowledge. However, Prof. Effros has played a much greater role in my graduate studies. She has been an excellent mentor in all aspects of professional life, and I learned the systematic way of doing high-quality research from her. For that I am forever grateful.

I would like to thank Prof. Michael Langberg, with whom I had the great fortune of collaborating right from the start. I learned a lot through my interactions with Prof. Langberg on both technical and non-technical fronts; from the power of combinatorial thinking, to the importance of counterexamples in research, to writing with simplicity and clarity. Without him, this work would have been unnecessarily more complicated and would have taken much longer. Working with Prof. Langberg has been a great pleasure for me.

In addition to Prof. Effros and Prof. Langberg, Professors Victoria Kostina, Shuki Bruck, Babak Hassibi, and Katrina Ligett were on my committee. I have learned so much from each of them through various interactions in classes, seminars, and during my candidacy. From Prof. Kostina, I learned the essentials of finite blocklength information theory and had the great opportunity to collaborate with her on an interesting problem. From Prof. Bruck, I learned about the principles of great teaching, in addition to constantly assessing the implications of one's work in the development of new technologies. From Prof. Hassibi, I learned about statistical signal processing and random matrices, in addition to the fundamental connection between control and communication. From Prof. Ligett, I learned to keep an open mind when dealing with models. What is simply assumed in one field, may be the main point of contention in another. I feel very fortunate to have had a committee of such high caliber.

I would also like to thank Dr. Tracey Ho, who was my co-advisor during the first year and with whom I had the opportunity to collaborate on my first paper.

I am deeply grateful to Howard and Jan Oringer for their support of my studies during my first year at Caltech. Their work on tackling some of the most important social issues of our time has been a great inspiration for me.

I would like to thank my friend and colleague Ming Fai Wong. Ming Fai helped me a lot when I started doing research and was a great resource for my endless questions on network coding. He was always available when I wanted to discuss research problems or even non-research related life issues. I thank Ming Fai for his support and friendship.

I would like to thank my labmate Sahin Lale for many interesting discussions on information theory and its future applications. I would also like to thank my EE and CMS colleagues, including Srikanth Tenneti, Ramya Korlakai Vinayak, Wentao Huang, Wael Halbawi, Fariborz Salehi, Siddharth Jain, Armeen Taeb, Josh Brake, Ehsan Abbasi, Abhinav Agarwal, Yorie Nakahira, and Yoke Peng Leong for their support and friendship.

As part of the Caltech Graduate Student Council, I had the opportunity to participate in a number of amazing projects. Many thanks go to Natalie Higgins, Andrew Robbins, Henry Ngo, Alicia Lanz, and Allison Strom, with whom I frequently collaborated. I thank them for their friendship and for being inspiring colleagues.

One of the features of Caltech is that it is not too hard to find great friends outside of one's own department. I would like to thank Hamed Hamze for many interesting discussions regarding theoretical research, finance, economics, careers, startups, and many other topics. I thank Tami Khazaei for being a great friend and for many uplifting conversations. I would like to thank Emad Nasrollahpoursamami and William Chan for many inspiring discussions about mathematics, careers, and life in general.

Many thanks go to Caltech EE Staff, especially Tanya Owen, Shirley Slattery, and Katie Pichotta, who through their constant care, help, and support, made the administrative side of graduate work very seamless.

I am very grateful to Laura Flower Kim and Daniel Yoder, from the Caltech International Offices, who from day one were incredibly friendly and supportive.

I would also like to acknowledge the important role of my college professors for their guidance during the graduate application process. Specifically, I would like to thank Professors Mohammad Bagher Asadi, Ali Olfat, Amir Masoud

Rabiei, Maryam Sabbaghian, and Mehrdad Shahshahani. They believed in me early on, and I am grateful for that. I would also like to thank my high school math teacher, Mani Rezaei, from whom I first learned about rigorous thinking and proof writing in mathematics.

Finally, I would like to thank my parents and my sister, to whom this thesis is dedicated. Without their continuous love, support, and emphasis on long-term thinking, this work would not have been possible.

ABSTRACT

Communication over a point-to-point link is relatively well understood. However, when such a link is part of a larger network, our understanding is far from complete. Nonetheless, progress in this area has important consequences in both the theoretical and practical aspects of communication networks.

In this work, we focus on the role of a single link in networks that in addition to point-to-point links, contain "multi-terminal components." An example of a network consisting of a single multi-terminal component is the uplink in a wireless communication network where multiple transmitters communicate with a single receiver over a shared medium. We demonstrate the existence of a class of such networks where a finite capacity link results in a rate gain for each source that far exceeds the capacity of that link. This is an example of a "network effect": the phenomenon where a resource, here link capacity, is significantly more valuable in a network than in isolation. Here we measure the "value" of the finite capacity link by the sum-capacity gain per source that it enables.

The central idea behind the construction of networks that exhibit such effects is the introduction of a node, referred to as the "cooperation facilitator" (CF), that allows other network nodes to work together to reduce interference. In the setting of the classical multiple access channel (MAC), an example of a CF is a node that receives rate-limited information from each transmitter and broadcasts rate-limited information back to the transmitters through a common bottleneck link. Let the "cooperation rate" be the capacity of the CF bottleneck link. We show that for a class of MACs, the presence of a CF leads to a sum-capacity gain that, as a function of the cooperation rate, has an infinite slope at cooperation rate zero. This means that the bottleneck link of the CF is significantly more valuable in some networks than in isolation. This class of MACs includes well-known examples such as the Gaussian MAC and the binary adder MAC.

In addition to sum-capacity gain, cooperation under the CF model also improves reliability. Specifically, in the case of the MAC with two transmitters, whenever the CF has full access to both messages, the maximal- and average-error capacity regions coincide. This effect is observed even when the coop-

eration rate is "negligible"; that is, the cooperation rate grows sub-linearly in the number of channel uses. An implication of this result is the existence of a network whose maximal-error sum-capacity is not continuous with respect to the capacities of its edges; this means that in some networks, even a negligible cooperation rate leads to a positive sum-capacity gain.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] P. Noorzad, M. Effros, and M. Langberg, "Can negligible rate increase network reliability?", *IEEE Transactions on Information Theory*, To appear.
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

[2] P. Noorzad, M. Effros, and M. Langberg, "The benefit of encoder cooperation in the presence of state information", *Proceedings of the IEEE International Symposium on Information Theory*, 2017, To appear.
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

[3] P. Noorzad, M. Effros, and M. Langberg, "The unbounded benefit of encoder cooperation for the k-user MAC", in *Proceedings of the IEEE International Symposium on Information Theory*, 2016. DOI: 10.1109/ISIT.2016.7541317,
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

[4] P. Noorzad, M. Effros, and M. Langberg, "Can negligible cooperation increase network reliability?", in *Proceedings of the IEEE International Symposium on Information Theory*, 2016. DOI: 10.1109/ISIT.2016.7541606,
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

[5] P. Noorzad, M. Effros, and M. Langberg, "On the cost and benefit of cooperation", in *Proceedings of the IEEE International Symposium on Information Theory*, 2015. DOI: 10.1109/ISIT.2015.7282412,
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

[6] P. Noorzad, M. Effros, M. Langberg, and T. Ho, "On the power of cooperation: Can a little help a lot?", in *Proceedings of the IEEE International Symposium on Information Theory*, 2014. DOI: 10.1109/ISIT.2014.6875411,
P.N. participated in the conception of the project, in the derivation of the proofs, and in the writing of the manuscript.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Communication networks play an increasingly important role in our society. While demand is growing, the resources used to meet those demands (e.g., bandwidth and power in wireless communication) are not. This leaves us with one possible solution: we have to use network resources as efficiently as possible.

One way to increase efficiency is to allow network nodes to work together, or "cooperate." Through cooperation, over-constrained regions of the network can take advantage of the resources available in less constrained regions. This results in a reduction in interference and thus an increase in throughput at a given power.

There are many ways network nodes can cooperate. In fact, many well-studied mechanisms such as feedback or relay nodes can be thought of as cooperative designs [1]. Here we consider cooperative strategies that are enabled through finite capacity noiseless links among network nodes. Such designs can take two possible forms: direct and indirect. (See Figure 1.1.) Our focus is on indirect cooperation.

A simple instance of indirect cooperation is the scenario where two or more nodes are connected via a pair of incoming and outgoing links to a common node, which we refer to as a "cooperation facilitator" (CF). Each node sends some information to the CF; using this information, the CF then helps each node improve its transmission rate. The amount of information sent and received by the CF is limited by the capacities of its input and output links.

1.1 The Cooperation Facilitator

We begin by considering cooperation in a network consisting of two encoders, a CF, a multiple access channel (MAC), and a decoder. In this network, the aim of each encoder is to transmit its message to the decoder with small probability of error.

For $i \in \{1, 2\}$, encoder i sends rate-limited information to the CF over a link of capacity C_{in}^i. The CF, using the information it receives from the encoders,

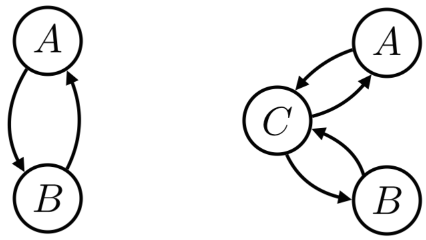

Figure 1.1: Left: Cooperation model where nodes A and B communicate *directly*. Right: Cooperation model where nodes A and B communicate *indirectly* through a third party C.

sends rate-limited information back to encoder i over a link of capacity C^i_{out}. The communication between the CF and the encoders can continue for a finite number of rounds, where in each round, both the CF and the encoders may use any information received in prior rounds. A CF with input link capacities $\mathbf{C}_{\text{in}} = (C^1_{\text{in}}, C^2_{\text{in}})$ and output link capacities $\mathbf{C}_{\text{out}} = (C^1_{\text{out}}, C^2_{\text{out}})$ is referred to as a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. (See Figure 1.2.)

Once the communication between the encoders and the CF is complete, each encoder, based on its message and the information it receives from the CF, transmits a codeword over the MAC. Upon receiving the output, the decoder seeks to find the transmitted messages.

In this work, we consider two benefits of cooperation: rate and reliability. We discuss each separately below.

1.2 Rate Benefit

To measure the rate benefit of cooperation, we calculate the sum-capacity gain resulting from cooperation. Sum-capacity is the maximum of the sum of transmission rates for all sources.

Our aim is to understand how the sum-capacity of a network consisting of a MAC and a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF depends on $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$. As we show in Chapter 3, however, we observe the most interesting behavior of sum-capacity when we fix $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$ and study it solely as a function of $\mathbf{C}_{\text{out}}$.

To guide our study, we briefly discuss prior work. When $\mathbf{C}_{\text{out}} = \mathbf{0}$, no cooperation is possible, and the sum-capacity is obtained from the capacity region of the classical MAC as derived by Ahlswede [2], [3] and Liao [4]. The sum-capacity of this setting serves as a baseline for measuring the rate benefit of

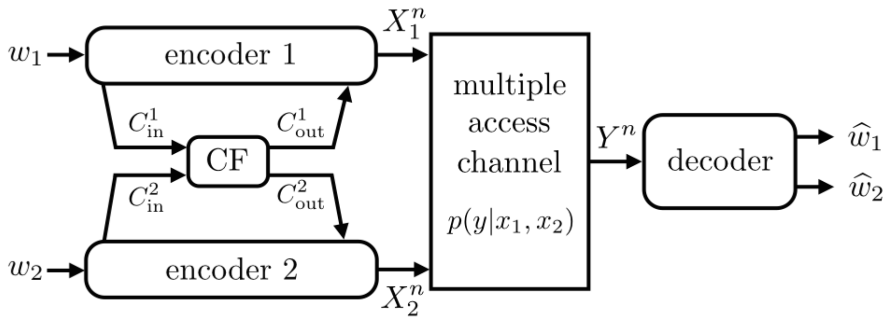

Figure 1.2: The network model for the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. Here w_1 and w_2 are the messages, n is the blocklength, X_1^n and X_2^n are the codewords, and Y^n is the MAC output.

cooperation; specifically, the "cooperation gain" is defined as the difference between sum-capacity in the presence of cooperation and the sum-capacity of the classical MAC.

While our aim is to understand *indirect* cooperation, the *direct* cooperation model of Willems [5], referred to as the "conferencing encoders model," proves important for our study. In Willems' model, which was also originally introduced for the MAC, the encoders are connected via directed links of capacities C_{12} and C_{21}. The main contribution of [5] is that it determines the capacity region of this setup. (See Figure 1.3.)

Using Willems' region [5], we see that the sum-capacity gain of (C_{12}, C_{21})-conferencing is bounded from above by $C_{12} + C_{21}$, which is the total amount of information the encoders receive per time step as a result of cooperation.

In the CF setting, the encoders receive a total of $C_{\text{out}}^1 + C_{\text{out}}^2$ bits per time step in the cooperation process. Does a result similar to conferencing hold in this case? That is, does there exist a *universal* constant K, independent of the MAC, such that the sum-capacity gain is always bounded from above by

$$K(C_{\text{out}}^1 + C_{\text{out}}^2)? \tag{1.1}$$

In Chapter 2, we demonstrate that no such universal constant exists; that is, for (1.1) to be an upper bound for the sum-capacity gain, K must depend on the MAC. This is done by constructing a sequence of MACs where for the mth MAC in the sequence, setting

$$C_{\text{out}}^1 = C_{\text{out}}^2 = \log\big(m \log m\big)$$

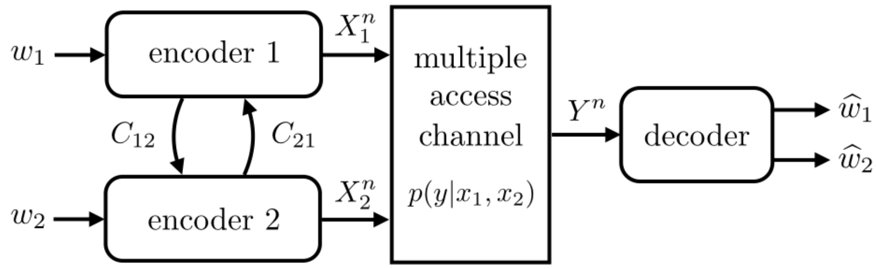

Figure 1.3: Network model for the MAC with (C_{12}, C_{21})-conferencing [5].

leads to a sum-capacity gain that is linear in m.

In Chapter 3, we extend the ideas of Chapter 2 in a number of directions. First, we consider the more general setting where for some $k \geq 2$, there are k encoders in our network. The encoders are all connected to a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, where $\mathbf{C}_{\text{in}}$ and $\mathbf{C}_{\text{out}}$ are vectors in $\mathbb{R}^k_{\geq 0}$. Second, we give a single-letter representation of a class of MACs for which the sum-capacity gain is not bounded by *any* linear function of $\mathbf{C}_{\text{out}}$. In addition, this class of MACs includes many well-known channels, such as the Gaussian MAC [6].

Even though a simple coding strategy suffices for the purposes of Chapter 2, we require random coding arguments in Chapter 3. In addition to techniques from Ulrey [7], who extended Ahlswede's and Liao's results to the k-user MAC, we apply a coding strategy due to Marton [8]. While Marton initially developed this coding strategy for the broadcast channel, it is particularly well-suited for the MAC in the presence of a CF. In fact, this is the strategy that is responsible for the large sum-capacity gain mentioned earlier.

In Chapter 4, we return to the setting of the 2-user MAC. Here we also study the sum-capacity gain, but in the presence of "channel state information" at the encoders and the decoder. This model is especially important as it arises in applications such as wireless communication with fading [9]. Our work in this chapter relies on prior work by Jafar [10] and Permuter, Shamai, and Somekh-Baruch [11].

Based on the encoder cooperation gain we demonstrate here, it is natural to ask whether in networks with multiple decoders, a similar gain is possible through decoder cooperation. The answer is negative; in particular, in the two-receiver broadcast channel, decoder cooperation via a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF cannot increase the sum-capacity by more than $C^1_{\text{out}} + C^2_{\text{out}}$. This limitation of decoder cooperation first appears in the work of Jalali, Effros, and Ho [12]. We provide a more detailed discussion of this result in Appendix D.

1.3 Reliability Benefit

In Chapter 5, we study the reliability benefit of cooperation.[1] The reliability of a code is a measure of the distribution of the error probabilities of the message vectors; in particular, we say that a code with small maximal probability of error is more reliable than a code with the same rate that has small average probability of error.

In Shannon's original work [14], which marks the beginning of information theory, the capacity of a point-to-point channel is defined as the supremum of all rates that are achievable with small *maximal* probability of error. In [14], Shannon defines the *average* probability of error as well, and proves that the capacity of a point-to-point channel is the same under both maximal and average probability of error constraints.

Three decades after Shannon, Dueck [15] demonstrated that Shannon's result does not extend to networks; specifically, Dueck constructed a MAC for which the maximal-error capacity region is strictly contained in the average-error capacity region. More than a decade later, however, Willems [16] proved that Shannon's result *does* extend to the broadcast channel; that is, the broadcast channel maximal- and average-error capacity regions coincide.

Given the results of Shannon [14], Dueck [15], and Willems [16], one may wonder where the MAC with rate-limited encoder cooperation lies; that is, what is the relationship between the average- and maximal-error capacity regions of the MAC with a CF? We answer this question in Chapter 5. Specifically, we show that in the case where the CF has full access to the encoders' messages, the maximal- and average-error capacity regions are the same as long as the capacities of the CF output links are positive. So in this scenario, our result falls in line with Shannon [14] and Willems [16].

In the case where the CF only has partial access to the messages, however, using Dueck's construction [15], we show that the maximal-error capacity re-

[1] We remark here that our use of the term "reliability" differs from prior work in the literature where "reliability" is used to describe the rate of decay of the average error probability of a sequence of codes with strictly increasing blocklength [13, p. 160].

gion is not necessarily equal to the average-error region. Nonetheless, there is still a reliability benefit, but the maximal probability of error constraint is too strict for its description. To address this problem, we define a family of error probability constraints that lie between the maximal and average error probabilities. We discuss this idea in detail in Chapter 5.

1.4 The Edge Removal Problem

Our work here is strongly motivated by results from the network coding literature. The network coding literature studies networks that solely consist of noiseless links of finite capacity [17]. The determination of the capacity region of such networks is open in general; in fact, even the impact of a single edge on the capacity region, which we next discuss, is unknown.

The "edge removal problem" [12], [18]–[20] is the problem of determining the effect of removing a finite capacity edge from a network. For simplicity, we focus on the effect of removing an edge on sum-capacity rather than on the entire capacity region. In the study of this problem, various versions of the "edge removal property" are introduced to capture the effect of edge removal in different networks. In this work, we encounter the following three variations. Note that our definitions for the weak and strong edge removal properties are consistent with [21].

In the following definitions, consider a network $\mathcal{N}$ and an edge e in $\mathcal{N}$ of capacity δ. Each variation of the edge removal property is defined for the fixed pair $(\mathcal{N}, e)$. We say the network $\mathcal{N}$ satisfies a certain edge removal property if for all edges e of $\mathcal{N}$, the pair $(\mathcal{N}, e)$ satisfies that property.

1.4.1 The Weak Edge Removal Property

The pair $(\mathcal{N}, e)$ satisfies the "weak edge removal property" if the sum-capacity of $\mathcal{N}$ is continuous in δ at $\delta = 0^+$.

The first study of the weak edge removal property appears in the work of Gu, Effros, and Bakshi [18], who conjecture that any network consisting solely of noiseless links satisfies this property. A similar conjecture also appears in the work of Chan and Grant [19].

1.4.2 The Strong Edge Removal Property

The pair $(\mathcal{N}, e)$ satisfies the "strong edge removal property" if there exists a constant $K > 0$ such that for all $\delta > 0$, removing e from $\mathcal{N}$ reduces the sum-capacity of $\mathcal{N}$ by at most $K\delta$.

Lee, Langberg, and Effros [22] demonstrate that this property holds for the linear programming bounds given by Yeung [23].

1.4.3 The Universal Edge Removal Property

This property is similar to the previous definition, with the difference that the constant K equals the number of sources. Formally, $(\mathcal{N}, e)$ satisfies the "universal edge removal property" if removing e reduces the sum-capacity by at most

$$(\text{the number of sources of } \mathcal{N}) \times \delta.$$

Ho, Effros, and Jalali introduce this property in [12], [20]. The intuition behind their definition is that each source can send at most δ bits of information over an edge of capacity δ; thus the removal of this edge should not affect the rate of each source by more than δ bits. Among other examples, they prove that this property holds for networks with co-located sources, networks where cutset outer bounds fully characterize the capacity region, and networks where linear codes achieve the capacity region.

1.4.4 Edge Removal and Other Open Problems

In addition to being a fundamental question in the research on network capacity, the edge removal problem is also related to a relatively large collection of problems that are either equivalent to or are implied by some variation of the edge removal property. We next mention some of these problems.

Gu and Effros [24], [25], and more recently, Kosut and Kliewer [21], establish connections between different variations of the edge removal property and strong converse results. In [26], Langberg and Effros study the connection between the edge removal problem and the zero- and ϵ-error network coding capacity regions. In [27], the same authors study the connection between the edge removal problem and the network source coding problem for dependent sources. Connections between the edge removal problem and a number of other problems in network coding are further explored by Wong, Effros, and Langberg [28]–[30].

1.4.5 Edge Removal in Noisy Networks.

"Noisy networks" are networks that in addition to point-to-point channels and noiseless links, also contain multi-terminal components such as the MAC or the broadcast channel. This definition is motivated by the fact that in any network, noisy point-to-point channels may be replaced by noiseless links of the same capacity without affecting the network capacity region [31]. In this work, we mainly consider the edge removal problem in noisy networks. This additional freedom allows us to settle problems that to date are still open in the network coding domain.

Specifically, in Chapter 2, we give an example of a network that does not satisfy the universal edge removal property. On the other hand, in Appendix D, we show that in any network, any ingoing edge of a node that has no outgoing edges satisfies the universal edge removal property. In Chapter 3, we extend the results of Chapter 2 by constructing a class of networks that do not satisfy the strong edge removal property. In Chapter 5, we exhibit a class of networks that do not satisfy the weak edge removal property. Despite this, the results in Chapter 5 are not stronger than those in Chapter 3. This is due to the fact that the counterexample in Chapter 5 is with respect to the *maximal-error* capacity region whereas Chapter 3 is concerned with the *average-error* capacity region. Our example in Chapter 2, however, applies to both the maximal- and average-error capacity regions. Finally, in Appendix C, we show that the network consisting of any MAC with a CF that has access to both messages *does satisfy* the weak edge removal property with respect to the average-error capacity region. The following table indicates, for a given variation of the edge removal property and capacity region definition, the chapter with the most relevant result.

Table 1.1: Location of the main result for each variation of the edge removal property.

	weak	strong	universal
maximal-error	Ch. 5	Ch. 5	Ch. 2
average-error	App. C	Ch. 3	Ch. 2/App. D

Chapter 2

A SIMPLE EXAMPLE

In this chapter, we study a simplified version of the MAC with CF model where the CF has access to both messages and communicates with the encoders through a shared bottleneck link. (See Figure 2.1.) Based on this model, we describe a sequence of MACs with increasing alphabet sizes and set the cooperation rate for each channel as a function of its alphabet size. We then show that the increase in sum-capacity that results from cooperation grows more quickly than any *universal* linear function of the cooperation rate.

We begin by reviewing the conferencing model and its capacity region as presented by Willems [5]. We give a formal introduction to the simplified CF model in Section 2.2.

2.1 Conferencing Encoders

Consider the MAC

$$\left(\mathcal{X}_1 \times \mathcal{X}_2, p_{Y|X_1,X_2}(y|x_1,x_2), \mathcal{Y}\right),$$

where $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ are finite sets and $p_{Y|X_1,X_2}(y|x_1,x_2)$ denotes the conditional distribution of the output, Y, given the inputs, X_1 and X_2. To simplify notation, we suppress the subscript of the probability distributions when the corresponding random variables are clear from context. For example, we write $p(x)$ instead of $p_X(x)$.

For every real number $a \geq 1$, $[a] := \{1, \ldots, \lfloor a \rfloor\}$.

There are two sources, source 1 and source 2, whose outputs are the messages $W_1 \in \mathcal{W}_1 = [2^{nR_1}]$ and $W_2 \in \mathcal{W}_2 = [2^{nR_2}]$, respectively. The random variables W_1 and W_2 are independent and uniformly distributed over their corresponding alphabets. The nonnegative real numbers R_1 and R_2 are called the *message rates.*

This material is based upon work supported by the National Science Foundation under Grant Numbers CCF-1321129, CCF-1018741, CCF-1038578, and CNS-0905615. It originally appears in [32].

In the absence of cooperation, each encoder only has access to its corresponding message. The encoders are represented by the functions

$$f_{1n} : \mathcal{W}_1 \to \mathcal{X}_1^n$$
$$f_{2n} : \mathcal{W}_2 \to \mathcal{X}_2^n.$$

We denote the output of the encoders by $x_1^n = f_{1n}(w_1)$ and $x_2^n = f_{2n}(w_2)$. Let Y^n be the output of the channel when the pair (x_1^n, x_2^n) is transmitted. Using Y^n, the decoder estimates the original messages via a decoding function $g_n : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2$.

A $(2^{nR_1}, 2^{nR_2}, n)$ code for the MAC is defined as the triple (f_{1n}, f_{2n}, g_n). Assuming that the messages are uniformly distributed, the average probability of error for this code is given by

$$P_e^{(n)} = \Pr\big(g_n(Y^n) \neq (W_1, W_2)\big).$$

The rate pair (R_1, R_2) is "achievable" if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $P_e^{(n)}$ tends to zero as the blocklength, n, approaches infinity. The capacity region, $\mathscr{C}$, is the closure of the set of all achievable rate pairs.

For a given capacity region $\mathscr{C} \subseteq \mathbb{R}^2_{\geq 0}$, the "sum-capacity," C_{sum}, is defined as

$$C_{\text{sum}} = \max\left\{R_1 + R_2 \middle| (R_1, R_2) \in \mathscr{C}\right\}. \tag{2.1}$$

In the absence of cooperation [2]–[4], the sum-capacity is given by

$$C_{\text{sum}} = \max_{p(x_1)p(x_2)} I\left(X_1, X_2; Y\right).$$

In the conferencing model, each encoder shares some information regarding its message with the other encoder prior to transmission over the channel. This sharing of information is achieved through a K*-round conference* over noiseless links of capacities C_{12} and C_{21}. A K-round conference consists of two sets of functions, $\{h_{11}, \ldots, h_{1K}\}$ and $\{h_{21}, \ldots, h_{2K}\}$, which for every message pair (w_1, w_2), recursively define the random vectors $v_1^K := (v_{11}, \ldots, v_{1K})$ and $v_2^K := (v_{21}, \ldots, v_{2K})$ as

$$v_{1k} = h_{1k}(w_1, v_2^{k-1})$$
$$v_{2k} = h_{2k}(w_2, v_1^{k-1})$$

for $k \in [K]$. In step k, encoder 1 (encoder 2) computes v_{1k} (v_{2k}) and sends it to encoder 2 (encoder 1). Since the noiseless links between the two encoders are of capacities C_{12} and C_{21}, respectively, we require

$$\sum_{k=1}^{K} \log|\mathcal{V}_{1k}| \le nC_{12}$$
$$\sum_{k=1}^{K} \log|\mathcal{V}_{2k}| \le nC_{21},$$

where $\mathcal{V}_{ik}$ is the round-k alphabet of the conferencing output of encoder i for $i \in \{1,2\}$ and $k \in [K]$. The outputs of the encoders, x_1^n and x_2^n, are given by

$$x_1^n = f_{1n}(w_1, v_2^K)$$
$$x_2^n = f_{2n}(w_2, v_1^K),$$

where f_{1n} and f_{2n} are deterministic functions.

Denote the sum-capacity of the MAC with (C_{12}, C_{21})-conferencing with

$$C_{\text{sum}}(C_{12}, C_{21}).$$

By studying the capacity region [5], we deduce

$$C_{\text{sum}}(C_{12}, C_{21}) \le C_{\text{sum}}(0,0) + C_{12} + C_{21}.$$

Thus, with conferencing, the sum-capacity increases at most linearly in (C_{12}, C_{21}) over the sum-capacity in the absence of cooperation.

2.2 The Simplified CF Model

In the CF model, cooperation is made possible not through finite capacity links between the encoders, but instead through a third party, the CF, which receives information from both encoders and transmits a single description of that information back to both (Figure 2.1). The CF is represented by the function

$$\varphi_n : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{Z},$$

where the CF output alphabet $\mathcal{Z} = [2^{n\delta}]$ is determined by the cooperation rate δ. The output of the CF, $z = \varphi_n(w_1, w_2)$, is available to both encoders. Each encoder chooses a blocklength-n codeword as a function of its own message

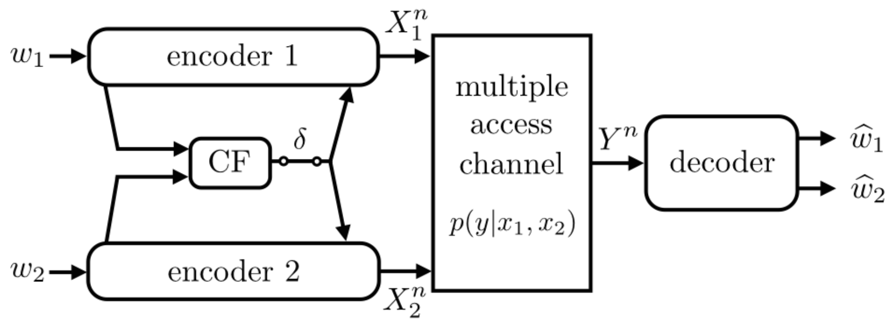

Figure 2.1: Network model for the MAC with a simplified CF. The *cooperation rate* is the capacity of the output link of the CF which we denote by δ.

and z and sends that codeword to the receiver using n transmissions. Hence the two encoders are represented by the functions

$$f_{1n} : \mathcal{W}_1 \times \mathcal{Z} \to \mathcal{X}_1^n$$
$$f_{2n} : \mathcal{W}_2 \times \mathcal{Z} \to \mathcal{X}_2^n.$$

The definitions of the decoder, probability of error, and capacity region are similar to the classical MAC discussed in the previous section and are omitted. We denote the sum-capacity of this network by $C_{\text{sum}}(\delta)$.

Given a pair of functions $f, g : \mathbb{Z}^+ \to \mathbb{R}$, we say $f = o(g)$ if $\lim_{m\to\infty} \frac{f(m)}{g(m)} = 0$. We say $f = \omega(g)$ if $g = o(f)$.

For a sequence of MACs

$$\left\{\left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}, p^{(m)}(y|x_1, x_2), \mathcal{Y}^{(m)}\right)\right\}_m,$$

$C_{\text{sum}}^{(m)}(\delta)$ denotes the CF sum-capacity of the mth channel when the cooperation rate is δ. We define the sum-capacity gain of the mth channel for all cooperation rates $\delta \geq 0$ as

$$G^{(m)}(\delta) := C_{\text{sum}}^{(m)}(\delta) - C_{\text{sum}}^{(m)}(0).$$

In the next theorem, which is the main result of this chapter, we see that for a sequence of MACs, the ratio of the sum-capacity gain to the cooperation rate grows without bound; this demonstrates the existence of a network that does not satisfy the universal edge removal property. In what follows, $\log(x)$ is the base-2 logarithm of x.

Theorem 2.2.1. *For every sequence of cooperation rates* $(\delta_m)_{m=1}^{\infty}$ *satisfying* $\delta_m = \log m + \omega(1)$ *and* $\delta_m \leq m$ *and every* $\epsilon \in (0,1)$*, there exists a sequence of MACs with input alphabets* $\mathcal{X}_1^{(m)} = \mathcal{X}_2^{(m)} = [2^m]$*, such that for sufficiently large* m*,*

$$\left(3 - \sqrt{5+4\epsilon}\right)m - \delta_m \leq G^{(m)}(\delta_m) \leq m + \delta_m.$$

In the above theorem, the choice of δ_m is constrained only by $\delta_m = \log m + \omega(1)$ and $\delta_m \leq m$. For example, for the sequence of MACs in Theorem 2.2.1 a cooperation rate of $\delta_m = \log(m \log m)$ leads to an increase in sum-capacity that is linear in m, giving a capacity benefit that is "almost" exponential in the cooperation rate.

In the next section, we prove the existence of a sequence of MACs with properties that are essential for the proof of Theorem 2.2.1. In Section 2.4, we show that for the sequence of channels of Section 2.3,

$$2m - \delta_m \leq C_{\text{sum}}^{(m)}(\delta_m) \leq 2m. \tag{2.2}$$

In Section 2.5, we show

$$m - \delta_m \leq C_{\text{sum}}^{(m)}(0) \leq (\sqrt{5+4\epsilon} - 1)m. \tag{2.3}$$

Combining these two results gives Theorem 2.2.1.

2.3 Channel Construction

For a fixed positive integer m, the channel

$$\left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}, p^{(m)}(y|x_1,x_2), \mathcal{Y}^{(m)}\right)$$

used in the proof of Theorem 2.2.1 has input alphabets $\mathcal{X}_1^{(m)} = \mathcal{X}_2^{(m)} = [2^m]$ and output alphabet

$$\mathcal{Y}^{(m)} = \left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}\right) \cup \{(E,E)\},$$

where "E" denotes an erasure symbol. For each $(x_1,x_2,y) \in \mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)} \times \mathcal{Y}^{(m)}$, $p^{(m)}(y|x_1,x_2)$ is defined as a function of the corresponding entry $b_{x_1x_2}$ of a $(0,1)$-matrix $B_m = (b_{x_1x_2})_{x_1,x_2=1}^{2^m}$. Precisely,

$$p^{(m)}(y|x_1,x_2) := \begin{cases} 1 - b_{x_1x_2}, & \text{if } y = (x_1,x_2) \\ b_{x_1x_2}, & \text{if } y = (E,E). \end{cases} \tag{2.4}$$

That is, when (x_1, x_2) is transmitted, $y = (x_1, x_2)$ is received if $b_{x_1 x_2} = 0$, and $y = (E, E)$ is received if $b_{x_1 x_2} = 1$. Thus, we interpret the 0 and 1 entries of B_m as "good" and "bad" entries, respectively. Let $\mathcal{X}^{(m)} = \{1, \ldots, 2^m\}$. We define the sets

$$0_{B_m} := \big\{(x_1, x_2) \big| b_{x_1, x_2} = 0\big\}$$
$$1_{B_m} := \big\{(x_1, x_2) \big| b_{x_1, x_2} = 1\big\},$$

to be the set of good and bad entries of $\mathcal{X}^{(m)} \times \mathcal{X}^{(m)}$, respectively.

For every $S, T \subseteq [2^m]$, let $B_m(S, T)$ be the submatrix obtained from B_m by keeping the rows with indices in S and columns with indices in T. For every $x \in [2^m]$, let $B_m(x, T) := B_m(\{x\}, T)$ and $B_m(S, x) := B_m(S, \{x\})$.

The proof of Theorem 2.2.1 requires that B_m satisfies two properties. One is that every sufficiently large submatrix of B_m should have a large fraction of bad entries. This property ensures that the sum-capacity of our channel without cooperation is small (Section 2.5). The second property is that every submatrix of a specific type should have at least one good entry. This property enables a significantly higher sum-capacity under low-rate cooperation using the cooperation facilitator model (Section 2.4). Lemma 2.3.1 demonstrates that these two properties can be simultaneously achieved. A proof of this and all subsequent lemmas can be found in Section 2.6.

Lemma 2.3.1. *Let $f, g : \mathbb{Z}^+ \to \mathbb{Z}^+$ be two functions such that $f(m) = \omega(m)$ and $g(m) = \log m + \omega(1)$. Then for every $\epsilon \in (0, 1)$, there exists a sequence of $2^m \times 2^m$ $(0, 1)$-matrices $(B_m)_{m=1}^{\infty}$ such that*

(1) for every $S, T \subseteq [2^m]$ that satisfy $|S|, |T| \geq f(m)$,

$$\frac{|(S \times T) \cap 1_{B_m}|}{|S|\,|T|} > 1 - \epsilon;$$

that is, in every sufficiently large submatrix of B_m, the fraction of bad entries is larger than $1 - \epsilon$, and

(2) for every $x \in [2^m]$ and $k \in [2^{m-g(m)}]$, each of $B_m(x, \mathcal{X}_{m,k})$ and $B_m(\mathcal{X}_{m,k}, x)$ contains at least one good entry, where

$$\mathcal{X}_{m,k} := \Big\{(k-1)2^{g(m)} + \ell \Big| \ell \in [2^{g(m)}]\Big\};$$

that is, if we break each row or column into consecutive blocks of size $2^{g(m)}$, each block contains at least one good entry.

Channel Definition: Choose functions f and g that satisfy the constraints $f(m) = \omega(m)$, $g(m) = \log m + \omega(1)$, and $\log f(m) = o(m)$. Fix a sequence of channels as defined by (2.4) using matrices $(B_m)_{m=1}^{\infty}$ satisfying the properties proved possible in Lemma 2 for the chosen functions f and g.

2.4 Inner and Outer Bounds for the CF Capacity Region

For the mth channel, using cooperation rate $\delta_m = g(m)$, we show the achievability of the rate pairs $(m, m - g(m))$ and $(m - g(m), m)$. For each, we employ a blocklength-1 code ($n = 1$). Time sharing between these codes results in an inner bound for the capacity region given by

$$R_1, R_2 \leq m$$
$$R_1 + R_2 \leq 2m - g(m).$$

If $R_1 = m$, $R_2 = m - g(m)$, and $n = 1$, then the message alphabets are given by $\mathcal{W}_1 = [2^m]$ and $\mathcal{W}_2 = [2^{m-g(m)}]$. By the second property of our channel in Lemma 2, for every $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$, the submatrix $B_m(w_1, \mathcal{X}_{m,w_2-1})$ contains at least one good entry. Let $z = \varphi(w_1, w_2)$, the output of the CF, be an element of $\mathcal{Z} = [2^{g(m)}]$ such that $(w_1, (w_2 - 1)2^{g(m)} + z)$ is a good entry of $B_m(w_1, \mathcal{X}_{m,w_2-1})$. If there's more than one good entry, pick the one that results in the smallest z.

To transmit message pair (w_1, w_2), encoder 1 sends $x_1 = w_1$ and encoder 2 sends $x_2 = (w_2 - 1)2^{g(m)} + z$. By the definition of our channel (2.4), the channel output equals the channel input (x_1, x_2) with probability one, and hence zero error decoding is possible. Thus the rate pair $(m, m - g(m))$ is achievable. Note that for this achievability scheme to work, only the second encoder needs to know the value of z. A similar argument proves the achievability of $(m - g(m), m)$ and the lower bound of (2.2) follows.

To find an outer bound for the capacity region, we use the capacity region of the conferencing model [5] in a special case. Consider the situation in which encoder 1 has access to both messages and can transmit information to encoder 2 on a noiseless link of capacity δ_m. Then it is easy to see that the capacity region of this network contains the capacity region of the CF model. This situation, however, is equivalent to the conferencing model [5] for $C_{12} = \delta_m$ and $C_{21} = \infty$. Hence an outer bound for the capacity region is given by the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, U) + \delta_m, \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned}$$

for some distribution $p(u)p(x_1|u)\,p(x_2|u)$. Note that

$$\begin{aligned} I(X_1; Y|X_2, U) &\leq H(X_1) \leq m, \\ I(X_1, X_2; Y) &\leq H(X_1, X_2) \leq 2m, \end{aligned}$$

and $\delta_m = g(m)$, so the region

$$\begin{aligned} R_1 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m \end{aligned}$$

is an outer bound for the CF model. Note that if we switch the roles of encoders 1 and 2, we get the outer bound

$$\begin{aligned} R_2 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m. \end{aligned}$$

Since the intersection of two outer bounds is also an outer bound, the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1, R_2 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m \end{aligned}$$

is an outer bound for the CF model as well and the upper bound of (2.2) follows.

2.5 Inner and Outer Bounds in the Absence of Cooperation

Consider the mth channel of the construction in Section 2.3. In the case where there is no cooperation, we show that the set of all rate pairs (R_1, R_2) satisfying

$$R_1 + R_2 \leq m - g(m)$$

is an inner bound for the capacity region. To this end, we show the achievability of the rate pairs $(m - g(m), 0)$ and $(0, m - g(m))$. The achievability of all other rate pairs in the inner bound follows by time-sharing between the encoders. Similar to the achievability result of the previous section, let $n = 1$.

Then $\mathcal{W}_1 = [2^{m-g(m)}]$ and $\mathcal{W}_2 = \{1\}$. By our channel construction, for every $w \in \mathcal{W}_1$, $B_m(\mathcal{X}_{m,w-1}, 1)$ contains at least one good entry. This means that the first column of B_m contains at least $|\mathcal{W}_1| = 2^{m-g(m)}$ good entries. Suppose encoder 1 transmits uniformly on these $2^{m-g(m)}$ good entries and encoder 2 transmits $x_2 = 1$. Then the input is always on a good entry and the channel output is the same as the channel input. Thus the pair $(m-g(m), 0)$ is achievable. A similar argument shows that the pair $(0, m-g(m))$ is achievable and the inner bound follows by time-sharing.

We next find an outer bound for the capacity region of our network in the absence of cooperation.

Let y_1 and y_2 be the components of output y; that is, when $y = (x_1, x_2)$, then $y_1 = x_1$ and $y_2 = x_2$, and when $y = (E, E)$, then $y_1 = y_2 = E$. Note that $y_1, y_2 \in [2^m] \cup \{E\}$. In the absence of cooperation, given an input distribution $p(x_1)p(x_2)$, the distribution $p(y_1)$ is given by

$$p(y_1) = \begin{cases} \gamma_{y_1} & y_1 \in \mathcal{X} \\ 1-\gamma & y_1 = E, \end{cases} \tag{2.5}$$

where

$$\forall x_1 \in [2^m] : \gamma_{x_1} = p(x_1) \sum_{x_2 : b_{x_1 x_2} = 0} p(x_2),$$

and $\gamma := \sum_{x_1} \gamma_{x_1}$. Let $\mathscr{R}_m$ denote the set of all pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I\left(X_1; Y | X_2\right), \\ R_2 &\leq I\left(X_2; Y | X_1\right), \\ R_1 + R_2 &\leq I\left(X_1, X_2; Y\right) \end{aligned}$$

for some distribution $p(x_1)p(x_2)p(y|x_1, x_2)$ and let $\operatorname{conv}(A)$ denote the convex hull of the set A. Then the capacity region in the absence of cooperation is given by the closure of $\operatorname{conv}(\mathscr{R}_m)$ [2]–[4].

If for all pairs $(R_1, R_2) \in \operatorname{conv}(\mathscr{R}_m)$, one of R_1 or R_2 is smaller than or equal to $\log 2f(m)$, then the upper bound in (2.3) follows, since

$$R_1 + R_2 \leq m + \log 2f(m),$$

and $\log f(m) = o(m)$. On the other hand, if there exist rate pairs $(R_1, R_2) \in \operatorname{conv}(\mathscr{R}_m)$ such that

$$R_1, R_2 > \log 2f(m), \tag{2.6}$$

then by the definition of $\mathscr{R}_m$ and (2.6),

$$H(X_1), H(X_2) > \log 2f(m). \tag{2.7}$$

Using (2.7), the following argument shows

$$R_1 + R_2 \leq (\sqrt{5+4\epsilon} - 1)m.$$

For our channel, Y, Y_1, and Y_2 are deterministic functions of (X_1, X_2), (X_1, Y_2) and (Y_1, X_2), respectively. Thus the bounds defining $\mathscr{R}_m$ simplify as

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) = H(Y_1|X_2) \leq H(Y_1), \\ R_2 &\leq I(X_2; Y|X_1) = H(Y_2|X_1) \leq H(Y_2). \end{aligned} \tag{2.8}$$

To bound $H(Y_1)$, we apply the following lemma, which bounds the probability that a random variable X falls in a specific set T; the bound is given as a function of the entropy of X and the cardinality of T. For any set T, we denote its indicator function by $\mathbf{1}_T$.

Lemma 2.5.1. *Let X be a discrete random variable with distribution $p : \mathcal{X} \to \mathbb{R}_{\geq 0}$, and let T be a subset of $\mathcal{X}$. If $q : T \to \mathbb{R}_{\geq 0}$ is a function and $\alpha := \sum_{x \in T} q(x)$, then*

$$-\sum_{x \in T} q(x) \log q(x) \leq \alpha \log |T| - \alpha \log \alpha. \tag{2.9}$$

When $q(x) = p(x)\mathbf{1}_T(x)$, the above inequality implies

$$\alpha = \sum_{x \in T} p(x) \leq K \left(1 - \frac{H(X) - 1}{\log |\mathcal{X}|}\right), \tag{2.10}$$

where $K = \left(1 - \frac{\log |T|}{\log |\mathcal{X}|}\right)^{-1}$.

By (2.5),

$$H(Y_1) = -\sum_{x_1} \gamma_{x_1} \log \gamma_{x_1} - (1-\gamma)\log(1-\gamma).$$

Applying (2.9) from Lemma 2.5.1,

$$H(Y_1) \leq \gamma m + h(\gamma) \leq \gamma m + 1, \tag{2.11}$$

where $h(\gamma)$ denotes the binary entropy function and is given by

$$h(\gamma) = \gamma \log \frac{1}{\gamma} + (1-\gamma) \log \frac{1}{1-\gamma}.$$

We next bound γ. To this end, we write each of the input distributions as a particular convex combination of uniform distributions using the next lemma.

Lemma 2.5.2. *If X is a random variable with finite alphabet $\mathcal{X}$, then there exists a positive integer k, a sequence of positive numbers $(\alpha_j)_{j=1}^k$, and a sequence of non-empty subsets of $\mathcal{X}$, $(S_j)_{j=1}^k$, such that the following properties are satisfied.*

(a) For every $j \in [k-1]$, $S_{j+1} \subsetneq S_j$.

(b) For all $x \in \mathcal{X}$,

$$p(x) = \sum_{j=1}^{k} \alpha_j \frac{\mathbf{1}_{S_j}(x)}{|S_j|}.$$

(c) For every C, $0 < C < |\mathcal{X}|$,

$$\sum_{j:|S_j| \le C} \alpha_j \le K \left(1 - \frac{H(X) - 1}{\log |\mathcal{X}|}\right),$$

where $K = \left(1 - \frac{\log C}{\log |\mathcal{X}|}\right)^{-1}$.

Using Lemma 2.5.2, we write $p(x_1)$ and $p(x_2)$ as

$$p(x_1) = \sum_{i=1}^{k} \alpha_i^{(1)} \frac{\mathbf{1}_{S_i^{(1)}}(x_1)}{|S_i^{(1)}|}$$
$$p(x_2) = \sum_{j=1}^{l} \alpha_j^{(2)} \frac{\mathbf{1}_{S_j^{(2)}}(x_2)}{|S_j^{(2)}|}.$$

Then

$$\gamma = \sum_{x_1, x_2 : b_{x_1 x_2} = 0} p(x_1)p(x_2) = \sum_{i=1}^{k} \sum_{j=1}^{l} \alpha_i^{(1)} \alpha_j^{(2)} \beta_{ij},$$

where

$$\beta_{ij} := \sum_{x_1, x_2 : b_{x_1 x_2} = 0} \frac{\mathbf{1}_{S_i^{(1)}}(x_1) \mathbf{1}_{S_j^{(2)}}(x_2)}{|S_i^{(1)}||S_j^{(2)}|}$$
$$= \frac{\left|\left(S_i^{(1)} \times S_j^{(2)}\right) \cap 0_{B_m}\right|}{|S_i^{(1)}||S_j^{(2)}|}.$$

For every i and j, $\beta_{ij} \le 1$. If, however, $\min\{|S_i^{(1)}|, |S_j^{(2)}|\} \ge f(m)$, then by the first property of our channel (Property (1) in Lemma 2.3.1), $\beta_{ij} \le \epsilon$. Thus by part (c) of Lemma 2.5.2 and (2.7),

$$\begin{aligned}
\gamma &< \epsilon + \sum_{i,j:\min\{|S_i^{(1)}|,|S_j^{(2)}|\}<f(m)} \alpha_i^{(1)}\alpha_j^{(2)} \\
&= \epsilon + 1 - \sum_{i,j:\min\{|S_i^{(1)}|,|S_j^{(2)}|\}\geq f(m)} \alpha_i^{(1)}\alpha_j^{(2)} \\
&= \epsilon + 1 \\
&\quad - \left(1 - \sum_{i:|S_i^{(1)}|<f(m)} \alpha_i^{(1)}\right)\left(1 - \sum_{j:|S_j^{(2)}|<f(m)} \alpha_j^{(2)}\right) \\
&\leq \epsilon + 1 - \left(1 - K_m\left(1 - \frac{H(X_1)-1}{m}\right)\right) \\
&\quad \times \left(1 - K_m\left(1 - \frac{H(X_2)-1}{m}\right)\right),
\end{aligned}$$

where $K_m = \left(1 - \frac{\log f(m)}{m}\right)^{-1}$. Note that $K_m \to 1$ as $m \to \infty$ since $\log f(m) = o(m)$ by assumption. Furthermore, the definition of $\mathscr{R}_m$ and (2.6) imply $\log 2f(m) \leq R_i \leq H(X_i)$ for $i \in \{1,2\}$. Thus

$$\begin{aligned}
\gamma &< \epsilon + 1 - \left(1 - K_m\left(1 - \frac{R_1 - 1}{m}\right)\right) \\
&\quad \times \left(1 - K_m\left(1 - \frac{R_2 - 1}{m}\right)\right) \\
&= \epsilon + K_m\left(2 - \frac{R_1 + R_2 - 2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1 - 1}{m}\right)\left(1 - \frac{R_1 - 1}{m}\right).
\end{aligned}$$

Combining the previous inequality with (2.8) and (2.11) results in

$$\begin{aligned}
\frac{R_1}{m} &\leq \epsilon + \frac{1}{m} + K_m\left(2 - \frac{R_1 + R_2 - 2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1 - 1}{m}\right)\left(1 - \frac{R_1 - 1}{m}\right) \\
&= \epsilon + \frac{1}{m} + K_m\left(2 - \frac{R_1 + R_2 - 2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1 + R_2 - 2}{m} + \frac{(R_1 - 1)(R_2 - 1)}{m^2}\right).
\end{aligned}$$

If we let $x = \frac{1}{m}R_1$ and $y = \frac{1}{m}R_2$, then the previous inequality can be written as

$$x \leq \epsilon + \frac{1}{m} + K_m\left(2 + \frac{2}{m} - x - y\right)$$
$$- K_m^2\left(1 + \frac{2}{m} - x - y + \left(x - \frac{1}{m}\right)\left(y - \frac{1}{m}\right)\right),$$

or

$$(x - a_m)(y + b_m) \leq c_m, \tag{2.12}$$

where

$$\begin{aligned} a_m &= 1 + \frac{1}{m} - \frac{1}{K_m}, \\ b_m &= -1 - \frac{1}{m} + \frac{1}{K_m} + \frac{1}{K_m^2}, \\ c_m &= -1 - \frac{2}{m} - \frac{1}{m^2} + \left(2 + \frac{2}{m}\right)\frac{1}{K_m} \\ &\quad + \left(\epsilon + \frac{1}{m}\right)\frac{1}{K_m^2} - a_m b_m. \end{aligned}$$

By symmetry, we can also show

$$(x + b_m)(y - a_m) \leq c_m. \tag{2.13}$$

Note that

$$\begin{aligned} a &:= \lim_{m\to\infty} a_m = 0 \\ b &:= \lim_{m\to\infty} b_m = 1 \\ c &:= \lim_{m\to\infty} c_m = 1 + \epsilon. \end{aligned}$$

Let S_m be the set of all nonnegative x, y that satisfy (2.12) and (2.13) and $\mathscr{S}_m$ be the set of all (mx, my) such that $(x, y) \in S_m$. Then by the arguments of this section, every $(R_1, R_2) \in \mathscr{R}_m$ that satisfies $R_1, R_2 > \log 2f(m)$ is in $\mathscr{S}_m$. As the capacity region is given by the closure of $\mathrm{conv}(\mathscr{R}_m)$, the definition of sum-capacity (2.1) implies

$$\begin{aligned} \frac{1}{m}C_{\mathrm{sum}}^{(m)}(0) &\leq \frac{1}{m}\max_{(R_1,R_2)\in\mathrm{conv}(\mathscr{S}_m)}(R_1 + R_2) \\ &= \max_{(x,y)\in\mathrm{conv}(S_m)}(x + y). \end{aligned}$$

Thus

$$\limsup_{m\to\infty}\frac{1}{m}C_{\mathrm{sum}}^{(m)}(0) \leq \lim_{m\to\infty}\max_{(x,y)\in\mathrm{conv}(S_m)}(x + y). \tag{2.14}$$

To find the limit on the right hand side, we make use of the following lemma.

Lemma 2.5.3. *Suppose* $(a_m)_{m=1}^{\infty}$, $(b_m)_{m=1}^{\infty}$, *and* $(c_m)_{m=1}^{\infty}$ *are three convergent sequences of real numbers with limits* a, b, *and* c, *respectively. The limits satisfy*

$$b, c, a+b, ab+c > 0,$$

and

$$\sqrt{(a+b)^2+4c} > b + \frac{c}{b}.$$

For every positive integer m, *let* S_m *be the set of all nonnegative* x, y *that satisfy (2.12) and (2.13). Then*

$$\lim_{m\to\infty} \max_{(x,y)\in \mathrm{conv}(S_m)} (x+y) = a - b + \sqrt{(a+b)^2+4c}.$$

It is easy to see that the sequences above satisfy the assumptions of Lemma 2.5.3. Thus

$$\limsup_{m\to\infty} \frac{1}{m} C_{\mathrm{sum}}^{(m)}(0) \le \sqrt{5+4\epsilon} - 1,$$

Therefore, for all but finitely many m,

$$C_{\mathrm{sum}}^{(m)}(0) \le (\sqrt{5+4\epsilon} - 1)m.$$

2.6 Proofs

2.6.1 Proof of Lemma 2.3.1

We use the probabilistic method [33]. We assign a probability to every $2^m \times 2^m$ $(0,1)$-matrix and show that the probability of a matrix having both properties is positive for sufficiently large m; hence, there exists at least one such matrix. Fix $\epsilon \in (0,1)$, and let $B_m = (b_{ij})_{i,j=1}^{2^m}$ be a random matrix with $b_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}\,(p)$, where $1-\epsilon < p < 1$. Let

$$\Gamma_m := \left\{ S \middle| S \subseteq [2^m], |S| \ge f(m) \right\}.$$

For every $S, T \in \Gamma_m$, define the event

$$\mathcal{E}_{S,T}^{(m)} := \left\{ \frac{|(S\times T)\cap 1_{B_m}|}{|S|\,|T|} \le 1-\epsilon \right\}.$$

By the union bound,

$$\begin{aligned}
\Pr\left(\bigcup_{S,T\in\Gamma_m}\mathcal{E}_{S,T}^{(m)}\right) &\le \sum_{S,T\in\Gamma_m}\Pr\left(\mathcal{E}_{S,T}^{(m)}\right)\\
&= \sum_{S,T\in\Gamma_m}\Pr\left(\left|(S\times T)\cap 1_{B_m}\right| \le (1-\epsilon)\,|S||T|\right)\\
&= \sum_{S,T\in\Gamma_m}\sum_{k=0}^{\lfloor(1-\epsilon)|S||T|\rfloor}\binom{|S||T|}{k}p^k(1-p)^{|S||T|-k}\\
&= \sum_{i,j=f(m)}^{2^m}\binom{2^m}{i}\binom{2^m}{j}\sum_{k=0}^{\lfloor(1-\epsilon)ij\rfloor}\binom{ij}{k}p^k(1-p)^{ij-k}.
\end{aligned}$$

Suppose $\{X_\ell\}_{\ell=1}^L$ is a set of independent random variables such that for each ℓ, $X_\ell \in [0,1]$ with probability one. If $S=\sum_{\ell=1}^L X_\ell$, from Hoeffding's inequality [34] it follows that for all $u \le \mathbb{E}S$,

$$\Pr(S \le u) \le e^{-\frac{2}{L}(\mathbb{E}S-u)^2}.$$

If $\{X_\ell\}_{\ell=1}^{ij}$ is a set of ij independent Bernoulli(p) random variables, then for every ℓ, $0 \le X_\ell \le 1$, and

$$(1-\epsilon)ij < pij = \mathbb{E}\left[\sum_{l=1}^{ij}X_l\right].$$

Thus Hoeffding's inequality implies

$$\begin{aligned}
\sum_{k=0}^{\lfloor(1-\epsilon)ij\rfloor}\binom{ij}{k}p^k(1-p)^{ij-k} &= \Pr\left(\sum_{l=1}^{ij}X_l \le (1-\epsilon)ij\right)\\
&\le e^{-2(p-1+\epsilon)^2 ij}.
\end{aligned}$$

Since $\binom{2^m}{i} \le 2^{mi}$,

$$\begin{aligned}
&\sum_{i,j=f(m)}^{2^m}\binom{2^m}{i}\binom{2^m}{j}\sum_{k=0}^{\lfloor(1-\epsilon)ij\rfloor}\binom{ij}{k}p^k(1-p)^{ij-k}\\
&\quad\le \sum_{i,j=f(m)}^{2^m}2^{m(i+j)}e^{-2(p-1+\epsilon)^2 ij}\\
&\quad= \sum_{i,j=f(m)}^{2^m}e^{(i+j)m\ln 2-2(p-1+\epsilon)^2 ij}.
\end{aligned}$$

Define $h : \mathbb{Z}^2 \to \mathbb{R}$ as

$$h(i,j) := (i+j)m\ln 2 - 2(p-1+\epsilon)^2 ij.$$

Then for $j \geq f(m)$,

$$\begin{aligned} &h(i+1,j) - h(i,j) \\ &\qquad = m\ln 2 - 2(p-1+\epsilon)^2 j \\ &\qquad \leq m\ln 2 - 2(p-1+\epsilon)^2 f(m) \\ &\qquad = f(m)\left(\frac{m}{f(m)}\ln 2 - 2(p-1+\epsilon)^2\right). \end{aligned}$$

By assumption,

$$\lim_{m\to\infty} \frac{m}{f(m)} = 0,$$

so there exists M_1 such that for all $m > M_1$,

$$\frac{m}{f(m)} < \frac{2}{\ln 2}(p-1+\epsilon)^2.$$

Therefore, for $m > M_1$ and $y \geq f(m)$, h is decreasing with respect to i. As h is symmetric with respect to i and j, for $m > M_1$ and $i \geq f(m)$, we also have $h(i,j+1) - h(i,j) < 0$. Thus h is a decreasing function in i and j for $m > M_1$ and $i,j \geq f(m)$. Hence for $m > M_1$,

$$\begin{aligned} &\sum_{i,j=f(m)}^{2^m} e^{(i+j)m\ln 2 - 2(p-1+\epsilon)^2 ij} \\ &\qquad \leq \left(2^m - f(m) + 1\right)^2 e^{2mf(m)\ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} \\ &\qquad < e^{2m(1+f(m))\ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} \\ &\qquad = e^{2(f(m))^2\left(\left(1+\frac{1}{f(m)}\right)\frac{m}{f(m)}\ln 2 - 2(p-1+\epsilon)^2\right)}. \end{aligned}$$

The exponent of the right hand side of the previous inequality,

$$2\left(f(m)\right)^2\left(\left(1+\frac{1}{f(m)}\right)\frac{m}{f(m)}\ln 2 - 2(p-1+\epsilon)^2\right),$$

goes to $-\infty$ as m approaches infinity. Thus

$$\lim_{m\to\infty} \Pr\left(\bigcup_{S,T\in\Gamma_m} \mathcal{E}_{S,T}^{(m)}\right) = 0.$$

This means that the probability that the fraction of bad entries in a sufficiently large submatrix is less than $1-\epsilon$ is going to zero.

Next, we calculate the probability that B_m doesn't satisfy the second property. For every $i \in [2^m]$ and $j \in [2^{m-g(m)}]$, define the event

$$\mathcal{E}_{i,j}^{(m)} := \left\{ 0_{B_m(i,\mathcal{X}_{m,j})} \cap 0_{B_m} = \varnothing \right\},$$

where

$$\mathcal{X}_{m,j} = \left\{ (j-1)2^{g(m)} + \ell \middle| \ell \in [2^{g(m)}] \right\}.$$

The probability that for every i and j, the set $B_m(i, \mathcal{X}_{m,j})$ doesn't have at least one good entry equals

$$\begin{aligned} \Pr\left(\bigcup_{i,j} \mathcal{E}_{i,j}^{(m)} \right) &\leq \sum_{i=1}^{2^m} \sum_{j=1}^{2^{m-g(m)}} \Pr\left(0_{B_m(i,\mathcal{X}_{j,k})} \cap 0_{B_m} = \varnothing \right) \\ &= 2^{2m-g(m)} p^{2^{g(m)}} \\ &= 2^{2^{g(m)}\left(\frac{2m-g(m)}{2^{g(m)}} + \log p \right)}. \end{aligned}$$

Since $m = o(2^{g(m)})$, the exponent of the right hand side of the previous inequality,

$$2^{g(m)} \left(\frac{2m - g(m)}{2^{g(m)}} + \log p \right),$$

goes to $-\infty$ as $m \to \infty$. This implies that

$$\Pr\left(\bigcup_{i,j} \mathcal{E}_{i,j}^{(m)} \right)$$

goes to zero as $m \to \infty$. Similarly, the probability that there exists (i, j) such that $B_m(\mathcal{X}_{m,j}, i)$ doesn't have at least one good entry goes to zero as m tends to infinity. Thus, by the union bound, the probability that the matrix doesn't satisfy either of these properties is going to zero. Therefore, for large enough m, almost every $(0, 1)$-matrix satisfies all the required properties, though we only need one such matrix.

2.6.2 Proof of Lemma 2.5.1

We first prove (2.9). If $\alpha = 0$, then $q(x) = 0$ for every $x \in T$ and both sides equal zero. Otherwise,

$$\begin{aligned} -\sum_{x \in T} q(x) \log q(x) &= -\alpha \sum_{x \in T} \frac{q(x)}{\alpha} \log \frac{q(x)}{\alpha} - \alpha \log \alpha \\ &\leq \alpha \log |T| - \alpha \log \alpha, \end{aligned}$$

since $q(x)/\alpha$ is a probability mass function with entropy $\sum_{x\in T} \frac{q(x)}{\alpha} \log \frac{\alpha}{q(x)}$ and alphabet size $|T|$.

We next prove (2.10). If

$$q(x) = p(x)\mathbf{1}_T(x),$$

then by the previous inequality,

$$\begin{aligned}
-\sum_{x\in T} p(x)\log p(x) &= -\sum_{x\in T} q(x)\log q(x) \\
&\le \alpha \log |T| - \alpha \log \alpha,
\end{aligned}$$

where

$$\alpha = \sum_{x\in T} q(x).$$

Similarly, replacing $\mathcal{X} \setminus T$ with T results in

$$\begin{aligned}
&-\sum_{x\in\mathcal{X}\setminus T} p(x)\log p(x) \\
&\quad \le (1-\alpha)\log|\mathcal{X}\setminus T| - (1-\alpha)\log(1-\alpha).
\end{aligned}$$

Adding the previous two inequalities gives

$$\begin{aligned}
H(X) &\le \alpha \log |T| + (1-\alpha)\log|\mathcal{X}\setminus T| + H(\alpha) \\
&\le \alpha \log |T| + (1-\alpha)\log|\mathcal{X}| + 1.
\end{aligned}$$

Therefore,

$$\frac{H(X)}{\log|\mathcal{X}|} \le 1 + \frac{1}{\log|\mathcal{X}|} - \left(1 - \frac{\log|T|}{\log|\mathcal{X}|}\right)\alpha,$$

and

$$\alpha \le \frac{1 - \frac{H(X)-1}{\log|\mathcal{X}|}}{1 - \frac{\log|T|}{\log|\mathcal{X}|}}.$$

2.6.3 Proof of Lemma 2.5.2

Let k be the cardinality of the range of $p : \mathcal{X} \to \mathbb{R}$. Then there exists a sequence $(x_j)_{j=1}^k$ such that

$$\{p(x) | x \in \mathcal{X}\} = \{p(x_j) | j \in [k]\},$$

and

$$0 < p(x_1) < \cdots < p(x_k) \le 1.$$

For $j \in [k]$, define

$$S_j := \{x \in \mathcal{X} | p(x) \geq p(x_j)\},$$

and let $S_{k+1} := \varnothing$. Then for $j \in [k]$, $S_{j+1} \subseteq S_j$, and

$$S_j \setminus S_{j+1} = \{x \in \mathcal{X} | p(x) = p(x_j)\} \neq \varnothing.$$

Thus the number of $x \in \mathcal{X}$ such that $p(x) = p(x_j)$ equals $|S_j \setminus S_{j+1}|$. For $j \in \{2, \ldots, k\}$, define

$$\alpha_j := |S_j| \Big(p(x_j) - p(x_{j-1}) \Big),$$

and set $\alpha_1 := |S_1| p(x_1)$. We have

$$\begin{aligned} \sum_{j=1}^{k} \alpha_j \frac{\mathbf{1}_{S_j}(x)}{|S_j|} &= \sum_{j=1}^{k} \left(p(x_j) - p(x_{j-1}) \right) \mathbf{1}_{S_j}(x) \\ &= \sum_{j=1}^{k} p(x_j) \mathbf{1}_{S_j \setminus S_{j+1}}(x) \\ &= p(x). \end{aligned}$$

This proves (a) and (b).

In (c), if the set

$$\Big\{ j \in [k] \Big| |S_j| \leq C \Big\}$$

is empty, then there's nothing to prove. Otherwise, it's a nonempty subset of $[k]$ and thus has a minimum, which we call j^*. Then

$$\begin{aligned} \sum_{j : |S_j| \leq C} \alpha_j &= \sum_{j=j^*}^{k} \alpha_j \\ &= \sum_{j=j^*}^{k} |S_j| \big(p(x_j) - p(x_{j-1}) \big) \\ &= \sum_{j=j^*}^{k} |S_j \setminus S_{j+1}| p(x_j) - |S_{j^*}| p(x_{j^*-1}) \\ &= \sum_{x \in S_{j^*}} p(x) - |S_{j^*}| p(x_{j^*-1}) \\ &\leq \sum_{x \in S_{j^*}} p(x). \end{aligned}$$

By (2.10),

$$\sum_{x\in S_{j^*}} p(x) \le \frac{1}{1-\frac{\log|S_{j^*}|}{\log|\mathcal{X}|}}\left(1-\frac{H(X)-1}{\log|\mathcal{X}|}\right)$$
$$\le \frac{1}{1-\frac{\log C}{\log|\mathcal{X}|}}\left(1-\frac{H(X)-1}{\log|\mathcal{X}|}\right),$$

since $|S_{j^*}| \le C$.

2.6.4 Proof of Lemma 2.5.3

Prior to proving Lemma 2.5.3, we state and prove the following lemma.

Lemma 2.6.1. *Suppose $a \in \mathbb{R}$ and*

$$b, c, a+b, ab+c \in \mathbb{R}_{>0}.$$

In addition,

$$\sqrt{(a+b)^2+4c} > b + \frac{c}{b}.$$

Let S be the set of all pairs (x, y) such that $x, y \ge 0$, and

$$\begin{cases}(x-a)(y+b) & \le c\\ (x+b)(y-a) & \le c.\end{cases}$$

If x_0 is the unique positive solution to the equation

$$(x-a)(x+b) = c,$$

then

$$\max_{(x,y)\in\mathrm{conv}(S)} (x+y) = 2x_0.$$

Proof. Since

$$(x-a)(y+b) - (x+b)(y-a) = (a+b)(x-y)$$

and $a+b>0$, the set S can be written as $S = S_1 \cup S_2$ (Figure 2.2), where S_1 is the set of all pairs (x, y) such that $0 \le x \le y$ and

$$(x+b)(y-a) \le c,$$

and S_2 is the set of all pairs (x, y) such that $0 \le y \le x$ and

$$(x-a)(y+b) \le c.$$

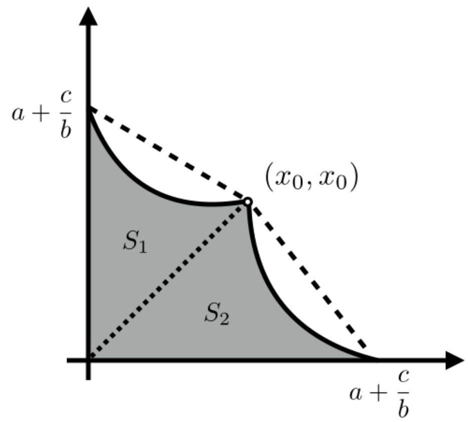

Figure 2.2: The sets S_1 and S_2 (gray area), and their convex hulls.

The intersection of S_1 and S_2 consists of all pairs (x, x) such that $0 \le x \le x_0$, where

$$x_0 = \frac{a - b + \sqrt{(a+b)^2 + 4c}}{2}.$$

Note that since b, c and $ab + c$ are positive,

$$\sqrt{(a+b)^2 + 4c} < a + b + \frac{2c}{b},$$

so $0 < x_0 < a + \frac{c}{b}$. The convex hull of S_1 consists of all pairs (x, y) such that $0 \le x \le y$ and

$$\left(a + \frac{c}{b} - x_0\right) x + x_0 y \le \left(a + \frac{c}{b}\right) x_0,$$

and the convex hull of S_2 consists of all pairs (x, y) such that $0 \le y \le x$ and

$$x_0 x + \left(a + \frac{c}{b} - x_0\right) y \le \left(a + \frac{c}{b}\right) x_0.$$

Note that $\operatorname{conv}(S_1) \cup \operatorname{conv}(S_2)$ is the region bounded by the axes $y = 0$, $x = 0$, and the piecewise linear function

$$h(x) = \begin{cases} \frac{x_0 - a - \frac{c}{b}}{x_0} x + a + \frac{c}{b} & 0 \le x \le x_0, \\ \frac{x_0}{x_0 - a - \frac{c}{b}} x - \frac{\left(a + \frac{c}{b}\right) x_0}{x_0 - a - \frac{c}{b}} & x_0 < x \le a + \frac{c}{b}. \end{cases}$$

Since $2x_0 \ge a + \frac{c}{b}$ by assumption,

$$\frac{x_0 - a - \frac{c}{b}}{x_0} \ge \frac{x_0}{x_0 - a - \frac{c}{b}}.$$

This means the slope of h is decreasing, or equivalently, h is a concave function. Thus $\operatorname{conv}(S_1) \cup \operatorname{conv}(S_2)$ is convex. But

$$S \subseteq \operatorname{conv}(S_1) \cup \operatorname{conv}(S_2) \subseteq \operatorname{conv}(S),$$

so

$$\mathrm{conv}(S) = \mathrm{conv}(S_1) \cup \mathrm{conv}(S_2).$$

This implies

$$\max_{(x,y)\in\mathrm{conv}(S)} (x+y) = 2x_0.$$

□

Using this lemma, we may prove Lemma 2.5.3. There exists a positive M such that for every $m \geq M$,

$$b_m, c_m, a_m + b_m, a_m b_m + c_m > 0$$

and

$$\sqrt{(a_m+b_m)^2 + 4c_m} - b_m - \frac{c_m}{b_m} > 0.$$

Let $x_0^{(m)}$ and x_0 be the unique positive solutions to the equations

$$(x_0^{(m)} - a_m)(x_0^{(m)} + b_m) = c_m$$

and

$$(x_0 - a)(x_0 + b) = c.$$

Since $x_0^{(m)}$ and x_0 are continuous functions of (a_m, b_m, c_m) and (a, b, c), respectively, we have

$$\lim_{m\to\infty} x_0^{(m)} = x_0.$$

Thus by Lemma 2.6.1,

$$\begin{aligned}
\lim_{m\to\infty} \max_{(x,y)\in\mathrm{conv}\left(S^{(m)}\right)} (x+y) &= \lim_{m\to\infty} 2x_0^{(m)} \\
&= 2x_0 \\
&= a - b + \sqrt{(a+b)^2 + 4c}.
\end{aligned}$$

Chapter 3

THE RATE BENEFIT

In the classical k-user multiple access channel (MAC) [7], there are k encoders and a single decoder. Each encoder has a private message which it transmits over n channel uses to the decoder. The decoder, once it receives n output symbols, finds the messages of all k encoders with small average probability of error. In this model, the encoders cannot cooperate, since each encoder only has access to its own message.

We now consider an alternative scenario where our k-user MAC is part of a larger network. In this network, there is a node that is connected to all k encoders and acts as a "cooperation facilitator" (CF). Specifically, for every $j \in [k]$, there is a link of capacity $C_{\text{in}}^j \geq 0$ going from encoder j to the CF and a link of capacity $C_{\text{out}}^j \geq 0$ going back. The CF helps the encoders exchange information before they transmit their codewords over the MAC. Figure 3.1 depicts a network consisting of a k-user MAC and a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, where $\mathbf{C}_{\text{in}} = (C_{\text{in}}^j)_{j\in[k]}$ and $\mathbf{C}_{\text{out}} = (C_{\text{out}}^j)_{j\in[k]}$ denote the capacities of the CF input and output links. In this figure, $X_{[k]}^n = (X_1^n, \ldots, X_k^n)$ is the vector of the channel inputs of the k encoders, and $\widehat{w}_{[k]} = (\widehat{w}_1, \ldots, \widehat{w}_k)$ is the vector of message reproductions at the decoder.

The main result of this chapter (Theorem 3.2.3) determines a set of MACs where the benefit of encoder cooperation through a CF grows very quickly with $\mathbf{C}_{\text{out}}$. Specifically, we find a class of MACs $\mathcal{C}^*$, where every MAC in $\mathcal{C}^*$ has the property that for any fixed $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^k$, the sum-capacity of that MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF has an infinite derivative in the direction of every $\mathbf{v} \in \mathbb{R}_{>0}^k$ at $\mathbf{C}_{\text{out}} = \mathbf{0}$. In other words, as a function of $\mathbf{C}_{\text{out}}$, the sum-capacity grows faster than any function with bounded derivative at $\mathbf{C}_{\text{out}} = \mathbf{0}$. This means that for any MAC in $\mathcal{C}^*$, sharing a small number of bits with each encoder leads to a large gain in sum-capacity.

An important implication of this result is the existence of a memoryless network that does not satisfy the strong edge removal property (Chapter 1, Sec-

This material is based upon work supported by the National Science Foundation under Grant Numbers 1527524, 1526771, and 1321129. It originally appears in [35].

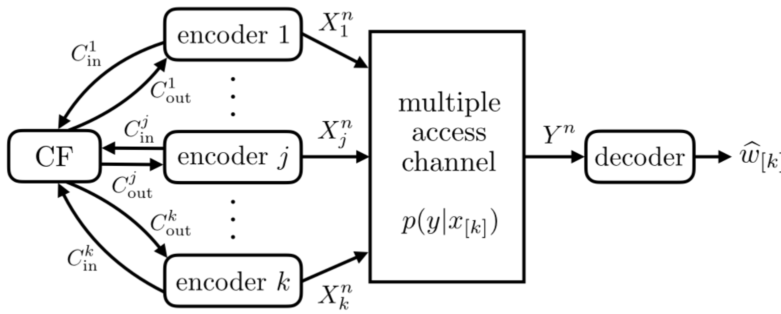

Figure 3.1: The network consisting of a k-user MAC and a CF. For $j \in [k]$, encoder j has access to message $w_j \in [2^{nR_j}]$, which is omitted in this figure.

tion 1.4). Recall that a network satisfies the strong edge removal property if removing an edge of capacity $\delta > 0$ decreases sum-capacity by at most a linear function of δ. Now consider a network consisting of a MAC in $\mathcal{C}^*$ and a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, where $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^k$. Our main result in this chapter (Theorem 3.2.3) implies that for small $\mathbf{C}_{\text{out}}$, removing all the output edges reduces sum-capacity by an amount much larger than any linear function of $\mathbf{C}_{\text{out}}$. Thus there exist memoryless networks that do not satisfy the edge removal property.

We introduce the coding scheme that leads to Theorem 3.2.3 in Section 3.3. This scheme combines forwarding, coordination, and classical MAC coding. In forwarding, each encoder sends part of its message to all other encoders by passing that information through the CF.[1] The coordination strategy is a modified version of Marton's coding scheme for the broadcast channel [8], [37]. To implement this strategy, the CF shares information with the encoders that enables them to transmit codewords that are jointly typical with respect to a *dependent* distribution; this is proven using a multivariate version of the covering lemma [38, p. 218]. The multivariate covering lemma is stated for strongly typical sets in [38]. In Appendix A, using the proof for the 2-user case from [38] and techniques from [39], we prove this lemma for weakly typical sets [40, p. 251]. Using weakly typical sets in our achievability proof allows our results to extend to continuous (e.g., Gaussian) channels without the need for quantization. Finally, the classical MAC strategy is Ulrey's [7] extension of Ahlswede's [2], [3] and Liao's [4] coding strategy to the k-user MAC.

[1]While it is possible to consider encoders that send *different* parts of their messages to different encoders using Han's result for the MAC with correlated sources [36], we avoid these cases for simplicity.

Using techniques from Willems [5], we derive an outer bound (Proposition 3.2.5) for the capacity region of the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. This outer bound does not capture the dependence of the capacity region on $\mathbf{C}_{\text{out}}$ and is thus loose for some values of $\mathbf{C}_{\text{out}}$. However, if the entries of $\mathbf{C}_{\text{out}}$ are sufficiently larger than the entries of $\mathbf{C}_{\text{in}}$, then our inner and outer bounds agree and we obtain the capacity region (Corollary 3.2.6).

In Section 3.4, we apply our results to the 2-user Gaussian MAC with a CF that has access to the messages of both encoders and has links of output capacity C_{out}. We show that for small C_{out}, the achievable sum-rate approximately equals a constant times $\sqrt{C_{\text{out}}}$. A similar approximation holds for a weighted version of the sum-rate as well, as we see in Proposition 3.4.2. This result implies that at least for the 2-user Gaussian MAC, the benefit of cooperation is not limited to sum-capacity and applies to other capacity region metrics as well.

In Appendix 3.6, we consider the extension of Willems' conferencing model [5] from 2 to k users. We apply our outer bound for the k-user MAC with a CF to obtain an outer bound for the k-user MAC with conferencing. The resulting outer bound is tight when $k = 2$ and coincides with the conferencing capacity region [5] . A special case of this model with $k = 3$ is studied in [41] for the Gaussian MAC.

In the next section, we formally define the capacity region of the network consisting of a k-user MAC and a CF.

3.1 Cooperation over the k-user MAC

Consider a network with k encoders, a CF, a k-user MAC, and a decoder (Figure 3.1). For each $j \in [k]$, encoder j communicates with the CF using noiseless links of capacities $C_{\text{in}}^j \geq 0$ and $C_{\text{out}}^j \geq 0$ going to and from the CF, respectively. The k encoders communicate with the decoder through a MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$, where

$$\mathcal{X}_{[k]} = \prod_{j \in [k]} \mathcal{X}_j,$$

and an element of $\mathcal{X}_{[k]}$ is denoted by $x_{[k]}$. We say a MAC is discrete if $\mathcal{X}_{[k]}$ and $\mathcal{Y}$ are either finite or countably infinite, and $p(y|x_{[k]})$ is a probability mass function on $\mathcal{Y}$ for every $x_{[k]} \in \mathcal{X}_{[k]}$. We say a MAC is continuous if for some positive integers $\ell_1, \ldots, \ell_{k+1}$, $\mathcal{X}_j = \mathbb{R}^{\ell_j}$ for $j \in [k]$, $\mathcal{Y} = \mathbb{R}^{\ell_{k+1}}$, and $p(y|x_{[k]})$ is a probability density function on $\mathcal{Y}$ for all $x_{[k]}$. In addition, we assume that our channel is memoryless and stationary, so that for every positive integer n, the nth extension channel of our MAC is given by $p(y^n|x_{[k]}^n)$, where

$$\forall (x_{[k]}^n, y^n) \in \mathcal{X}_{[k]}^n \times \mathcal{Y}^n : p(y^n|x_{[k]}^n) = \prod_{t=1}^{n} p(y_t|x_{[k]t}).$$

An example of a continuous MAC is the k-user Gaussian MAC with noise variance $N > 0$, where

$$p(y|x_{[k]}) = \frac{1}{\sqrt{2\pi N}} \exp\Big[-\frac{1}{2N}\Big(y - \sum_{j\in[k]} x_j\Big)^2\Big]. \tag{3.1}$$

Henceforth, all MACs are memoryless and stationary and either discrete or continuous.

We next describe a

$$\big((2^{nR_1}, \ldots, 2^{nR_k}), n, L\big)\text{-code}$$

for the MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF with cost functions $(b_j)_{j\in[k]}$ and cost constraint vector $\mathbf{B} = (B_j)_{j\in[k]} \in \mathbb{R}_{\geq 0}^k$. For each $j \in [k]$, cost function b_j is a fixed mapping from $\mathcal{X}_j$ to $\mathbb{R}_{\geq 0}$. Each encoder $j \in [k]$ wishes to transmit a message $w_j \in [2^{nR_j}]$ to the decoder. This is accomplished by first exchanging information with the CF and then transmitting across the MAC. Communication with the CF occurs in L rounds. For each $j \in [k]$ and $\ell \in [L]$, sets $\mathcal{U}_{j\ell}$ and $\mathcal{V}_{j\ell}$, respectively, describe the alphabets of symbols that encoder j can send to and receive from the CF in round ℓ. These alphabets satisfy the link capacity constraints

$$\begin{aligned}\sum_{\ell\in[L]} \log|\mathcal{U}_{j\ell}| &\leq nC_{\text{in}}^j \\ \sum_{\ell\in[L]} \log|\mathcal{V}_{j\ell}| &\leq nC_{\text{out}}^j.\end{aligned} \tag{3.2}$$

The operation of encoder j and the CF, respectively, in round ℓ are given by

$$\begin{aligned}\varphi_{j\ell} &: [2^{nR_j}] \times \mathcal{V}_j^{\ell-1} \to \mathcal{U}_{j\ell} \\ \psi_{j\ell} &: \prod_{i\in[k]} \mathcal{U}_i^{\ell} \to \mathcal{V}_{j\ell},\end{aligned}$$

where $\mathcal{U}_j^\ell = \prod_{\ell'=1}^{\ell} \mathcal{U}_{j\ell'}$ and $\mathcal{V}_j^\ell = \prod_{\ell'=1}^{\ell} \mathcal{V}_{j\ell'}$. After its exchange with the CF, encoder j applies a function

$$f_j : [2^{nR_j}] \times \mathcal{V}_j^L \to \mathcal{X}_j^n,$$

to choose a codeword, which it transmits across the channel. In addition, every x_j^n in the range of f_j satisfies the cost constraint

$$\sum_{t\in[n]} b_j(x_{jt}) \leq nB_j.$$

The decoder receives channel output Y^n and applies

$$g : \mathcal{Y}^n \to \prod_{j\in[k]} [2^{nR_j}]$$

to obtain estimate $\hat{W}_{[k]}$ of the message vector $w_{[k]}$.

The encoders, CF, and decoder together define a

$$\big((2^{nR_1}, \ldots, 2^{nR_k}), n, L\big)\text{-code.}$$

The average error probability of the code is defined by

$$P_e^{(n)} := \Pr\big\{g(Y^n) \neq W_{[k]}\big\},$$

where $W_{[k]}$ is the transmitted message vector and is uniformly distributed on $\prod_{j\in[k]}[2^{nR_j}]$. A rate vector $R_{[k]} = (R_1, \ldots, R_k)$ is *achievable* if there exists a sequence of $\big((2^{nR_1}, \ldots, 2^{nR_k}), n, L\big)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region, $\mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, is defined as the closure of the set of all achievable rate vectors.

3.2 Inner and Outer Bounds

In this section, we describe the key results. In Subsection 3.2.1, we present our inner bound. In Subsection 3.2.2, we state the main result of this chapter, which proves the existence of a class of MACs with large cooperation gain. Finally, in Subsection 3.2.3, we discuss our outer bound.

3.2.1 Inner Bound

Using the coding scheme we introduce in Section 3.3, we obtain an inner bound for the capacity region of the k-user MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. The following definitions are useful for describing that bound. Choose vectors $\mathbf{C}_0 = (C_{j0})_{j=1}^k$ and $\mathbf{C}_d = (C_{jd})_{j=1}^k$ in $\mathbb{R}_{\geq 0}^k$ such that for all $j \in [k]$,

$$C_{j0} \leq C_{\text{in}}^j \tag{3.3}$$

$$C_{jd} + \sum_{i\neq j} C_{i0} \leq C_{\text{out}}^j. \tag{3.4}$$

Here C_{j0} is the number of bits per channel use encoder j sends directly to the other encoders via the CF and C_{jd} is the number of bits per channel use the CF transmits to encoder j to implement the coordination strategy. Subscript "d" in C_{jd} alludes to the dependence created through coordination. The set

$$S_d(\mathbf{C}_d) := \Big\{ j \in [k] \Big| C_{jd} \neq 0 \Big\}$$

describes the encoders that participate in this dependence. We denote $S_d(\mathbf{C}_d)$ with S_d when $\mathbf{C}_d$ is clear from context.

Fix alphabets $\mathcal{U}_0, \mathcal{U}_1, \ldots, \mathcal{U}_k$. For every nonempty $S \subseteq [k]$, let $\mathcal{U}_S$ be the set of all $u_S = (u_j)_{j\in S}$ where $u_j \in \mathcal{U}_j$ for all $j \in S$. Define the set $\mathcal{X}_S$ similarly. Let $\mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$ be the set of all distributions on $\mathcal{U}_0 \times \mathcal{U}_{[k]} \times \mathcal{X}_{[k]}$ that are of the form

$$p(u_0) \cdot \prod_{i\in S_d^c} p(u_i|u_0) \cdot p(u_{S_d}|u_0, u_{S_d^c}) \cdot \prod_{j\in[k]} p(x_j|u_0, u_j), \tag{3.5}$$

satisfy the dependence constraints[2]

$$\zeta_S := \sum_{j\in S} C_{jd} - \sum_{j\in S} H(U_j|U_0) + H(U_S|U_0, U_{S_d^c}) > 0 \qquad \forall\, \varnothing \subsetneq S \subseteq S_d,$$

and cost constraints

$$\mathbb{E}\big[b_j(X_j)\big] \leq B_j \qquad \forall j \in [k]. \tag{3.6}$$

Here U_0 encodes the "common message," which, for every $j \in [k]$, contains nC_{j0} bits from the message of encoder j and is shared with all other encoders through the CF; each random variable U_j captures the information encoder j receives from the CF to create dependence with the codewords of other encoders. The random variable X_j represents the symbol encoder j transmits over the channel.

We next state our inner bound for the k-user MAC with encoder cooperation via a CF. The coding strategy that achieves this inner bound uses only a single round of cooperation ($L = 1$) and is given in Section 3.3. The error analysis is presented in Subsection 3.5.1.

[2]The constraint on ζ_S is imposed by the multivariate covering lemma (Appendix A), which we use in the proof of our inner bound.

Theorem 3.2.1 (Inner Bound)**.** *Consider a MAC* $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$. *Fix* $\mathbf{C}_0, \mathbf{C}_d \in \mathbb{R}^k_{\geq 0}$ *satisfying (3.3) and (3.4) and* $p \in \mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$. *If the rate vector* $R_{[k]} := (R_1, \ldots, R_k)$ *satisfies*

$$\sum_{j\in[k]} R_j < I(X_{[k]}; Y) - \zeta_{S_d}, \tag{3.7}$$

and for every $S, T \subseteq [k]$ *there exist sets* A *and* B *such that* $S \cap S_d^c \subseteq A \subseteq S$, $S^c \cap S_d^c \subseteq B \subseteq S^c$, *and*

$$\begin{aligned} &\sum_{j\in A}(R_j - C_{j0})^+ + \sum_{j\in B\cap T}(R_j - C_{\text{in}}^j)^+ \\ &\qquad < I\big(U_A, X_{A\cup(B\cap T)}; Y \big| U_0, U_B, X_{B\setminus T}\big) - \zeta_{(A\cup B)\cap S_d}, \end{aligned} \tag{3.8}$$

then $R_{[k]} \in \mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$.[3]

The next corollary treats the case where the CF transmits the bits it receives from each encoder to all other encoders without change. In this case, our coding strategy simply combines forwarding with classical MAC encoding. We obtain this result from Theorem 3.2.1 by setting, for all $j \in [k]$, $C_{jd} = 0$ and $|\mathcal{U}_j| = 1$, in addition to choosing $A = S$ and $B = S^c$ for every $S, T \subseteq [k]$ in (3.8). In Corollary 3.2.2, $\mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$ is the set of all distributions $p(u_0)\prod_{j\in[k]} p(x_j|u_0)$ that satisfy the cost constraints (3.6).

Corollary 3.2.2 (Forwarding Inner Bound)**.** *For any MAC,* $\mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ *contains the set of all rate vectors that for some* $\mathbf{C}_0 \in \mathbb{R}^k_{\geq 0}$ *(satisfying (3.3) and (3.4) with* $\mathbf{C}_d = \mathbf{0}$*) and distribution* $p \in \mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$, *satisfy*

$$\begin{aligned} &\sum_{j\in S} R_j < I\big(X_S; Y | U_0, X_{S^c}\big) + \sum_{j\in S} C_{j0} \qquad \forall\, \varnothing \neq S \subseteq [k] \\ &\sum_{j\in[k]} R_j < I(X_{[k]}; Y). \end{aligned}$$

3.2.2 Sum-Capacity Gain

We wish to understand when cooperation leads to a benefit that exceeds the resources employed to enable it. Therefore, we compare the gain in sum-capacity obtained through cooperation to the number of bits shared with the encoders to enable that gain. Formally, for any k-user MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, define the sum-capacity as

$$C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) := \max_{\mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})} \sum_{j\in[k]} R_j.$$

[3] For every real number x, $x^+ := \max\{x, 0\}$.

Let $\mathcal{X}_1, \ldots, \mathcal{X}_k, \mathcal{Y}$ be finite sets. Fix cost functions $(b_j)_{j\in[k]}$ and cost constraints $(B_j)_{j\in[k]}$, and let $\mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})$ be the set of all independent distributions

$$p(x_{[k]}) = \prod_{j\in[k]} p(x_j)$$

on $\mathcal{X}_{[k]}$ that satisfy the cost constraints (3.6). Similarly, $\mathcal{P}(\mathcal{X}_{[k]})$ is the set of *all* distributions on $\mathcal{X}_{[k]}$ that satisfy (3.6). We say a MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ is in $\mathcal{C}^*$, if for some $p_{\text{ind}} \in \mathcal{P}_{\text{ind}}(X_{[k]})$ that satisfies

$$I_{\text{ind}}(X_{[k]};Y) = \max_{p\in\mathcal{P}_{\text{ind}}(X_{[k]})} I(X_{[k]};Y),$$

there exists $p_{\text{dep}} \in \mathcal{P}(X_{[k]})$ whose support is contained in the support of p_{ind} and satisfies

$$I_{\text{dep}}(X_{[k]};Y) + D\big(p_{\text{dep}}(y)\|p_{\text{ind}}(y)\big) > I_{\text{ind}}(X_{[k]};Y). \tag{3.9}$$

In the above equation, $p_{\text{dep}}(y)$ and $p_{\text{ind}}(y)$ are the output distributions corresponding to the input distributions $p_{\text{dep}}(x_{[k]})$ and $p_{\text{ind}}(x_{[k]})$, respectively. We remark that (3.9) is equivalent to

$$\mathbb{E}_{\text{dep}}\Big[D\big(p(y|X_{[k]})\|p_{\text{ind}}(y)\big)\Big] > \mathbb{E}_{\text{ind}}\Big[D\big(p(y|X_{[k]})\|p_{\text{ind}}(y)\big)\Big],$$

where the expectations are with respect to $p_{\text{dep}}(x_{[k]})$ and $p_{\text{ind}}(x_{[k]})$, respectively.

Using these definitions, we state the main result of this chapter which captures a family of MACs for which the slope of the gain function is infinite in every direction at $\mathbf{C}_{\text{out}} = \mathbf{0}$. The proof appears in Subsection 3.5.2.

Theorem 3.2.3 (Sum-capacity). *Consider a finite alphabet MAC in $\mathcal{C}^*$ and fix $\mathbf{C}_{\text{in}} \in \mathbb{R}^k_{>0}$. Then for any vector $\mathbf{v} \in \mathbb{R}^k_{>0}$,*

$$\lim_{h\to 0^+} \frac{C_{\text{sum}}(\mathbf{C}_{\text{in}}, h\mathbf{v}) - C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{0})}{h} = \infty. \tag{3.10}$$

For continuous MACs, when $b_j(x) = |x|^2$ for all $j \in [k]$, cost constraints are referred to as power constraints. In addition, for every $j \in [k]$, the variable P_j is commonly used instead of B_j. Our next proposition states that the Gaussian MAC satisfies (3.10). The proof appears in Subsection 3.5.3.

Proposition 3.2.4. *For the k-user Gaussian MAC with power constraint vector $\mathbf{P} = (P_j)_{j\in[k]} \in \mathbb{R}^k_{>0}$, defined by (3.1), (3.10) holds.*

3.2.3 Outer Bound

We next describe our outer bound. While we only make use of a single round of cooperation in our inner bound (Theorem 3.2.1), the outer bound applies to all coding schemes regardless of the number of rounds.

Proposition 3.2.5 (Outer Bound)**.** *For any MAC, $\mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is a subset of the closure of the set of all rate vectors that for some distribution $p \in \mathcal{P}_{\text{ind}}(\mathcal{U}_0, \mathcal{X}_{[k]})$ satisfy*

$$\sum_{j\in S} R_j \leq I\big(X_S; Y|U_0, X_{S^c}\big) + \sum_{j\in S} C_{\text{in}}^j \qquad \forall\, \varnothing \neq S \subseteq [k] \tag{3.11}$$

$$\sum_{j\in[k]} R_j \leq I(X_{[k]}; Y). \tag{3.12}$$

The proof of this proposition appears in Subsection 3.5.4. Our proof uses ideas similar to the proof of the converse for the 2-user MAC with conferencing [5].

If the capacities of the CF output links are sufficiently large, our inner and outer bounds coincide and we obtain the capacity region. This follows by setting $C_{j0} = C_{\text{in}}^j$ for all $j \in [k]$ in our forwarding inner bound (Corollary 3.2.2) and comparing it with the outer bound given in Proposition 3.2.5.

Corollary 3.2.6. *For the MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$ with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, if*

$$\forall j \in [k] : C_{\text{out}}^j \geq \sum_{i:i\neq j} C_{\text{in}}^i,$$

then our inner and outer bounds agree.

3.3 Coding Strategy

Choose nonnegative constants $(C_{j0})_{j=1}^k$ and $(C_{jd})_{j=1}^k$ such that (3.3) and (3.4) hold for all $j \in [k]$. Fix a distribution $p \in \mathcal{P}(\mathcal{U}_0, \mathcal{U}_{[k]}, \mathcal{X}_{[k]}, S_d)$ and constants $\epsilon, \delta > 0$. Let

$$\begin{aligned} R_{j0} &= \min\{R_j, C_{j0}\} \\ R_{jd} &= \min\{R_j, C_{\text{in}}^j\} - R_{j0} \\ R_{jj} &= R_j - R_{j0} - R_{jd} = (R_j - C_{\text{in}}^j)^+. \end{aligned}$$

For every $j \in [k]$, split the message of encoder j as $w_j = (w_{j0}, w_{jd}, w_{jj})$, where $w_{j0} \in [2^{nR_{j0}}]$, $w_{jd} \in [2^{nR_{jd}}]$, $w_{jj} \in [2^{nR_{jj}}]$. For all $j \in [k]$, encoder j sends (w_{j0}, w_{jd}) noiselessly to the CF. This is possible since $R_{j0} + R_{jd}$ is less than or equal to C_{in}^j. The CF sends w_{j0} to all other encoders via its output links and uses w_{jd} to implement the coordination strategy to be descibed below. Due to the CF rate constraints, encoder j cannot share the remaining part of its message, w_{jj}, with the CF. Instead, it transmits w_{jj} over the channel using the classical MAC strategy.

Let $\mathcal{W}_0 := \prod_{j\in[k]}[2^{nR_{j0}}]$. For every $w_0 \in \mathcal{W}_0$, let $U_0^n(w_0)$ be drawn independently according to

$$\Pr\left\{U_0^n(w_0) = u_0^n\right\} := \prod_{t=1}^{n} p(u_{0t}).$$

Given $U_0^n(w_0) = u_0^n$, for every $j \in [k]$, $w_{jd} \in [2^{nR_{jd}}]$, and $z_j \in [2^{nC_{jd}}]$, let $U_j^n(w_{jd}, z_j|u_0^n)$ be drawn independently according to

$$\Pr\left\{U_j^n(w_{jd}, z_j|u_0^n) := u_j^n \middle| U_0^n(w_0) = u_0^n\right\} = \prod_{t=1}^{n} p(u_{jt}|u_{0t}). \tag{3.13}$$

For $j \in [k]$, we denote a mapping from $[2^{nC_{jd}}]$ to $\mathcal{U}_j^n$ by μ_j^n.[4] For every $(w_1, \ldots, w_k)$, define $\mathcal{E}(u_0^n, \mu_1, \ldots, \mu_k)$ as the event where $U_0^n(w_0) = u_0^n$ and for every $j \in [k]$,

$$U_j^n(w_{jd}, \cdot|u_0^n) = \mu_j^n(\cdot). \tag{3.14}$$

Let $\mathcal{A}(u_0^n, \mu_{[k]}^n) \subseteq \prod_{j\in[k]}[2^{nC_{jd}}]$ be the set

$$\mathcal{A}(u_0^n, \mu_{[k]}^n) := \left\{(z_1, \ldots, z_k) \middle| \big(u_0^n, \mu_1(z_1), \ldots, \mu_k(z_k)\big) \in A_\delta^{(n)}(U_0, U_{[k]})\right\}, \tag{3.15}$$

where $A_\delta^{(n)}(U_0, U_{[k]})$ is the weakly typical set with respect to the distribution $p(u_0, u_{[k]})$. If $\mathcal{A}(u_0^n, \mu_{[k]}^n)$ is empty, set $Z_j = 1$ for all $j \in [k]$. Otherwise, let the k-tuple $Z_{[k]} = (Z_1, \ldots, Z_k)$ be the smallest element of $\mathcal{A}(u_0^n, \mu_{[k]}^n)$ with respect to the lexicographical order. Finally, given $U_0^n(w_0) = u_0^n$ and $U_j^n(w_{jd}, Z_j|u_0^n) = u_j^n$, for each $w_{jj} \in [2^{nR_{jj}}]$, let $X_j^n(w_{jj}|u_0^n, u_j^n)$ be a random vector drawn independently according to

$$\begin{aligned}\Pr&\left\{X_j^n(w_{jj}|u_0^n, u_j^n) = x_j^n \middle| U_0^n(w_0) = u_0^n, U_j^n(w_{jd}, Z_j) = u_j^n\right\}\\ &:= \prod_{t=1}^{n} p(x_{jt}|u_{0t}, u_{jt}).\end{aligned}$$

We next describe the encoding and decoding processes.

[4] Note the difference between u_j^n and μ_j^n; the former denotes an *element* of $\mathcal{U}_j^n$, while the latter denotes a *mapping* from $[2^{nC_{jd}}]$ to $\mathcal{U}_j^n$.

Encoding. For every $j \in [k]$, encoder j sends the pair (w_{j0}, w_{jd}) to the CF. The CF sends $((w_{i0})_{i \neq j}, Z_j)$ back to encoder j. Encoder j, having access to $w_0 = (w_{i0})_{i\in[k]}$ and Z_j, transmits $X_j^n(w_{jj}|U_0^n(w_0), U_j^n(w_{jd}, Z_j))$ over the channel.

Decoding. The decoder, upon receiving Y^n, maps Y^n to the unique k-tuple $\hat{W}_{[k]}$ such that

$$\Big(U_0^n(\hat{W}_0), \big(U_j^n(\hat{W}_{jd}, \hat{Z}_j|U_0^n)\big)_j, \big(X_j^n(\hat{W}_{jj}|U_0^n, U_j^n)\big)_j, Y^n\Big) \\ \in A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y). \tag{3.16}$$

If such a k-tuple does not exist, the decoder sets its output to the k-tuple $(1, 1, \ldots, 1)$. Note that in (3.16), $\hat{Z}_j$ is the CF output to encoder j corresponding to the CF input vector $(\hat{W}_{j0}, \hat{W}_{jd})_{j\in[k]}$.

The analysis of the expected error probability for the proposed random code appears in Subsection 3.5.1.

3.4 Case Study: 2-User Gaussian MAC

In this section, we study the network consisting of the 2-user Gaussian MAC with power constraints and a CF whose input link capacities are sufficiently large so that the CF has full access to the messages and whose output link capacities both equal C_{out}. We show that in this scenario, the benefit of cooperation extends beyond sum-capacity; that is, capacity metrics other than sum-capacity also exhibit an infinite slope at $C_{\text{out}} = 0$. In addition, we show that the behavior of these metrics (including sum-capacity) is bounded from below by a constant multiplied by $\sqrt{C_{\text{out}}}$.

The following corollary for the 2-user MAC follows from Theorem 3.2.1.

Corollary 3.4.1. *Consider a MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$ with a CF that has access to both messages and has output link capacities C_{out}^1 and C_{out}^2. The capacity region of this network contains the set of all rate pairs (R_1, R_2) that satisfy*

$$\begin{aligned} R_1 &\leq \max\{I(X_1; Y|U_0) - C_{1d}, I(X_1; Y|X_2, U_0)\} + C_{10} \\ R_2 &\leq \max\{I(X_2; Y|U_0) - C_{2d}, I(X_2; Y|X_1, U_0)\} + C_{20} \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U_0) + C_{10} + C_{20} \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned}$$

for some nonnegative constants $C_{1d} \le C^1_{\text{out}}$, $C_{2d} \le C^2_{\text{out}}$,

$$C_{10} := C^1_{\text{out}} - C_{2d}$$
$$C_{20} := C^2_{\text{out}} - C_{1d},$$

and some distribution $p(u_0)p(x_1, x_2|u_0)$ *that satisfies* $\mathbb{E}[X_i^2] \le P_i$ *for* $i \in \{1, 2\}$ *and*

$$C_{1d} + C_{2d} = I(X_1; X_2|U_0).$$

By (3.1), the 2-user Gaussian MAC can be represented as

$$Y = X_1 + X_2 + Z,$$

where Z is independent of (X_1, X_2) and is distributed as $Z \sim \mathcal{N}(0, N)$ for some noise variance $N > 0$. Let $U_0 \sim \mathcal{N}(0, 1)$, and (X_1', X_2') be a pair of random variables independent of U_0 and jointly distributed as $\mathcal{N}(\mu, \Sigma)$, where

$$\mu := \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ and } \Sigma := \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}$$

for some $\rho_0 \in [0, 1]$. Finally, for $i \in \{1, 2\}$, set

$$\frac{1}{\sqrt{P_i}} X_i = \rho_i X_i' + \sqrt{1 - \rho_i^2} U_0,$$

for some $\rho_i \in [0, 1]$. Calculating the region described in Corollary 3.4.1 for the Gaussian MAC using the joint distribution of (U_0, X_1, X_2) and setting $\gamma_i := P_i/N$ for $i \in \{1, 2\}$ and $\bar{\gamma} := \sqrt{\gamma_1 \gamma_2}$, gives the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned}
&R_1 - C_{10} \\
&\qquad \le \max\left\{ \frac{1}{2} \log \frac{1 + \rho_1^2\gamma_1 + \rho_2^2\gamma_2 + 2\rho_0\rho_1\rho_2\bar{\gamma}}{1 + (1 - \rho_0^2)\rho_2^2\gamma_2} - C_{1d}, \frac{1}{2} \log\left(1 + (1 - \rho_0^2)\rho_1^2\gamma_1\right) \right\} \\
&R_2 - C_{20} \\
&\qquad \le \max\left\{ \frac{1}{2} \log \frac{1 + \rho_1^2\gamma_1 + \rho_2^2\gamma_2 + 2\rho_0\rho_1\rho_2\bar{\gamma}}{1 + (1 - \rho_0^2)\rho_1^2\gamma_1} - C_{2d}, \frac{1}{2} \log\left(1 + (1 - \rho_0^2)\rho_2^2\gamma_2\right) \right\}
\end{aligned}$$

and

$$\begin{aligned}
R_1 + R_2 &\le \frac{1}{2} \log\left(1 + \rho_1^2\gamma_1 + \rho_2^2\gamma_2 + 2\rho_0\rho_1\rho_2\bar{\gamma}\right) + C_{10} + C_{20} \\
R_1 + R_2 &\le \frac{1}{2} \log\left(1 + \gamma_1 + \gamma_2 + 2\big(\rho_0\rho_1\rho_2 + \sqrt{(1 - \rho_1^2)(1 - \rho_2^2)}\big)\bar{\gamma}\right)
\end{aligned}$$

for some $\rho_1, \rho_2 \in [0,1]$ and $\rho_0 = \sqrt{1 - 2^{-2(C_{1d}+C_{2d})}}$. Denote this region with $\mathscr{C}_{\text{ach}}(C_{\text{out}})$.

We next introduce a lower bound for the weighted version of the sum-capacity. Denote the capacity region of this network with $\mathscr{C}(C_{\text{out}})$. For every $\alpha \in [0,1]$, define

$$C_\alpha(C_{\text{out}}) = \max_{(R_1,R_2)\in\mathscr{C}(C_{\text{out}})} (\alpha R_1 + (1-\alpha)R_2).$$

Note that $C_\alpha(C_{\text{out}})$ is a generalization of the notion of sum-capacity where the weighted sum of the encoders' rates is considered. The main result of this section demonstrates that for small C_{out}, $C_\alpha(C_{\text{out}})$ is bounded from below by a constant times $\sqrt{C_{\text{out}}}$ when C_{out} is small. The proof is given in Subsection 3.5.5.

Proposition 3.4.2. *For the Gaussian MAC*

$$Y = X_1 + X_2 + Z$$

with $Z \sim \mathcal{N}(0,N)$ and SNRs (γ_1, γ_2), we have

$$C_\alpha(C_{\text{out}}) - C_\alpha(0) \geq \frac{2\sqrt{\gamma_1\gamma_2 \cdot \log e}}{1+\gamma_1+\gamma_2} \cdot \min\{\alpha, 1-\alpha\} \cdot \sqrt{C_{\text{out}}} + o(\sqrt{C_{\text{out}}}).$$

In particular, for every $\alpha \in (0,1)$,

$$\left.\frac{dC_\alpha}{dC_{\text{out}}}\right|_{C_{\text{out}}=0^+} = \infty.$$

In Figure 3.2, using [42], we plot the sum-rate of the region $\mathscr{C}_{\text{ach}}(C_{\text{out}})$ and the forwarding inner bound (Corollary 3.2.2) for $\gamma_1 = \gamma_2 = 100$. We also plot the $\sqrt{C_{\text{out}}}$-term in the lower bound given by Proposition 3.4.2. Notice that the forwarding inner bound provides a cooperation gain that is at most linear in C_{out}.

3.5 Proofs

3.5.1 Proof of Theorem 3.2.1 (Inner bound)

Fix $\eta > 0$, and choose a distribution $p(u_0, u_{[k]}, x_{[k]})$ on $\mathcal{U}_0 \times \mathcal{U}_{[k]} \times \mathcal{X}_{[k]}$ of the form

$$p(u_0) \cdot \prod_{i\in S_d^c} p(u_i|u_0) \cdot p(u_{S_d}|u_0, u_{S_d^c}) \cdot \prod_{j\in[k]} p(x_j|u_0, u_j),$$

that satisfies the dependence constraints

$$\zeta_S := \sum_{j\in S} C_{jd} - \sum_{j\in S} H(U_j|U_0) + H(U_S|U_0, U_{S_d^c}) > 0 \qquad \forall\, \varnothing \subsetneq S \subseteq S_d,$$

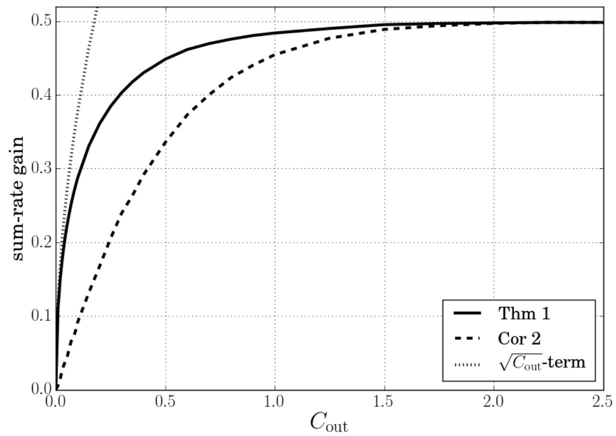

Figure 3.2: Plot of the achievable sum-rate gain given by Theorem 3.2.1 and Corollary 3.2.2 (forwarding inner bound) for Gaussian input distributions, and the $\sqrt{C_{\text{out}}}$-term given in Proposition 3.4.2. Here $\gamma_1 = \gamma_2 = 100$.

and cost constraints

$$\mathbb{E}\big[b_j(X_j)\big] \le B_j - \eta \qquad \forall j \in [k]. \tag{3.17}$$

Let $(w_1, \dots, w_k)$ denote the transmitted message vector and $(\hat{W}_1, \dots, \hat{W}_k)$ denote the output of the decoder. To simplify notation, denote

$$U_0^n(w_0), U_j^n(w_{jd}, Z_j|U_0^n), X_j^n(w_{jj}|U_0^n, U_j^n)$$

with U_0^n, U_j^n, and X_j^n, respectively. Similarly, define $\hat{U}_0^n$, $\hat{U}_j^n$, and $\hat{X}_j^n$ as

$$U_0^n(\hat{W}_0), U_j^n(\hat{W}_{jd}, Z_j|U_0^n), X_j^n(\hat{W}_{jj}|U_0^n, U_j^n).$$

Here $\hat{W}_0$, $\hat{W}_{jd}$, and $\hat{W}_{jj}$ are defined in terms of $(\hat{W}_j)_j$ similar to the definitions of w_0, w_{jd}, and w_{jj} in Section 3.3. Let Y^n denote the channel output when $X_{[k]}^n$ is transmitted. Then the joint distribution of $(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n)$ is given by

$$p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) := p(u_0^n) p_{\text{code}}(u_{[k]}^n|u_0^n) p(x_{[k]}^n|u_0^n, u_{[k]}^n) p(y^n|x_{[k]}^n), \tag{3.18}$$

where

$$p_{\text{code}}(u_{[k]}^n|u_0^n) = \sum_{\mu_{[k]}^n} p(\mu_1|u_0^n) \cdots p(\mu_k|u_0^n) p(u_{[k]}^n|u_0^n, \mu_{[k]}^n) \tag{3.19}$$

and $p(\mu_j^n|u_0^n)$ and $p(u_{[k]}^n|u_0^n, \mu_{[k]}^n)$ are calculated according to

$$p(\mu_j^n|u_0^n) = \prod_{z_j \in [2^{nC_{jd}}]} p(\mu_j^n(z_j)|u_0^n) = \prod_{z_j \in [2^{nC_{jd}}]} \prod_{t=1}^{n} p(\mu_{jt}(z_j)|u_{0t}),$$

and

$$p(u_{[k]}^n|u_0^n, \mu_{[k]}^n) = \sum_{z_{[k]}} p(z_{[k]}|u_0^n, \mu_{[k]}^n) \prod_{j\in[k]} \mathbf{1}\{\mu_j^n(z_j) = u_j^n\}.$$

Furthermore, $p(z_{[k]}|u_0^n, \mu_{[k]}^n)$ is given by

$$p(z_{[k]}|u_0^n, \mu_{[k]}^n) = \begin{cases} \mathbf{1}\big\{z_{[k]} = \min \mathcal{A}(u_0^n, \mu_{[k]}^n)\big\} & \text{if } \mathcal{A}(u_0^n, \mu_{[k]}^n) \neq \varnothing \\ \mathbf{1}\big\{z_{[k]} = 1_{[k]}\big\} & \text{otherwise,} \end{cases}$$

where the minimum is calculated according to the lexicographical ordering. Define the distribution $p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)$ as

$$p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) := p(u_0^n) p(x_{[k]}^n|u_0^n, u_{[k]}^n) p(y^n|x_{[k]}^n) \prod_{j\in[k]} p(u_j^n|u_0^n), \tag{3.20}$$

which is the joint input-output distribution if independent codewords are transmitted. We next review some results regarding weakly typical sets [40, pp. 520-524] that are required for our error analysis.

For any $S \subseteq [k]$, let $A_\delta^{(n)}(U_0, U_S)$ denote the weakly typical set with respect to the distribution $p(u_0, u_S)$, a marginal of $p(u_0, u_{[k]})$. In addition, for every $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$, let $A_\delta^{(n)}(u_0^n, u_S^n)$ be the set of all $u_{S^c}^n$ such that

$$(u_0^n, u_{[k]}^n) \in A_\delta^{(n)}(U_0, U_{[k]}).$$

Similarly, let $A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y)$ be the weakly typical set with respect to the distribution $p(u_0, u_{[k]}, x_{[k]}) p(y|x_{[k]})$, where $p(y|x_{[k]})$ is given by the channel definition. For subsets $S, T \subseteq [k]$, $A_\epsilon^{(n)}(U_0, U_S, X_T, Y)$ is the weakly typical set with respect to the marginal distribution $p(u_0, u_S, x_T, y)$. For $(u_0^n, u_S^n, x_T^n, y^n) \in A_\epsilon^{(n)}(U_0, U_S, X_T, Y)$, $A_\epsilon^{(n)}(u_0^n, u_S^n, x_T^n, y^n)$ is the set of all $(u_{S^c}^n, x_{T^c}^n)$ such that

$$\big(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n\big) \in A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y).$$

Furthermore, we have [40, p. 523]

$$\log |A_\epsilon^{(n)}(u_0^n, u_S^n, x_T^n, y^n)| \leq n\big(H(U_{S^c}, X_{T^c}|U_0, U_S, X_T, Y) + 2\epsilon\big). \tag{3.21}$$

Finally, under fairly general conditions described in Appendix B,[5] there exists an increasing function $I : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that if $(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n) \overset{\text{iid}}{\sim} p(u_0, u_{[k]}, x_{[k]}, y)$, then

$$\Pr\Big\{(U_0^n, U_{[k]}^n, X_{[k]}^n, Y^n) \in A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y)\Big\} \geq 1 - 2^{-nI(\epsilon)}. \tag{3.22}$$

[5]Distributions that satisfy these conditions include any distribution on a finite alphabet and the Gaussian distribution.

Fix any such function I.

We next study the relationship between p_{code} and p_{ind} given by (3.18) and (3.20), respectively. Note that the main difference arises from the conditional marginal $p_{\text{code}}(u_{[k]}^n|u_0^n)$ and $p_{\text{ind}}(u_{[k]}^n|u_0^n)$. Our first lemma provides an upper bound for p_{code} in terms of p_{ind}.

Lemma 3.5.1. *For every nonempty $S \subseteq [k]$ and all (u_0^n, u_S^n),*

$$\frac{1}{n} \log \frac{p_{\text{code}}(u_S^n|u_0^n)}{p_{\text{ind}}(u_S^n|u_0^n)} \leq nC_{Sd},$$

where $C_{Sd} := \sum_{j\in S} C_{jd}$.

Proof. By (3.19),

$$p_{\text{code}}(u_S^n|u_0^n) = \sum_{\mu_{[k]}^n} p(u_S^n|u_0^n, \mu_{[k]}^n) \prod_{j\in[k]} p(\mu_j^n|u_0^n).$$

To bound $p_{\text{code}}(u_S^n|u_0^n)$, note that

$$p(u_S^n|u_0^n, \mu_{[k]}^n) \leq \prod_{j\in S} \mathbf{1}\big\{(\mu_j^n)^{-1}(u_j^n) \neq \varnothing\big\},$$

where

$$(\mu_j^n)^{-1}(u_j^n) := \left\{z_j \in [2^{nC_{jd}}] \middle| \mu_j^n(z_j) = u_j^n\right\}.$$

For every $j \in S$,

$$\begin{aligned}\sum_{\mu_j^n} p(\mu_j^n|u_0^n)\mathbf{1}\big\{(\mu_j^n)^{-1}(u_j^n) \neq \varnothing\big\} &= \Pr\left\{\exists z_j : U_j^n(z_j) = u_j^n \middle| U_0^n = u_0^n\right\} \\ &\leq 2^{nC_{jd}} p(u_j^n|u_0^n).\end{aligned}$$

Thus

$$\begin{aligned}p_{\text{code}}(u_S^n|u_0^n) &\leq \sum_{\mu_S} \prod_{j\in S} p(\mu_j^n|u_0^n)\mathbf{1}\big\{(\mu_j^n)^{-1}(u_j^n) \neq \varnothing\big\} \\ &= \prod_{j\in S}\Big(\sum_{\mu_j^n} p(\mu_j^n|u_0^n)\mathbf{1}\big\{(\mu_j^n)^{-1}(u_j^n) \neq \varnothing\big\}\Big) \\ &\leq 2^{n\sum_{j\in S} C_{jd}} p_{\text{ind}}(u_S^n|u_0^n).\end{aligned}$$

□

Our second lemma provides an upper bound for $p_{\text{ind}}(u_S^n|u_0^n)$ when (u_0^n, u_S^n) is typical.

Lemma 3.5.2. *For all nonempty $S_d^c \subseteq S \subseteq [k]$ and $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$,*

$$\frac{1}{n}\log\frac{p_{\text{ind}}(u_S^n|u_0^n)}{p(u_S^n|u_0^n)} \leq -\sum_{j\in S\cap S_d} H(U_j|U_0) + H(U_{S\cap S_d}|U_0, U_{S_d^c}) + 2(|S\cap S_d|+1)\delta.$$

Proof. Recall that

$$p(u_{[k]}^n|u_0^n) = p(u_{S_d}^n|u_0^n, u_{S_d^c}^n)\prod_{j\in S_d^c} p(u_j^n|u_0^n).$$

Thus for all $S \supseteq S_d^c$, we have

$$p(u_S^n|u_0^n) = p(u_{S\cap S_d}^n|u_0^n, u_{S_d^c}^n)\prod_{j\in S_d^c} p(u_j^n|u_0^n).$$

Therefore,

$$\begin{aligned}\frac{p_{\text{ind}}(u_S^n|u_0^n)}{p(u_S^n|u_0^n)} &= \frac{p_{\text{ind}}(u_{S\cap S_d}^n|u_0^n)}{p(u_{S\cap S_d}^n|u_0^n, u_{S_d^c}^n)} \\ &= \frac{\prod_{j\in S\cap S_d} p(u_j^n|u_0^n)}{p(u_{S\cap S_d}^n|u_0^n, u_{S_d^c}^n)}.\end{aligned}$$

The proof now follows from the definition of $A_\delta^{(n)}(U_0, U_S)$. □

Combining the previous two lemmas results in the next corollary, which we use in our error analysis.

Corollary 3.5.3. *For every nonempty S satisfying $S_d^c \subseteq S \subseteq [k]$ and all $(u_0^n, u_S^n) \in A_\delta^{(n)}(U_0, U_S)$,*

$$\frac{1}{n}\log\frac{p_{\text{code}}(u_S^n|u_0^n)}{p(u_S^n|u_0^n)} \leq \zeta_{S\cap S_d} + 2\big(|S\cap S_d|+1\big)\delta.$$

Let $\mathcal{E}$ denote the event where either the output of an encoder does not satisfy the corresponding cost constraint, or the output of the decoder differs from the transmitted k-tuple of messages; that is $(\hat{W}_j)_{j=1}^k \neq (w_j)_{j=1}^k$. Denote the former event with $\mathcal{E}_{\text{cost}}$ and the latter event with $\mathcal{E}_{\text{dec}}$. When $\mathcal{E}_{\text{dec}}$ occurs, it is either the case that $(w_j)_{j=1}^k$ does not satisfy (3.16) (denote this event with $\mathcal{E}_{\text{typ}}$), or that there is another k-tuple, $(\hat{W}_j)_{j=1}^k \neq (w_j)_{j=1}^k$, that also satisfies (3.16). If the latter event occurs, we either have $\hat{W}_0 \neq w_0$ (denote this event with $\mathcal{E}_{\varnothing,\varnothing}$), or $\hat{W}_0 = w_0$. When $\hat{W}_0 = w_0$, define the subsets $S, T \subseteq [k]$ as

$$\begin{aligned}S &:= \big\{j\big|\hat{W}_{jd} \neq w_{jd}\big\}\\ T &:= \big\{j\big|\hat{W}_{jj} \neq w_{jj}\big\}.\end{aligned}$$

Now for every pair of subsets $S, T \subseteq [k]$ such that $S \cup T \neq \varnothing$, define $\mathcal{E}_{S,T}$ as the event where there exists a $(\hat{W}_j)_{j=1}^k$ that satisfies (3.16), $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. Thus we may write

$$\mathcal{E} \subseteq \mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}} \cup \bigcup_{S,T \subseteq [k]} \mathcal{E}_{S,T}.$$

The union over all $\mathcal{E}_{S,T}$ also contains the event $\mathcal{E}_{\varnothing,\varnothing}$. By the union bound,

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}) + \sum_{S,T \subseteq [k]} \Pr(\mathcal{E}_{S,T}).$$

Thus to find a set of achievable rates for our random code design, it suffices to find conditions under which $\Pr(\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}})$ and each $\Pr(\mathcal{E}_{S,T})$ go to zero as $n \to \infty$.

We begin with the event $\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}$. Define $\mathcal{E}_{\text{enc}}$ as the event where

$$\left(U_0^n, U_{[k]}^n\right) \notin A_\delta^{(n)}(U_0, U_{[k]}).$$

We bound the probability of $\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}$ by

$$\Pr(\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}) \leq \Pr(\mathcal{E}_{\text{enc}} \cup \mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}) \leq \Pr(\mathcal{E}_{\text{enc}}) + \Pr(\mathcal{E}_{\text{cost}} \setminus \mathcal{E}_{\text{enc}}) + \Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}). \tag{3.23}$$

The event $\mathcal{E}_{\text{enc}}$ occurs if and only if $\mathcal{A}(U_0^n, U_{[k]}^n(.))$ (defined in Section 3.3) is empty. Thus

$$\Pr(\mathcal{E}_{\text{enc}}) = \Pr\left\{\mathcal{A}(U_0^n, U_{[k]}^n(.)) = \varnothing\right\}.$$

If $S_d = \varnothing$, then

$$p_{\text{code}}(u_{[k]}^n | u_0^n) = p_{\text{ind}}(u_{[k]}^n | u_0^n) = p(u_{[k]}^n | u_0^n),$$

which implies

$$\Pr\left(\mathcal{E}_{\text{enc}}^c\right) \geq \Pr\left(A_\delta^{(n)}(U_0, U_{[k]})\right).$$

Thus $\Pr(\mathcal{E}_{\text{enc}})$ goes to zero in this case. If $S_d \neq \varnothing$, recall that for every nonempty $S \subseteq S_d$, ζ_S is defined as

$$\zeta_S = \sum_{j \in S} C_{jd} - \sum_{j \in S} H(U_j | U_0) + H(U_S | U_0, U_{S_d^c}).$$

From the multivariate covering lemma (Appendix A), it follows that $\Pr(\mathcal{E}_{\text{enc}})$ goes to zero if for all nonempty $S \subseteq S_d$,

$$\zeta_S > (8|S_d| - 2|S| + 10)\delta. \tag{3.24}$$

Next, for $\mathcal{E}_{\text{cost}} \setminus \mathcal{E}_{\text{enc}}$ we have

$$\begin{aligned}
&\Pr\left(\mathcal{E}_{\text{cost}} \setminus \mathcal{E}_{\text{enc}}\right) && (3.25)\\
&= \sum_{\substack{(u_0^n, u_{[k]}^n, x_{[k]}^n):\\ (u_0^n, u_{[k]}^n) \in A_\delta^{(n)}}} p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n) \Pr\left\{\exists j \in [k] : \frac{1}{n}\sum_{t=1}^n b_j(x_{jt}) > B_j\right\}\\
&\le 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \sum_{\substack{(u_0^n, u_{[k]}^n, x_{[k]}^n):\\ (u_0^n, u_{[k]}^n) \in A_\delta^{(n)}}} p(u_0^n, u_{[k]}^n, x_{[k]}^n) \Pr\left\{\exists j \in [k] : \frac{1}{n}\sum_{t=1}^n b_j(x_{jt}) > B_j\right\}\\
&\le 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \sum_{x_{[k]}^n} p(x_{[k]}^n) \Pr\left\{\exists j \in [k] : \frac{1}{n}\sum_{t=1}^n b_j(x_{jt}) > B_j\right\}\\
&\le 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \sum_{j\in[k]} \sum_{x_j^n} p(x_j^n) \Pr\left\{\frac{1}{n}\sum_{t=1}^n b_j(x_{jt}) > B_j\right\}\\
&\le 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \sum_{j\in[k]} 2^{-n\theta(b_j(X_j),\eta)}, && (3.26)
\end{aligned}$$

where (3.26) follows from (B.3) in Appendix B. Thus $\Pr\left(\mathcal{E}_{\text{cost}} \setminus \mathcal{E}_{\text{enc}}\right) \to 0$ if

$$\zeta_{S_d} + 2(|S_d|+1)\delta) < \min_{j\in[k]} \theta(b_j(X_j), \eta). \tag{3.27}$$

We now find an upper bound for $\Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}})$. Let $B^{(n)}$ be the set of all $(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)$ such that $(u_0^n, u_{[k]}^n) \in A_\delta^{(n)}(U_0, U_{[k]})$ but $(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) \notin A_\epsilon^{(n)}(U_0, U_{[k]}, X_{[k]}, Y)$. Then

$$\begin{aligned}
\Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}) &= \sum_{B^{(n)}} p(u_0^n) p_{\text{code}}(u_{[k]}^n|u_0^n) p(x_{[k]}^n|u_0^n, u_{[k]}^n) p(y^n|x_{[k]}^n)\\
&\overset{(a)}{\le} 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \sum_{B^{(n)}} p(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n)\\
&\overset{(b)}{\le} 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} \Pr\left\{(A_\epsilon^{(n)})^c\right\} \overset{(c)}{\le} 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta - I(\epsilon))},
\end{aligned}$$

where (a) follows from Corollary 3.5.3, (b) holds since $B^{(n)} \subseteq (A_\epsilon^{(n)})^c$, and (c) follows from the definition of $I(\epsilon)$ given by (3.22). Thus $\Pr(\mathcal{E}_{\text{typ}} \setminus \mathcal{E}_{\text{enc}}) \to 0$ if

$$\zeta_{S_d} < I(\epsilon) - 2(|S_d|+1)\delta. \tag{3.28}$$

Therefore, if (3.24), (3.27), and (3.28) hold, then by (3.23), $\Pr(\mathcal{E}_{\text{cost}} \cup \mathcal{E}_{\text{typ}}) \to 0$.

We next study $\mathcal{E}_{\varnothing,\varnothing}$, which is the event where there exists a k-tuple $(\hat{W}_j)_j$ that satisfies (3.16) but $\hat{W}_0 \neq w_0$. If this event occurs, then $(\hat{U}_0^n, \hat{U}_{[k]}^n, \hat{X}_{[k]}^n)$ and Y^n are independent. By the union bound,

$$\Pr(\mathcal{E}_{\varnothing,\varnothing}) \leq 2^{n\sum_{j\in[k]} R_j} \sum_{A_\epsilon^{(n)}} p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n) p_{\text{code}}(y^n).$$

We rewrite the sum in the above inequality as

$$\sum_{A_\epsilon^{(n)}(Y)} p_{\text{code}}(y^n) \sum_{A_\epsilon^{(n)}(y^n)} p_{\text{code}}(u_0^n, u_{[k]}^n, x_{[k]}^n),$$

Using Corollary 3.5.3, we upper bound the inner sum by

$$\begin{aligned} &\sum_{A_\epsilon^{(n)}(y^n)} 2^{n(\zeta_{S_d}+2(|S_d|+1)\delta)} p(u_0^n, u_{[k]}^n, x_{[k]}^n) \\ &\quad \leq 2^{n(H(U_0,U_{[k]},X_{[k]}|Y)+2\epsilon)} 2^{n(\zeta_{|S_d|}+2(|S_d|+1)\delta)} 2^{-n(H(U_0,U_{[k]},X_{[k]})+\epsilon)}, \end{aligned} \tag{3.29}$$

where (3.29) follows from (3.21). This implies $\Pr(\mathcal{E}_{\varnothing,\varnothing}) \to 0$ if

$$\sum_{j\in[k]} R_j < I(X_{[k]};Y) - \zeta_{S_d} - 2(|S_d|+1)\delta - 3\epsilon.$$

Next, let $S, T \subseteq [k]$ be sets such that $S \cup T \neq \varnothing$ and consider the event $\mathcal{E}_{S,T}$. Recall that this is the event where there exists a k-tuple $(\hat{W}_j)_j$ that satisfies (3.16) and $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. For every $A \subseteq S$ and $B \subseteq S^c$, let $\mathcal{E}_{S,T}^{A,B} \subseteq \mathcal{E}_{S,T}$ be the event where there exists a k-tuple $(\hat{W}_j)_j$ that satisfies

$$\begin{aligned} &\Big(U_0^n(w_0), \big(U_j^n(\hat{W}_{jd}, \hat{Z}_j|U_0^n)\big)_{j\in A}, \big(U_j^n(w_{jd}, \hat{Z}_j|U_0^n)\big)_{j\in B}, \\ &\quad \big(X_j^n(\hat{W}_{jj}|U_0^n, \hat{U}_j^n)\big)_{j\in A\cup(B\cap T)}, \big(X_j^n(w_{jj}|U_0^n, \hat{U}_j^n)\big)_{j\in B\setminus T}, Y^n\Big) \in A_\epsilon^{(n)} \end{aligned} \tag{3.30}$$

and $\hat{W}_0 = w_0$, $\hat{W}_{jd} \neq w_{jd}$ if and only if $j \in S$, and $\hat{W}_{jj} \neq w_{jj}$ if and only if $j \in T$. If $\mathcal{E}_{S,T}$ occurs, then so does $\mathcal{E}_{S,T}^{A,B}$ for every $A \subseteq S$ and $B \subseteq S^c$. Thus

$$\mathcal{E}_{S,T} \subseteq \bigcap_{A,B} \mathcal{E}_{S,T}^{A,B}.$$

This implies

$$\Pr(\mathcal{E}_{S,T}) \leq \min_{A,B} \Pr\big(\mathcal{E}_{S,T}^{A,B}\big). \tag{3.31}$$

Therefore, to bound $\Pr(\mathcal{E}_{S,T})$, we find an upper bound on $\Pr(\mathcal{E}_{S,T}^{A,B})$ for any $A \subseteq S$ and $B \subseteq S^c$ such that $A \cup (B \cap T) \neq \varnothing$. This is the key difference between our error analysis here and the error analysis for the 2-user MAC with transmitter cooperation presented in [43]. For independent distributions, using the constraint that subsets of typical codewords are also typical does not lead to a larger region; the same may not be true when dealing with dependent distributions. That being said, to include all independent random variables in our error analysis, instead of calculating the minimum in (3.31) over all $A \subseteq S$ and $B \subseteq S^c$, we limit ourselves to subsets A and B that satisfy

$$S \cap S_d^c \subseteq A \subseteq S$$
$$S^c \cap S_d^c \subseteq B \subseteq S^c,$$

since all the random vectors $(U_j^n)_{j\in S_d^c}$ are independent given U_0^n. Choose any such A and B. Note that for every $j \in A \cup (B \cap T)$, either $\hat{W}_{jd} \neq w_{jd}$ or $\hat{W}_{jj} \neq w_{jj}$. In addition, in (3.30),

$$\Big(\big(U_j^n(\hat{W}_{jd}, \hat{Z}_j|U_0^n)\big)_{j\in A}, \big(U_j^n(w_{jd}, \hat{Z}_j|U_0^n)\big)_{j\in B}, \big(X_j^n(\hat{W}_{jj}|U_0^n, U_j^n)\big)_{j\in A\cup(B\cap T)}, \big(X_j^n(w_{jj}|U_0^n, U_j^n)\big)_{j\in B\setminus T}\Big)$$

is independent of Y^n given

$$\Big(U_0^n(w_0), \big(U_j^n(w_{jd}, .|U_0^n)\big)_{j\in S^c}, \big(X_j^n(w_{jj}|U_0^n, U_j^n(.))\big)_{j\in S^c\setminus T}\Big).$$

Therefore, by the union bound, $\Pr(\mathcal{E}_{S,T}^{A,B})$ is bounded from above by

$$2^{n\left(\sum_{j\in A} R_{jd}+\sum_{j\in A\cup(B\cap T)} R_{jj}\right)} \times \sum_{A_\epsilon^{(n)}} p(x_{A\cup(B\cap T)}^n|u_0^n, u_{A\cup(B\cap T)}^n) \times \sum_{\mu_{A\cup S^c}^n, \chi_{S^c\setminus T}^n} p(u_0^n, \mu_{S^c}^n, \chi_{S^c\setminus T}^n, y^n)p(\mu_A|u_0^n)p(u_{A\cup B}^n, x_{B\setminus T}^n|u_0^n, \mu_{A\cup S^c}^n, \chi_{S^c\setminus T}^n), \tag{3.32}$$

where the inner sum is over all mappings $\mu_j^n : [2^{nC_{jd}}] \to \mathcal{U}_j^n$ for $j \in A \cup S^c$ and $\chi_j^n : [2^{nC_{jd}}] \to \mathcal{X}_j^n$ for $j \in S^c \setminus T$. The distribution $p(u_0^n, \mu_{S^c}^n, \chi_{S^c\setminus T}^n, y^n)$ is a marginal of $p(u_0^n, \mu_{[k]}^n, \chi_{[k]}^n, y^n)$, which is defined as

$$p(u_0^n, \mu_{[k]}^n, \chi_{[k]}^n, y^n) = p(u_0^n, \mu_{[k]}^n)p(\chi_{[k]}^n|u_0^n, \mu_{[k]}^n)p(y^n|u_0^n, \mu_{[k]}^n, \chi_{[k]}^n),$$

where

$$\begin{aligned} p(\chi_{[k]}^n|u_0^n, \mu_{[k]}^n) &= \prod_{j\in[k]} p(\chi_j^n|u_0^n, \mu_j^n) \\ &= \prod_{j\in[k]} \prod_{z_j\in[2^{nC_{jd}}]} p(\chi_j^n(z_j)|u_0^n, \mu_j^n(z_j)), \end{aligned}$$

and

$$p(y^n|u_0^n, \mu_{[k]}^n, \chi_{[k]}^n) = \sum_{z_{[k]}} p(z_{[k]}|u_0^n, \mu_{[k]}^n) p(y^n|\chi_{[k]}^n(z_{[k]})).$$

We have

$$\begin{aligned}
&p(u_{A\cup B}^n, x_{B\setminus T}^n | u_0^n, \mu_{A\cup S^c}^n, \chi_{S^c\setminus T}^n) \\
&\leq \mathbf{1}\Big\{\exists (z_j)_{j\in B} \in \prod_{j\in B} [2^{nC_{jd}}] : (\forall j \in B : \mu_j^n(z_j) = u_j^n) \wedge (\forall j \in B \setminus T : \chi_j^n(z_j) = x_j^n)\Big\} \\
&\quad \times \mathbf{1}\Big\{\exists (z_j)_{j\in A} \in \prod_{j\in A} [2^{nC_{jd}}] : (\forall j \in A : \mu_j^n(z_j) = u_j^n)\Big\}. \qquad (3.33)
\end{aligned}$$

We can thus upper bound the inner sum in (3.32) as a product of the sums

$$\begin{aligned}
&\sum_{\mu_{S^c}^n, \chi_{S^c\setminus T}^n} p(u_0^n, \mu_{S^c}^n, \chi_{S^c\setminus T}^n, y^n) \\
&\times \mathbf{1}\Big\{\exists (z_j)_{j\in B} \in \prod_{j\in B} [2^{nC_{jd}}] : (\forall j \in B : \mu_j^n(z_j) = u_j^n) \wedge (\forall j \in B \setminus T : \chi_j^n(z_j) = x_j^n)\Big\}
\end{aligned}$$

and

$$\sum_{\mu_A} p(\mu_A|u_0^n)\mathbf{1}\Big\{\exists (z_j)_{j\in A} \in \prod_{j\in A} [2^{nC_{jd}}] : \forall j \in A, \mu_j^n(z_j) = u_j^n\Big\}.$$

We first find an upper bound for the first sum. Define

$$\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) := \sum_{\mu_{[k]}^n, \chi_{[k]}^n} p(u_0^n, \mu_{[k]}^n, \chi_{[k]}^n, y^n) \prod_{j\in[k]} \mathbf{1}\{\mu_j^n(1) = u_j^n, \chi_j^n(1) = x_j^n\}.$$

The following argument demonstrates that $\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n) = p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n)$,

$$\begin{aligned}
\tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n) &= \sum_{y^n} \tilde{p}(u_0^n, u_{[k]}^n, x_{[k]}^n, y^n) \\
&= \sum_{\mu_{[k]}^n, \chi_{[k]}^n} p(u_0^n, \mu_{[k]}^n, \chi_{[k]}^n) \prod_{j\in[k]} \mathbf{1}\{\mu_j^n(1) = u_j^n, \chi_j^n(1) = x_j^n\} \\
&= p(u_0^n) \prod_{j\in[k]} \sum_{\mu_j^n, \chi_j^n} p(\mu_j^n, \chi_j^n | u_0^n) \mathbf{1}\{\mu_j^n(1) = u_j^n, \chi_j^n(1) = x_j^n\} \\
&= p_{\text{ind}}(u_0^n, u_{[k]}^n, x_{[k]}^n). \qquad (3.34)
\end{aligned}$$

For every $z_B = (z_j)_{j\in B}$, where $z_j \in [2^{nC_{jd}}]$ for all $j \in B$, let $\mathcal{E}(z_B)$ denote the event where for all $j \in B$, $U_j^n(w_{jd}, z_j|U_0^n) = u_j^n$, and for all $j \in B \setminus T$,

$X_j^n(w_{jj}|U_0^n, U_j^n) = x_j^n$. Also, recall that by definition, $C_{Bd} = \sum_{j\in B} C_{jd}$. Then

$$\begin{aligned}
&\sum_{\mu^n_{S^c},\chi^n_{S^c\setminus T}} p(u_0^n, \mu^n_{S^c}, \chi^n_{S^c\setminus T}, y^n) \\
&\quad \times \mathbf{1}\Big\{\exists z_B \in \prod_{j\in B}[2^{nC_{jd}}] : (\forall j \in B : \mu_j^n(z_j) = u_j^n) \wedge (\forall j \in B\setminus T : \chi_j^n(z_j) = x_j^n)\Big\} \\
&\quad = \Pr\Big(\{U_0^n = u_0^n, Y^n = y^n\} \cap \bigcup_{z_B} \mathcal{E}(z_B)\Big) \\
&\quad = \Pr\Big(\bigcup_{z_B}(\{U_0^n = u_0^n, Y^n = y^n\} \cap \mathcal{E}(z_B))\Big) \\
&\quad \le 2^{nC_{Bd}} \Pr\big(\{U_0^n = u_0^n, Y^n = y^n\} \cap \mathcal{E}(1_B)\big) && (3.35) \\
&\quad = 2^{nC_{Bd}} \tilde{p}(u_0^n, u_B^n, x^n_{B\setminus T}, y^n) \\
&\quad = 2^{nC_{Bd}} p(u_0^n) p_{\text{ind}}(u_B^n, x^n_{B\setminus T}|u_0^n) \tilde{p}(y^n|u_0^n, u_B^n, x^n_{B\setminus T}), && (3.36)
\end{aligned}$$

where (3.35) follows by the union bound and (3.36) follows from (3.34). Using a similar argument we can show

$$\sum_{\mu_A} p(\mu_A|u_0^n)\mathbf{1}\Big\{\exists z_A \in \prod_{j\in A}[2^{nC_{jd}}] : \forall j \in A, \mu_j^n(z_j) = u_j^n\Big\} \le 2^{nC_{Ad}} p_{\text{ind}}(u_A^n|u_0^n). \tag{3.37}$$

Thus by (3.33), (3.36), and (3.37), the expression

$$\begin{aligned}
&2^{n\left(\sum_{j\in A} R_{jd} + \sum_{j\in A\cup(B\cap T)} R_{jj} + C_{Ad} + C_{Bd}\right)} \\
&\qquad \times \sum_{A_\epsilon^{(n)}} p(u_0^n) p_{\text{ind}}(u^n_{A\cup B}|u_0^n) p(x^n_{A\cup B}|u_0^n, u^n_{A\cup B}) \tilde{p}(y^n|u_0^n, u_B^n, x^n_{B\setminus T})
\end{aligned}$$

is an upper bound for (3.32). Applying Lemma 3.5.2 to $p_{\text{ind}}(u^n_{A\cup B}|u_0^n)$ and dropping the epsilon term, this expression can be further bounded from above by

$$\begin{aligned}
&2^{n\left(\sum_{j\in A} R_{jd} + \sum_{j\in A\cup(B\cap T)} R_{jj} + \zeta_{(A\cup B)\cap S_d}\right)} \\
&\qquad \times \sum_{A_\epsilon^{(n)}(U_0, U_B, X_{B\setminus T}, Y)} p(u_0^n, u_B^n, x^n_{B\setminus T}) \tilde{p}(y^n|u_0^n, u_B^n, x^n_{B\setminus T}) \\
&\qquad \times \sum_{A_\epsilon^{(n)}(u_0^n, u_B^n, x^n_{B\setminus T}, y^n)} p(u_A^n|u_0^n, u_B^n) p(x^n_{A\cup(B\cap T)}|u_0^n, u^n_{A\cup(B\cap T)})
\end{aligned}$$

Using (3.21), we can further upper bound the logarithm of this expression by

$$\begin{aligned} n\Bigg[\sum_{j\in A} R_{jd} + \sum_{j\in A\cup(B\cap T)} R_{jj} + \zeta_{(A\cup B)\cap S_d}\Bigg] \\ + \log \sum_{A_\epsilon^{(n)}(U_0,U_B,X_{B\setminus T},Y)} p(u_0^n, u_B^n, x_{B\setminus T}^n)\tilde{p}(y^n|u_0^n, u_B^n, x_{B\setminus T}^n) \\ - nH(U_A|U_0,U_B) - nH(X_{A\cup(B\cap T)}|U_0,U_{A\cup(B\cap T)}) \\ + nH(U_A, X_{A\cup(B\cap T)}|U_0,U_B,X_{B\setminus T},Y). \end{aligned}$$

Therefore, $\Pr(\mathcal{E}_{S,T}^{A,B}) \to 0$ if

$$\begin{aligned} \sum_{j\in A} R_{jd} + \sum_{j\in A\cup(B\cap T)} R_{jj} \\ < -\zeta_{(A\cup B)\cap S_d} + H(U_A|U_0,U_B) + H(X_{A\cup(B\cap T)}|U_0,U_{A\cup(B\cap T)}) \\ - H(U_A, X_{A\cup(B\cap T)}|U_0,U_B,X_{B\setminus T},Y) \\ = I(U_A, X_{A\cup(B\cap T)};Y|U_0,U_B,X_{B\setminus T}) - \zeta_{(A\cup B)\cap S_d}, \end{aligned}$$

where the last equality follows from the fact that

$$\begin{aligned} H(U_A|U_0,U_B) &= H(U_A|U_0,U_B,X_{B\setminus T}) + I(U_A;X_{B\setminus T}|U_0,U_B) \\ &= H(U_A|U_0,U_B,X_{B\setminus T}) \end{aligned}$$

and

$$\begin{aligned} H(X_{A\cup(B\cap T)}|U_0,U_{A\cup B}) &= H(X_{A\cup(B\cap T)}|U_0,U_{A\cup B},X_{B\setminus T}) \\ &\quad + I(X_{A\cup(B\cap T)};X_{B\setminus T}|U_0,U_{A\cup B}) \\ &= H(X_{A\cup(B\cap T)}|U_0,U_{A\cup B},X_{B\setminus T}). \end{aligned}$$

Thus $\Pr(\mathcal{E}_{S,T}) \to 0$ if for some $S\cap S_d^c \subseteq A \subseteq S$ and $S^c\cap S_d^c \subseteq B \subseteq S^c$ such that $A\cup(B\cap T) \neq \varnothing$,

$$\begin{aligned} \sum_{j\in A} R_{jd} + \sum_{j\in A\cup(B\cap T)} R_{jj} \\ < I(U_A, X_{A\cup(B\cap T)};Y|U_0,U_B,X_{B\setminus T}) - \zeta_{(A\cup B)\cap S_d}. \end{aligned} \tag{3.38}$$

The bounds we obtain above are in terms of $(R_{jd})_{j=1}^k$ and $(R_{jj})_{j=1}^k$. To convert these to bounds in terms of $(R_j)_{j=1}^k$, recall that $R_{j0} = \min\{C_{j0}, R_j\}$, $R_{jj} = (R_j - C_{\text{in}}^j)^+$, and

$$\begin{aligned} R_{jd} &= R_j - R_{j0} - R_{jj} \\ &= R_j - \min\{C_{j0}, R_j\} - R_{jj} = \max\{R_j - C_{j0}, 0\} - (R_j - C_{\text{in}}^j)^+ \\ &= (R_j - C_{j0})^+ - (R_j - C_{\text{in}}^j)^+. \end{aligned}$$

Thus (3.38) can be written as

$$\sum_{j\in A}(R_j - C_{j0})^+ + \sum_{j\in B\cap T}(R_j - C_{\text{in}}^j)^+$$
$$< I(U_A, X_{A\cup(B\cap T)}; Y|U_0, U_B, X_{B\setminus T}) - \zeta_{(A\cup B)\cap S_d}$$

3.5.2 Proof of Theorem 3.2.3 (Sum-capacity gain)

Fix any vector $\mathbf{v} \in \mathbb{R}^k_{>0}$, rate vector $\mathbf{C}_{\text{in}} \in \mathbb{R}^k_{>0}$, and cost constraint vector $\mathbf{B} \in \mathbb{R}^k_{\geq 0}$. For every $h \geq 0$, define $\mathbf{C}_{\text{out}}(h) = h\mathbf{v}$. In the achievable region defined in Section 3.1, let $\mathcal{U}_0 = \{0,1\}$, and for every $j \in [k]$, let $\mathcal{U}_j = \mathcal{X}_j$. Set $C_{j0} = 0$ and $C_{jd} = C_{\text{out}}^j(h)$ for every $j \in [k]$. For $h > 0$, let $\mathcal{P}(h)$ be the set of all distributions of the form

$$p(u_0, u_{[k]}) \cdot \prod_{j\in[k]} p(x_j|u_0, u_j)$$

that satisfy dependence constraints

$$\sum_{j\in S} C_{\text{out}}^j(h) - \sum_{j\in S} H(U_j|U_0) + H(U_S|U_0) > 0 \qquad \forall\, \varnothing \subsetneq S \subseteq [k],$$

and cost constraints

$$\mathbb{E}\big[b_j(X_j)\big] \leq B_j \qquad \forall j \in [k].$$

Since our MAC is in $\mathcal{C}^*$, for some distribution $p_a \in \mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})$ that satisfies

$$I_a(X_{[k]}; Y) = \max_{p\in\mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})} I(X_{[k]}; Y),$$

there exists a distribution $p_b \in \mathcal{P}(\mathcal{X}_{[k]})$ that satisfies

$$\mathbb{E}_b\Big[D\big(p(y|X_{[k]})\|p_a(y)\big)\Big] > \mathbb{E}_a\Big[D\big(p(y|X_{[k]})\|p_a(y)\big)\Big]$$

and whose support is contained in the support of p_a. Here we also assume that for all $j \in [k]$,

$$I_a(X_j; Y|X_{[k]\setminus\{j\}}) > 0.$$

At the end of the proof, we show that in the case where the last property does not hold, the same result follows by considering a MAC with a smaller number of users.

Choose $\mu \in (0,1)$ such that for every nonempty $S \subseteq [k]$,

$$\mu I_a(X_S; Y|X_{S^c}) < \sum_{j\in S} C_{\text{in}}^j. \tag{3.39}$$

For every $\lambda \in [0,1]$, define the distribution $p_\lambda(u_0, u_{[k]}, x_{[k]})$ as

$$p_\lambda(u_0, u_{[k]}, x_{[k]}) := p_\lambda(u_0) p_\lambda(u_{[k]}) p_\lambda(x_{[k]}|u_0, u_{[k]}),$$

where

$$p_\lambda(u_0) := \begin{cases} \mu & \text{if } u_0 = 1 \\ 1-\mu & \text{if } u_0 = 0, \end{cases}$$

and for every $u_{[k]} \in \mathcal{U}_{[k]}$ (recall $\mathcal{U}_{[k]} = \mathcal{X}_{[k]}$),

$$p_\lambda(u_{[k]}) := (1-\lambda) p_a(u_{[k]}) + \lambda p_b(u_{[k]}).$$

Finally, for every $(u_0, u_{[k]}, x_{[k]})$,

$$p_\lambda(x_{[k]}|u_0, u_{[k]}) := \prod_{j \in [k]} p_\lambda(x_j | u_0, u_j),$$

where for all $j \in [k]$,

$$p_\lambda(x_j|u_0, u_j) := \begin{cases} \mathbf{1}\{x_j = u_j\} & \text{if } u_0 = 1 \\ p_a(x_j) & \text{if } u_0 = 0. \end{cases}$$

Note that $p_\lambda(u_0)$ and $p_\lambda(x_{[k]}|u_0, u_{[k]})$ do not depend on λ. In addition, since p_a and p_b satisfy the cost constraints, and for all $j \in [k]$ and $\lambda \in [0,1]$,

$$p_\lambda(x_j) = (1-\lambda) p_a(x_j) + \lambda p_b(x_j),$$

p_λ satisfies the cost constraints as well.

We next find a function $\lambda^*(h)$ such that for sufficiently small h,

$$p_{\lambda^*(h)}(u_0, u_{[k]}, x_{[k]}) \in \mathcal{P}(h).$$

Fix $\epsilon > 0$, and define $h : [0,1] \to \mathbb{R}$ by

$$h(\lambda) := \frac{1}{\sum_{j \in [k]} v_j} \Big(\sum_{j \in [k]} H_\lambda(U_j) - H_\lambda(U_{[k]}) \Big) + \epsilon \lambda. \tag{3.40}$$

The following argument relies on Lemma 3.5.6, which appears at the end of this section. By Lemma 3.5.6 (i),

$$\frac{dh}{d\lambda} = \frac{1}{\sum_{j \in [k]} v_j} \sum_{u_{[k]}} \big(p_b(u_{[k]}) - p_a(u_{[k]}) \big) \log \frac{p_\lambda(u_{[k]})}{\prod_{j \in [k]} p_\lambda(u_j)} + \epsilon,$$

and by Lemma 3.5.6 (iii),

$$\frac{dh}{d\lambda}\Big|_{\lambda=0^+} = \epsilon > 0.$$

Thus $h(\lambda)$ is continuously differentiable and has positive derivative at $\lambda = 0^+$. Therefore, by the inverse function theorem, there exists a function $\lambda = \lambda^*(h)$ defined on $[0, h_0)$ for some $h_0 > 0$ that satisfies

$$h \sum_{j\in[k]} v_j = \sum_{j\in[k]} H_{\lambda^*(h)}(U_j) - H_{\lambda^*(h)}(U_{[k]}) + \epsilon\lambda^*(h) \sum_{j\in[k]} v_j,$$

and

$$\frac{d\lambda^*}{dh}\Big|_{h=0^+} = \frac{1}{\epsilon}. \tag{3.41}$$

Now for every nonempty $S \subseteq [k]$, define the function $\zeta_S : [0, h_0) \to \mathbb{R}$ as

$$\zeta_S(h) = \sum_{j\in S} C^j_{\text{out}}(h) - \sum_{j\in S} H_{\lambda^*}(U_j) + H_{\lambda^*}(U_S), \tag{3.42}$$

If we calculate the derivative of ζ_S at $h = 0^+$, by Lemma 3.5.6 (iii), we get

$$\frac{d\zeta_S}{dh}\Big|_{h=0^+} = \sum_{j\in S} v_j > 0.$$

This implies that there exists $0 < h_1 \leq h_0$ such that for every $0 < h < h_1$ and all nonempty $S \subseteq [k]$,

$$\zeta_S(h) > 0.$$

Therefore, for all sufficiently small h, $p_{\lambda^*(h)}(u_0, u_{[k]}, x_{[k]})$ is in $\mathcal{P}(h)$.

We next find a lower bound for the achievable sum-rate using the distribution $p_{\lambda^*}(u_0, u_{[k]}, x_{[k]})$ for small h. For every $S, T \subseteq [k]$, define the function $f_{S,T}$: $[0, h_1) \to \mathbb{R}$ as

$$f_{S,T}(h) := I_{\lambda^*(h)}(X_{S\cup T}; Y|U_0, U_{S^c}, X_{S^c\cap T^c}) + \sum_{j\in T\setminus S} C^j_{\text{in}} - \zeta_{[k]}(h). \tag{3.43}$$

This definition is motivated by the fact that by Lemma 3.5.4, any rate vector $(R_j)_{j\in[k]}$ that satisfies

$$\sum_{j\in S\cup T} R_j < f_{S,T}(h) \qquad \forall\, S, T \subseteq [k]$$

is in $\mathscr{C}(\mathbf{C}_{\text{in}}, h\mathbf{v})$. In (3.43), expanding the mutual information term with respect to U_0 gives

$$\begin{aligned} &I_{\lambda^*}(X_{S\cup T}; Y|U_0, U_{S^c}, X_{S^c\cap T^c}) \\ &\quad = \mu I_{\lambda^*}(X_S; Y|X_{S^c}) + (1-\mu) I_a(X_{S\cup T}; Y|X_{S^c\cap T^c}), \end{aligned}$$

where the term $I_{\lambda^*}(X_S;Y|X_{S^c})$ is calculated with respect to the distribution

$$p_{\lambda^*}(x_{[k]}) = (1-\lambda^*)p_a(x_{[k]}) + \lambda^* p_b(x_{[k]}).$$

Next, for every $S \subseteq [k]$, define the function $F_S : [0, h_1) \to \mathbb{R}$ as

$$F_S(h) := I_{\lambda^*}(X_S;Y|U_0, U_{S^c}, X_{S^c}) - \zeta_{[k]}(h).$$

The following argument shows that for sufficiently small h and for all $S, T \subseteq [k]$,

$$f_{S,T}(h) \geq F_{S\cup T}(h).$$

Consider some S and T for which $T \setminus S$ is not empty. Then

$$\begin{aligned} f_{S,T}(0) &= \mu I_a(X_S;Y|X_{S^c}) + (1-\mu)I_a(X_{S\cup T};Y|X_{S^c\cap T^c}) + \sum_{j\in T\setminus S} C_{\text{in}}^j \\ &> \mu I_a(X_S;Y|X_{S^c}) + (1-\mu)I_a(X_{S\cup T};Y|X_{S^c\cap T^c}) + \mu I_a(X_{T\setminus S};Y|X_{(T\setminus S)^c}) \qquad (3.44) \\ &\geq I_a(X_{S\cup T};Y|X_{S^c\cap T^c}) = F_{S\cup T}(0), \end{aligned}$$

where (3.44) follows from (3.39). Note that $f_{S,T}$ and $F_{S\cup T}$ are continuous functions of h for all S and T. Thus there exists $0 < h_2 \leq h_1$ such that for every $h \in [0, h_2)$ and $S, T \subseteq [k]$ with $T \setminus S \neq \varnothing$,

$$f_{S,T}(h) \geq F_{S\cup T}(h).$$

Next consider S and T for which $T \setminus S = \varnothing$; that is, $T \subseteq S$. In this case,

$$\begin{aligned} f_{S,T}(h) &= I_{\lambda^*}(X_{S\cup T};Y|U_0, U_{S^c}, X_{S^c\cap T^c}) + \sum_{j\in T\setminus S} C_{\text{in}}^j - \zeta_{[k]}(h) \\ &= I_{\lambda^*}(X_S;Y|U_0, U_{S^c}, X_{S^c}) - \zeta_{[k]}(h) \\ &= F_S(h) = F_{S\cup T}(h). \end{aligned}$$

Thus $f_{S,T}(h) \geq F_{S\cup T}(h)$ for all such S and T as well. Now fix $h \in [0, h_2)$. From the above argument, it follows that the region

$$\mathscr{C}_{\text{ach}}(h) := \Big\{ R_{[k]} \Big| \forall\, \varnothing \subsetneq S \subseteq [k] : 0 \leq \sum_{j\in S} R_j \leq F_S(h) \Big\}$$

is a subset of $\mathscr{C}(\mathbf{C}_{\text{in}}, h\mathbf{v})$. Now consider the region

$$\mathscr{C}_{\text{out}}(h) := \Big\{ R_{[k]} \Big| \forall\, \varnothing \subsetneq S \subseteq [k] : 0 \leq \sum_{j\in S} R_j \leq \Phi_S(h) \Big\},$$

where $\Phi_S(h)$ is defined as

$$\Phi_S(h) := F_S(h) + \zeta_{S^c}(h) + \sum_{j\in S} C^j_{\text{out}}(h).$$

Note that $\mathscr{C}_{\text{out}}(h)$ contains $\mathscr{C}_{\text{ach}}(h)$.

We next show that there exists $h_3 \in (0, h_2]$ such that for every $j \in [k]$ and all $h \in (0, h_3)$,

$$\Phi_{\{j\}}(h) > k \sum_{i\in[k]} C^i_{\text{out}}(h). \tag{3.45}$$

To see this, first note that the right hand side of the above equation equals zero at $h = 0$, while

$$\Phi_{\{j\}}(0) = I_a(X_j; Y|X_{[k]\setminus\{j\}}) > 0.$$

Inequality (3.45) now follows from the fact that both sides are continuous in h.

By Lemma 3.5.7, which appears at the end of this section, for a fixed h, the mapping $S \mapsto \Phi_S(h)$ is submodular and nondecreasing. Thus for every $j \in [k]$, there exists a rate vector $(R_i)_{i\in[k]}$ in $\mathscr{C}_{\text{out}}(h)$ such that

$$R_j > k \sum_{i\in[k]} C^i_{\text{out}}(h), \text{ and}$$
$$\sum_{i\in[k]} R_i = \Phi_{[k]}(h).$$

For example, for $j = 1$, consider the rate vector $(R_i)_{i\in[k]}$, where $R_1 = \Phi_{\{1\}}(h)$, and for all $1 < i \leq k$,

$$R_i = \Phi_{[i]} - \Phi_{[i-1]}.$$

From Lemma 3.5.8, it follows that the defined rate vector is in $\mathscr{C}_{\text{out}}(h)$. Since $\mathscr{C}_{\text{out}}(h)$ is a convex subset of $\mathbb{R}^k_{\geq 0}$, there exists a rate vector $(R^*_j(h))_j \in \mathscr{C}_{\text{out}}(h)$ such that

$$R^*_j(h) > \sum_{j\in[k]} C^j_{\text{out}}(h) \quad \forall\, j \in [k], \text{ and}$$
$$\sum_{j\in[k]} R^*_j(h) = \Phi_{[k]}(h).$$

On the other hand, from the definition of $\zeta_S(h)$, given by (3.42), it follows that

$$\Phi_S(h) \leq F_S(h) + \sum_{j\in[k]} C^j_{\text{out}}(h).$$

Thus

$$\left(R_j^*(h) - \sum_{i\in[k]} C_{\text{out}}^i(h)\right)_{j\in[k]} \in \mathscr{C}_{\text{ach}}(h).$$

This implies that the sum-rate

$$\begin{aligned} R_{\text{sum}}(h) &= \Phi_{[k]}(h) - k\sum_{j\in[k]} C_{\text{out}}^j(h) \\ &= \mu I_{\lambda^*}(X_{[k]};Y) + (1-\mu)I_a(X_{[k]};Y) - k\sum_{j\in[k]} C_{\text{out}}^j(h) \end{aligned}$$

is achievable. In addition, since by [36, Lemma 3.2] we have

$$R_{\text{sum}}(0) = I_a(X_{[k]};Y) = \max_{p\in\mathcal{P}_{\text{ind}}(\mathcal{X}_{[k]})} I(X_{[k]};Y) = C_{\text{sum}}(\mathbf{C}_{\text{in}},\mathbf{0}),$$

thus

$$C_{\text{sum}}(\mathbf{C}_{\text{in}},h\mathbf{v}) - C_{\text{sum}}(\mathbf{C}_{\text{in}},\mathbf{0}) \geq R_{\text{sum}}(h) - R_{\text{sum}}(0) \tag{3.46}$$

for all $h\in[0,h_3)$. Thus

$$\begin{aligned} &\liminf_{h\to 0^+} \frac{C_{\text{sum}}(\mathbf{C}_{\text{in}},h\mathbf{v}) - C_{\text{sum}}(\mathbf{C}_{\text{in}},\mathbf{0})}{h} \\ &\quad\geq \lim_{h\to 0^+} \frac{R_{\text{sum}}(h) - R_{\text{sum}}(0)}{h} \qquad (3.47)\\ &\quad= \mu\frac{d}{d\lambda^*} I_{\lambda^*}(X_{[k]};Y)\Big|_{\lambda^*=0^+} \times \frac{d\lambda^*}{dh}\Big|_{h=0^+} - k\sum_{j\in[k]} v_j \\ &\quad\geq \frac{\mu}{\epsilon}\left[\sum_{x_{[k]}} \big(p_b(x_{[k]}) - p_a(x_{[k]})\big) D\big(p(y|x_{[k]})\|p_a(y)\big)\right] - k\sum_{j\in[k]} v_j. \qquad (3.48) \end{aligned}$$

Here (3.47) follows from (3.46) and (3.48) is proved by combining (3.41) and Lemma 3.5.6 (ii), which appears at the end of this section. From our definitions of p_a and p_b, it follows that

$$\sum_{x_{[k]}} p_b(x_{[k]}) D\big(p(y|x_{[k]})\|p_a(y)\big) > \sum_{x_{[k]}} p_a(x_{[k]}) D\big(p(y|x_{[k]})\|p_a(y)\big).$$

Since ϵ is arbitrary, from (3.48) we get

$$\lim_{h\to 0^+} \frac{C_{\text{sum}}(\mathbf{C}_{\text{in}},h\mathbf{v}) - C_{\text{sum}}(\mathbf{C}_{\text{in}},\mathbf{0})}{h} = \infty.$$

This completes the proof for the case where

$$S_* := \left\{ j\in[k] \,\middle|\, I_a(X_j;Y|X_{[k]\setminus\{j\}}) > 0 \right\}$$

equals $[k]$. We next consider a MAC for which S_* is a strict subset of $[k]$ (i.e., $S_* \subsetneq [k]$).

For every $j \in [k]$, let $\mathcal{A}_j \subseteq \mathcal{X}_j$ denote the the support of $p_a(x_j)$. Then for nonempty $S \subseteq [k]$, the support of $p_a(x_S)$ is given by

$$\mathcal{A}_S := \prod_{j \in S} \mathcal{A}_j.$$

Note that

$$I_a(X_{S_*^c}; Y | X_{S_*}) \leq \sum_{j \in S_*^c} I_a(X_j; Y | X_{[k] \setminus \{j\}}) = 0.$$

Thus for every $x_{S_*} \in \mathcal{A}_{S_*}$,

$$I_a(X_{S_*^c}; Y | X_{S_*} = x_{S_*}) = 0,$$

which implies that for all $x_{[k]} \in \mathcal{A}_{[k]}$,

$$p(y|x_{[k]}) = p_a(y|x_{S_*}).$$

Note that since the support of p_b is contained in the support of p_a by assumption, it follows that for all nonempty $S \subseteq [k]$, the support of $p_b(x_S)$ is contained in $\mathcal{A}_S$.

Now consider the $|S_*|$-user MAC

$$\Big(\mathcal{A}_{S_*}, p_a(y|x_{S_*}), \mathcal{Y}\Big),$$

and the input distributions $p_{\text{ind}}(x_{S_*}) = p_a(x_{S_*})$ and $p_{\text{dep}}(x_{S_*}) = p_b(x_{S_*})$. Note that

$$I_{\text{ind}}(X_{S_*}; Y) = \max_{p \in \mathcal{P}(\mathcal{X}_{S_*})} I(X_{S_*}; Y),$$

and

$$\begin{aligned}
\mathbb{E}_{\text{dep}}\Big[D\big(p_a(y|X_{S_*}) \| p_{\text{ind}}(y)\big)\Big] &= \mathbb{E}_b\Big[D\big(p(y|X_{[k]}) \| p_a(y)\big)\Big] \\
&> \mathbb{E}_a\Big[D\big(p(y|X_{[k]}) \| p_a(y)\big)\Big] \\
&= \mathbb{E}_{\text{ind}}\Big[D\big(p_a(y|X_{S_*}) \| p_{\text{ind}}(y)\big)\Big].
\end{aligned}$$

Furthermore, for every $j \in S_*$,

$$I_{\text{ind}}(X_j; Y | X_{S_* \setminus \{j\}}) = I_a(X_j; Y | X_{[k] \setminus \{j\}}) > 0.$$

Thus this MAC satisfies all of the conditions under which we already proved Theorem 3.2.3. Suppose $\mathbf{v} = (v_j)_{j=1}^k \in \mathbb{R}_{>0}^k$. Define $\mathbf{v}^* = (v_j^*)_{j=1}^k \in \mathbb{R}_{>0}^k$ as

$$v_j^* := v_j \mathbf{1}\{j \in S_*\}.$$

Then

$$\lim_{h\to 0^+} \frac{C_{\text{sum}}(\mathbf{C}_{\text{in}}, h\mathbf{v}) - C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{0})}{h} = \lim_{h\to 0^+} \frac{C_{\text{sum}}(\mathbf{C}_{\text{in}}, h\mathbf{v}^*) - C_{\text{sum}}(\mathbf{C}_{\text{in}}, \mathbf{0})}{h} \stackrel{(\star)}{=} \infty,$$

where $(\star)$ follows from the fact that our $|S_*|$-user MAC satisfies all the required properties to imply an infinite directional derivative for sum-capacity.

We next provide the proofs for the lemmas we use in the above argument.

We begin by proving Lemma 3.5.4, which is used to obtain a lower bound on the sum-capacity.

Lemma 3.5.4. *For any MAC with a* $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$*-CF, let*

$$p(u_0, u_{[k]}) \cdot \prod_{j\in[k]} p(x_j|u_0, u_j)$$

be a distribution that satisfies

$$\zeta_S = \sum_{j\in S} C_{\text{out}}^j - \sum_{j\in S} H(U_j|U_0) + H(U_S|U_0) > 0 \qquad \forall\, \varnothing \subsetneq S \subseteq [k],$$

and $\mathbb{E}\big[b_j(X_j)\big] \le B_j$ *for all* $j \in [k]$*. Then* $\mathscr{C}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ *contains the set of all rate vectors* $(R_j)_{j\in[k]}$ *for which*

$$\sum_{j\in S\cup T} R_j < I(X_{S\cup T}; Y|U_0, U_{S^c}, X_{S^c\cap T^c}) + \sum_{j\in T\setminus S} C_{\text{in}}^j - \zeta_{[k]} \qquad \forall\, S, T \subseteq [k]$$

$$\sum_{j\in[k]} R_j < I(X_{[k]}; Y) - \zeta_{[k]}.$$

Proof. In Theorem 3.2.1, for every $j \in [k]$, set $C_{j0} = 0$ and $C_{jd} = C_{\text{out}}^j$. If in (3.8), for every $S, T \subseteq [k]$ with $S \cup T \ne \varnothing$, we choose $A = S$ and $B = S^c$, then we see that any rate vector $(R_j)_{j\in[k]}$ that satisfies

$$\sum_{j\in S\cup T} (R_j - C_{\text{in}}^j)^+ < I(X_{S\cup T}; Y|U_0, U_{S^c}, X_{S^c\cap T^c}) + \sum_{j\in T\setminus S} C_{\text{in}}^j - \zeta_{[k]} \qquad \forall\, S, T \subseteq [k]$$

$$\sum_{j\in[k]} R_j < I(X_{[k]}; Y) - \zeta_{[k]},$$

is achievable. To get the desired region, we apply the following lemma. The proof is simple and is omitted.

Lemma 3.5.5. *Let k be a positive integer. Fix $\gamma > 0$ and for every $j \in [k]$, let α_j be a real number. Then the vector $(x_j)_{j\in[k]}$ satisfies*

$$\sum_{j\in[k]} (x_j - \alpha_j)^+ < \gamma$$

if and only if for every nonempty $S \subseteq [k]$,

$$\sum_{j\in S} (x_j - \alpha_j) < \gamma.$$

□

The next lemma provides the derivative of the input-output mutual information and the total correlation [44], when calculated with respect to the convex combination of two distributions.

Lemma 3.5.6. *Consider two distributions p_a and p_b defined on the finite alphabet $\mathcal{X}_{[k]}$. For every $\lambda \in [0, 1]$, define the distribution p_λ on $\mathcal{X}_{[k]}$ as*

$$p_\lambda(x_{[k]}) = (1 - \lambda)p_a(x_{[k]}) + \lambda p_b(x_{[k]}).$$

Then the following statements are true.

(i) For every nonempty $S \subseteq [k]$, we have

$$\frac{d}{d\lambda} H_\lambda(X_S) = -\sum_{x_S} (p_b(x_S) - p_a(x_S)) \log p_\lambda(x_S).$$

(ii) For every finite alphabet k-user MAC $(\mathcal{X}_{[k]}, p(y|x_{[k]}), \mathcal{Y})$, we have

$$\frac{d}{d\lambda} I_\lambda(X_{[k]}; Y) = \sum_{x_{[k]}} \big(p_b(x_{[k]}) - p_a(x_{[k]})\big) D\big(p(y|x_{[k]}) \| p_\lambda(y)\big). \tag{3.49}$$

(iii) If p_a has the form

$$p_a(x_{[k]}) = \prod_{j\in[k]} p_a(x_j),$$

and the support of $p_a(x_{[k]})$ contains the support of $p_b(x_{[k]})$, then for every nonempty $S \subseteq [k]$,

$$\frac{d}{d\lambda}\Big(\sum_{j\in S} H_\lambda(X_j) - H_\lambda(X_S)\Big)\Big|_{\lambda=0^+} = 0. \tag{3.50}$$

Proof. Statement (i) follows by direct calculation.

For (ii), note that

$$p_\lambda(y) = (1-\lambda)p_a(y) + \lambda p_b(y).$$

Thus by (i),

$$\begin{aligned}\frac{d}{d\lambda}H_\lambda(Y) &= -\sum_y \big(p_b(y) - p_a(y)\big)(\log e + \log p_\lambda(y))\\ &= \sum_y \big(p_b(y) - p_a(y)\big)\log\frac{1}{p_\lambda(y)}\\ &= \sum_{x_{[k]}} \big(p_b(x_{[k]}) - p_a(x_{[k]})\big)\sum_y p(y|x_{[k]})\log\frac{1}{p_\lambda(y)}.\end{aligned}$$

Similarly,

$$\begin{aligned}&\frac{d}{d\lambda}H_\lambda(Y|X_{[k]})\\ &= \sum_{x_{[k]}} \big(p_b(x_{[k]}) - p_a(x_{[k]})\big)\sum_y p(y|x_{[k]})\log\frac{1}{p(y|x_{[k]})}.\end{aligned}$$

Taking the difference between these derivatives completes the proof of part (ii).

For part (iii), note that for every $j \in [k]$,

$$\begin{aligned}\frac{d}{d\lambda}H_\lambda(X_j) &= -\sum_{x_j}(p_b(x_j) - p_a(x_j))(\log e + \log p_\lambda(x_j))\\ &= -\sum_{x_j}(p_b(x_j) - p_a(x_j))\log p_\lambda(x_j).\end{aligned}$$

Hence

$$\begin{aligned}\frac{d}{d\lambda}\sum_{j\in S}H_\lambda(X_j) &= -\sum_{j\in S}\sum_{x_j}(p_b(x_j) - p_a(x_j))\log p_\lambda(x_j)\\ &= \sum_{x_S}(p_b(x_S) - p_a(x_S))\log\frac{1}{\prod_{j\in S}p_\lambda(x_j)}.\end{aligned}$$

On the other hand,

$$\frac{d}{d\lambda}H_\lambda(X_S) = -\sum_{x_S}(p_b(x_S) - p_a(x_S))\log p_\lambda(x_S).$$

Thus

$$\frac{d}{d\lambda}\Big(\sum_{j\in S}H_\lambda(X_j) - H_\lambda(X_S)\Big) = \sum_{x_S}(p_b(x_S) - p_a(x_S))\log\frac{p_\lambda(x_S)}{\prod_{j\in S}p_\lambda(x_j)}.$$

Equation (3.50) now follows from the fact that

$$p_a(x_S) = \prod_{j \in S} p_a(x_j),$$

and the support of p_b is contained in the support of p_a. $\square$

In the next lemma, we prove that for a fixed h, the mapping $S \mapsto \Phi_S(h)$ is nondecreasing and submodular. In the statement of this lemma, $2^{[k]}$ denotes the collection of all subsets of $[k]$. This lemma was proved for independent distributions by Han [36, Lemma 3.1].

Lemma 3.5.7. *Fix a distribution*

$$p(u_{[k]}) \cdot \prod_{j \in [k]} p(x_j|u_j) \cdot p(y|x_{[k]})$$

on $\mathcal{U}_{[k]} \times \mathcal{X}_{[k]} \times \mathcal{Y}$, *and define the function* $\Phi : 2^{[k]} \to \mathbb{R}$ *as*

$$\Phi(S) := I(X_S; Y|U_{S^c}X_{S^c}) + \sum_{j \in S} H(U_j) - H(U_S|U_{S^c})$$

for every $S \subseteq [k]$*. Then* Φ *is nondecreasing and submodular.*

Proof. Note that

$$\Phi(S) = H(Y|U_{S^c}X_{S^c}) - H(Y|X_{[k]}) + \sum_{j \in S} H(U_j) + H(U_{S^c}) - H(U_{[k]}).$$

For every $j \in [k]$, define $V_j := (U_j, X_j)$. Then for every $S \subseteq [k]$,

$$\begin{aligned}\sum_{j \in S} H(V_j) + H(V_{S^c}) - H(V_{[k]}) &= \sum_{j \in S} H(U_j, X_j) + H(U_{S^c}, X_{S^c}) - H(U_{[k]}, X_{[k]}) \\ &= \sum_{j \in S} H(U_j) + H(U_{S^c}) - H(U_{[k]}),\end{aligned}$$

since each X_j only depends on U_j. Thus

$$\begin{aligned}\Phi(S) &= H(Y|V_{S^c}) - H(Y|V_{[k]}) + \sum_{j \in S} H(V_j) + H(V_{S^c}) - H(V_{[k]}) \\ &= H(V_{S^c}|Y) + \sum_{j \in S} H(V_j) - H(V_{[k]}|Y).\end{aligned}$$

We first show Φ is nondecreasing; that is, we show $\Phi(S) \subseteq \Phi(T)$ whenever $S \subseteq T$. Suppose $S \subseteq T$. Then

$$\begin{aligned}
&H(V_{S^c}|Y) + \sum_{j\in S} H(V_j) \\
&\quad = H(V_{T^c}|Y) + H(V_{S^c\setminus T^c}|V_{T^c}, Y) + \sum_{j\in T} H(V_j) - \sum_{j\in T\setminus S} H(V_j) \\
&\quad \le H(V_{T^c}|Y) + \sum_{j\in T} H(V_j),
\end{aligned}$$

since

$$H(V_{S^c\setminus T^c}|V_{T^c}, Y) = H(V_{T\setminus S}|V_{T^c}, Y) \le \sum_{j\in T\setminus S} H(V_j).$$

Thus $\Phi(S) \subseteq \Phi(T)$.

We next show Φ is submodular. Fix $S, T \subseteq [k]$. Our aim is to prove

$$\Phi(S) + \Phi(T) \ge \Phi(S \cup T) + \Phi(S \cap T). \tag{3.51}$$

We have

$$\begin{aligned}
H(V_{S^c}|Y) + H(V_{T^c}|Y) &= H(V_{S^c\cap T^c}|Y) + H(V_{S^c\setminus T^c}|V_{S^c\cap T^c}, Y) \\
&\quad + H(V_{S^c\cup T^c}|Y) - H(V_{S^c\setminus T^c}|V_{T^c}, Y) \\
&= H(V_{S^c\cap T^c}|Y) + H(V_{S^c\cup T^c}|Y) \\
&\quad + I(V_{S^c\setminus T^c}; V_{T^c\setminus S^c}|V_{S^c\cap T^c}, Y) \\
&\ge H(V_{S^c\cap T^c}|Y) + H(V_{S^c\cup T^c}|Y).
\end{aligned}$$

This proves (3.51), since

$$\sum_{j\in S} H(V_j) + \sum_{j\in T} H(V_j) = \sum_{j\in S\cup T} H(V_j) + \sum_{j\in S\cap T} H(V_j).$$

□

For a rate region defined by submodular constraints, the next lemma gives an explicit formula for a rate vector that achieves the maximum sum-rate. It is a special case of [45, Corollary 44.3a, p. 772] and is included here for completeness.

Lemma 3.5.8. *Let $\Phi : 2^{[k]} \to \mathbb{R}_{\ge 0}$ be a nondecreasing submodular function with $\Phi(\varnothing) = 0$. Define the region $\mathscr{R}(\Phi) \subseteq \mathbb{R}^k_{\ge 0}$ as*

$$\mathscr{R}(\Phi) := \left\{ R_{[k]} \middle| \forall\, S \subseteq [k] : \sum_{j\in S} R_j \le \Phi(S) \right\}.$$

Then the rate vector $R_{[k]} = (R_1, \dots, R_k)$*, defined as*

$$R_j := \begin{cases} \Phi\big([j]\big) - \Phi\big([j-1]\big) & \text{if } j \in [k] \setminus \{1\} \\ \Phi\big(\{1\}\big) & \text{if } j = 1, \end{cases}$$

is in $\mathscr{R}(\Phi)$.

Proof. Using induction on $|S|$, we show that the rate vector $R_{[k]}$ satisfies

$$\forall\, S \subseteq [k] : \sum_{j \in S} R_j \leq \Phi(S).$$

If $|S| = 0$, then $S = \varnothing$ and the proof is clear. For nonempty $S \subseteq [k]$, let $j := \max S$. By the induction hypothesis,

$$\sum_{i \in S \setminus \{j\}} R_i \leq \Phi\big(S \setminus \{j\}\big).$$

Thus

$$\begin{aligned} \sum_{i \in S} R_i &\leq \Phi\big(S \setminus \{j\}\big) + R_j \\ &= \Phi\big(S \setminus \{j\}\big) + \Phi\big([j]\big) - \Phi\big([j-1]\big) \\ &\leq \Phi(S), \end{aligned} \tag{3.52}$$

where (3.52) follows from the fact that Φ is submodular. □

3.5.3 Proof of Proposition 3.2.4 (The k-user Gaussian MAC)

A close inspection of the proof of Theorem 3.2.3 (Subsection 3.5.2) reveals that the proof applies to any discrete or continuous MAC for which there exists a family of input distributions $(p_\lambda(x_{[k]}))_{\lambda \in [0,1]}$ such that

(i) for $\lambda = 0$,

$$\begin{aligned} p_0(x_{[k]}) &= \prod_{j \in [k]} p_0(x_j), \text{ and} \\ I_0(X_{[k]}; Y) &= \max_{p(x_1) \dots p(x_k)} I(X_{[k]}; Y); \end{aligned}$$

(ii) for every nonempty $S \subseteq [k]$,

$$\frac{d}{d\lambda} H_\lambda(X_S)\Big|_{\lambda = 0^+} = 0; \text{ and}$$

(iii) the input-output mutual information satisfies

$$\frac{d}{d\lambda} I_\lambda(X_{[k]};Y)\Big|_{\lambda=0^+} > 0.$$

For the Gaussian MAC with power constraints $(P_j)_{j\in[k]}$, define $p_\lambda(x_{[k]})$ as the probability density function of a multivariate Gaussian distribution with mean vector zero and covariance matrix Σ_λ, where for $i,j \in [k]$, the (i,j)-entry of Σ_λ is given by

$$\Sigma_\lambda(i,j) = \begin{cases} \lambda\sqrt{P_i P_j} & \text{if } i \neq j \\ P_i & \text{if } i = j. \end{cases}$$

Clearly, condition (i) is satisfied. Next, note that for every nonempty $S \subseteq [k]$, we have

$$H_\lambda(X_S) = \frac{1}{2}\log\Big[(2\pi e)^{|S|}\Big(\prod_{j\in[k]} P_j\Big)(1-\lambda)^{|S|-1}\big(1+(|S|-1)\lambda\big)\Big],$$

from which (ii) follows by a simple calculation. Finally, defining the SNR of encoder j as $\gamma_j := P_j/N$ for $j \in [k]$ gives

$$I_\lambda(X_{[k]};Y) = \frac{1}{2}\log\Big(1 + \sum_{i\in[k]}\gamma_i + \lambda\sum_{i,j:i\neq j}\sqrt{\gamma_i\gamma_j}\Big),$$

from which (iii) follows if there exist distinct $i,j \in [k]$ such that $\gamma_i\gamma_j > 0$.

3.5.4 Proof of Proposition 3.2.5 (Outer bound)

Consider a $\big((2^{nR_1},\ldots,2^{nR_k}),n,L\big)$-code for the MAC with a $(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$-CF. For every message vector $w_{[k]} = (w_1,\ldots,w_k)$, $j\in[k]$, and $\ell \in [L]$, define

$$\begin{aligned} u_{j\ell} &:= \varphi_{j\ell}(w_j, v_j^{\ell-1}) \\ v_{j\ell} &:= \psi_{j\ell}(u_1^\ell,\ldots,u_k^\ell), \end{aligned}$$

where $u_j^\ell := (u_{j1},\ldots,u_{j\ell})$ and $v_j^\ell := (v_{j1},\ldots,v_{j\ell})$. Also, for every nonempty $S \subseteq [k]$ and $\ell \in [L]$, let $u_{S\ell} = (u_{j\ell})_{j\in S}$ and $u_S^\ell = (u_j^\ell)_{j\in S}$. Finally, for every $j\in[k]$, $\ell\in[L]$, and $v_j^{\ell-1} \in \mathcal{V}_j^{\ell-1}$, define the mapping

$$\begin{aligned} \varphi^{-1}_{j\ell,v_j^{\ell-1}} : \mathcal{U}_{j\ell} &\to 2^{[2^{nR_j}]} \\ u_{j\ell} &\mapsto \Big\{w_j \Big| \varphi_{j\ell}(w_j,v_j^{\ell-1}) = u_{j\ell}\Big\}, \end{aligned}$$

where $2^{[2^{nR_j}]}$ denotes the set of all the subsets of $[2^{nR_j}]$.

Note that $v_{[k]}^L$ is a deterministic function of $u_{[k]}^L$. Thus for every $u_{[k]}^L$ and $j \in [k]$, the set

$$\mathcal{A}_j(u_{[k]}^L) := \bigcap_{\ell=1}^{L} \varphi_{j\ell, v_j^{\ell-1}}^{-1}(u_{j\ell})$$

is well-defined. It follows that for a fixed code and a given message vector $w_{[k]}$, the vector of all CF inputs is given by $u_{[k]}^L$ if and only if for every $j \in [k]$, $w_j \in \mathcal{A}_j(u_{[k]}^L)$.

Now consider a sequence of $\big((2^{nR_1}, \ldots, 2^{nR_k}), n, L\big)$-codes with $P_e^{(n)} \to 0$ as $n \to \infty$. By Fano's inequality [40, p. 38], there exists a sequence $(\epsilon_n)_{n=1}^{\infty}$ such that $\epsilon_n \to 0$ as $n \to \infty$ and

$$H(W_{[k]}|Y^n) \leq n\epsilon_n.$$

Thus for every nonempty subset $S \subseteq [k]$,

$$H(W_S|W_{S^c}, U_{[k]}^L, Y^n) \leq n\epsilon_n.$$

We have

$$\begin{aligned} n\sum_{j\in S} R_j &\leq H(W_S|W_{S^c}) \\ &= I(W_S; U_{[k]}^L, Y^n|W_{S^c}) + H(W_S|W_{S^c}, U_{[k]}^L, Y^n) \\ &\leq I(W_S; U_{[k]}^L|W_{S^c}) + I(W_S; Y^n|W_{S^c}, U_{[k]}^L) + n\epsilon_n. \end{aligned} \tag{3.53}$$

We next find an upper bound for each of the mutual information terms. For the first term, we have

$$\begin{aligned} I(W_S; U_{[k]}^L|W_{S^c}) &= H(U_{[k]}^L|W_{S^c}) && (3.54) \\ &= \sum_{\ell=1}^{L} H(U_{[k]\ell}|W_{S^c}, U_{[k]}^{\ell-1}) \\ &= \sum_{\ell=1}^{L} H(U_{S\ell}, U_{S^c\ell}|W_{S^c}, U_{[k]}^{\ell-1}) \\ &= \sum_{\ell=1}^{L} H(U_{S\ell}|W_{S^c}, U_{[k]}^{\ell-1}, U_{S^c\ell}) \leq \sum_{j\in S} C_{\text{in}}^j, && (3.55) \end{aligned}$$

where (3.54) follows from the fact that $U_{[k]}^L$ is a deterministic function of $W_{[k]}$, and (3.55) follows from the fact that $U_{S^c\ell}$ is a deterministic function of $(W_{S^c}, U_{[k]}^{\ell-1})$. For the second term in (3.53), we have

$$\begin{aligned} I(W_S; Y^n | W_{S^c}, U_{[k]}^L) &= H(Y^n | W_{S^c}, U_{[k]}^L) - H(Y^n | W_S, W_{S^c}, U_{[k]}^L) \\ &= H(Y^n | U_{[k]}^L, X_{S^c}^n) - H(Y^n | U_{[k]}^L, X_{[k]}^n) \\ &\le \sum_{t=1}^n \Big(H(Y_t | X_{S^c t}, U_{[k]}^L) - H(Y_t | U_{[k]}^L, X_{[k]t}) \Big) \\ &\le \sum_{t=1}^n I(X_{St}; Y_t | U_{[k]}^L, X_{S^c t}), \end{aligned}$$

where $X_{St} = (X_{jt})_{j \in S}$. We have

$$p(u_{[k]}^L) = \Pr\left\{ \forall j \in [k] : W_j \in \mathcal{A}_j(u_{[k]}^L) \right\} = \prod_{j \in [k]} \frac{|\mathcal{A}_j(u_{[k]}^L)|}{|\mathcal{W}_j|},$$

and

$$p(u_{[k]}^L | w_j) = \mathbf{1}\left\{ w_j \in \mathcal{A}_j(u_{[k]}^L) \right\} \prod_{i : i \ne j} \frac{|\mathcal{A}_i(u_{[k]}^L)|}{|\mathcal{W}_i|}.$$

Thus

$$p(w_j | u_{[k]}^L) = \frac{p(w_j) p(u_{[k]}^L | w_j)}{p(u_{[k]}^L)} = \frac{\mathbf{1}\left\{ w_j \in \mathcal{A}_j(u_{[k]}^L) \right\}}{|\mathcal{A}_j(u_{[k]}^L)|}$$

and

$$p(w_{[k]} | u_{[k]}^L) = \frac{p(w_{[k]}) p(u_{[k]}^L | w_{[k]})}{p(u_{[k]}^L)} = \frac{\prod_{j \in [k]} \mathbf{1}\left\{ w_j \in \mathcal{A}_j(u_{[k]}^L) \right\}}{\prod_{j \in [k]} |\mathcal{A}_j(u_{[k]}^L)|} = \prod_{j \in [k]} p(w_j | u_{[k]}^L).$$

Therefore, $W_1, \ldots, W_k$ are independent given $U_{[k]}^L$. Recall that at time $t \in [n]$, the output of encoder j is given by $X_{jt} := f_{jt}(W_j, V_j^L)$ for some mapping

$$f_{jt} : [2^{nR_j}] \times \mathcal{V}_j^L \to \mathcal{X}_j.$$

Also define $U_{0t} := U_{[k]}^L$ for all $t \in [n]$. Then

$$\begin{aligned} p(x_{[k]t} | u_{0t}) &= \sum_{w_{[k]}} p(w_{[k]} | u_{0t}) p(x_{[k]t} | w_{[k]}, u_{0t}) \\ &= \sum_{w_{[k]}} \prod_{j \in [k]} p(w_j | u_{0t}) p(x_{jt} | w_j, u_{0t}) \\ &= \prod_{j \in [k]} \sum_{w_j} p(w_j | u_{0t}) p(x_{jt} | w_j, u_{0t}) = \prod_{j \in [k]} p(x_{jt} | u_{0t}). \end{aligned}$$

Defining a time sharing random variable and applying the usual time sharing argument [40, p. 600] completes the proof.

3.5.5 Proof of Proposition 3.4.2 (The Gaussian MAC)

Consider any $\alpha \in [0, 1/2]$. In the region given in Section 3.4, set $C_{10} = C_{20} = 0$, $C_{1d} = C_{2d} = C_{\text{out}}$, $\rho_1 = \rho_2 = 1$, and

$$\rho_0 = \sqrt{1 - 2^{-4C_{\text{out}}}}.$$

Then the rate pair (R_1^*, R_2^*) given by

$$\begin{aligned} R_1^* &= \frac{1}{2} \log\Big(\frac{1 + \gamma_1 + \gamma_2 + 2\rho_0\bar{\gamma}}{1 + (1 - \rho_0^2)\gamma_2}\Big) - C_{\text{out}} \\ R_2^* &= \frac{1}{2} \log\big(1 + (1 - \rho_0^2)\gamma_2\big), \end{aligned}$$

is achievable. Since

$$\begin{aligned} C_\alpha(0) &= \alpha \times \frac{1}{2} \log\Big(\frac{1 + \gamma_1 + \gamma_2}{1 + \gamma_2}\Big) + (1 - \alpha) \times \frac{1}{2} \log(1 + \gamma_2) \\ &= \frac{\alpha}{2} \log(1 + \gamma_1 + \gamma_2) + \frac{1 - 2\alpha}{2} \log(1 + \gamma_2), \end{aligned}$$

we have

$$\begin{aligned} & C_\alpha(C_{\text{out}}) - C_\alpha(0) \\ & \quad \geq \alpha R_1^* + (1 - \alpha) R_2^* - C_\alpha(0) && (3.56) \\ & \quad = \frac{\alpha}{2} \log\Big(1 + \frac{2\rho_0\bar{\gamma}}{1 + \gamma_1 + \gamma_2}\Big) + \frac{1 - 2\alpha}{2} \log\Big(1 - \frac{\rho_0^2\gamma_2}{1 + \gamma_2}\Big) - C_{\text{out}}. && (3.57) \end{aligned}$$

Using the fact that $2^x = 1 + \frac{x}{\log e} + o(x)$ and $\sqrt{1 + o(1)} = 1 + o(1)$, we get

$$\begin{aligned} \rho_0 &= \sqrt{1 - 2^{-4C_{\text{out}}}} \\ &= \sqrt{\frac{4C_{\text{out}}}{\log e} + o(C_{\text{out}})} \\ &= \frac{2}{\sqrt{\log e}} \cdot \sqrt{C_{\text{out}}} + o(\sqrt{C_{\text{out}}}). \end{aligned}$$

In addition,

$$\rho_0^2 = \frac{4C_{\text{out}}}{\log e} + o(C_{\text{out}}) = o(\sqrt{C_{\text{out}}}).$$

Applying $\log(1 + x) = x \log e + o(x)$ to (3.56) completes the proof for $\alpha \in [0, 1/2]$. The proof for $\alpha \in (1/2, 1]$ follows similarly.

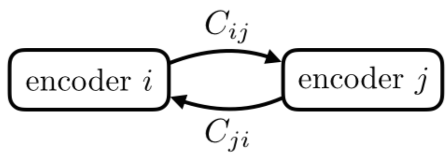

Figure 3.3: In k-user MAC with conferencing, for every $i, j \in [k]$, there are links of capacities C_{ij} and C_{ji} connecting encoders i and j.

3.6 Appendix: The k-User MAC with Conferencing Encoders

In this appendix, we extend Willems' conferencing encoders model [5] from the 2-user MAC to the k-user MAC and provide an outer bound on the capacity region.

Consider a k-user MAC where for every $i, j \in [k]$ (in this section, $i \neq j$ by assumption), there is a noiseless link of capacity $C_{ij} \geq 0$ going from encoder i to encoder j and a noiseless link of capacity $C_{ji} \geq 0$ going back (Figure 3.3). As in 2-user conferencing, the "conference" occurs over a finite number of rounds. In the first round, for every $i, j \in [k]$ with $C_{ij} > 0$, encoder i transmits some information to encoder j that is a function of its own message $w_i \in [2^{nR_i}]$. In each subsequent round, every encoder transmits information that is a function of its message and information it receives before that round. Once the conference is complete, each encoder transmits its codeword over the k-user MAC.

We next define a $\big((2^{nR_1}, \ldots, 2^{nR_k}), n, L\big)$-code for the k-user MAC with an L-round $(C_{ij})_{i,j=1}^k$-conference. For every $i, j \in [k]$ and $\ell \in [L]$, fix a set $\mathcal{V}_{ij}^{(\ell)}$ so that for every $i, j \in [k]$,

$$\sum_{\ell=1}^{L} \log |\mathcal{V}_{ij}^{(\ell)}| \leq nC_{ij}.$$

Here $\mathcal{V}_{ij}^{(\ell)}$ represents the alphabet of the symbol encoder i sends to encoder j in round ℓ of the conference. For every $\ell \in [L]$, define $\mathcal{V}_{ij}^{\ell} = \prod_{\ell'=1}^{\ell} \mathcal{V}_{ij}^{(\ell')}$. For $j \in [k]$, encoder j is represented by the collection of functions $\big(f_j, (h_{ji}^{(\ell)})_{i,\ell}\big)$, where

$$f_j : [2^{nR_j}] \times \prod_{i:i\neq j} \mathcal{V}_{ij}^{L} \to \mathcal{X}_j^n$$
$$h_{ji}^{(\ell)} : [2^{nR_j}] \times \prod_{i':i'\neq j} \mathcal{V}_{i'j}^{\ell-1} \to \mathcal{V}_{ji}^{(\ell)}.$$

The decoder is a mapping $g : \mathcal{Y}^n \to \prod_{j=1}^k [2^{nR_j}]$. The definitions of cost constraints, achievable rate vectors, and the capacity region are similar to those given in Section 3.1.

The next result compares the capacity region of a MAC with cooperation under the conferencing and CF models.

Proposition 3.6.1. *The capacity region of a MAC with an L-round $(C_{ij})_{i,j=1}^k$-conference is a subset of the capacity region of the same MAC with an L-round $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF cooperation if for all $j \in [k]$,*

$$C_{\text{in}}^j \geq \sum_{i:i\neq j} C_{ji} \quad \textit{and} \quad C_{\text{out}}^j \geq \sum_{i:i\neq j} C_{ij}.$$

Similarly, for every L, the capacity region of a MAC with L-round $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF cooperation is a subset of the capacity region of the same MAC with a single-round $(C_{ij})_{i,j=1}^k$-conference if for all $i, j \in [k]$, $C_{ij} \geq C_{\text{in}}^i$.

Proof. An L-round $(C_{ij})_{i,j=1}^k$-conference for a blocklength-n code is uniquely determined by a collection of sets $\{\mathcal{W}_{ij}^{(\ell)}\}_{i,j,\ell}$ and mappings

$$h_{ji}^{(\ell)} : [2^{nR_j}] \times \prod_{i':i'\neq j} \mathcal{W}_{i'j}^{\ell-1} \to \mathcal{W}_{ji}^{(\ell)}, \tag{3.58}$$

where $i, j \in [k]$ and $\ell \in [L]$. In (3.58), for every $\ell \in [L]$,

$$\mathcal{W}_{ij}^{\ell} := \prod_{\ell'=1}^{\ell} \mathcal{W}_{ij}^{(\ell')}.$$

Furthermore, for all $i, j \in [k]$ and $\ell \in [L]$, $W_{ij}^{(\ell)}$ satisfies

$$\sum_{\ell\in[L]} \log |\mathcal{W}_{ij}^{(\ell)}| \leq nC_{ij}.$$

Finally, for every message vector $(m_1, \ldots, m_k)$, where $m_j \in [2^{nR_j}]$, define $w_{ji}^{(\ell)}$ recursively as

$$w_{ji}^{(\ell)} = h_{ji}^{(\ell)}\Big(m_j, \big(w_{i'j}^{\ell-1}\big)_{i'\neq j}\Big).$$

Our aim is to construct a blocklength-n code for the same MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF that through L rounds of communication with the encoders, provides them with the same information as the L-round conference given above. To this end, for every $j \in [k]$ and $\ell \in [L]$ define the sets $\mathcal{U}_{j\ell}$ and $\mathcal{V}_{j\ell}$ as

$$\mathcal{U}_{j\ell} := \prod_{i:i\neq j} \mathcal{W}_{ji}^{(\ell)}$$
$$\mathcal{V}_{j\ell} := \prod_{i:i\neq j} \mathcal{W}_{ij}^{(\ell)}.$$

Then

$$\begin{aligned}\sum_{\ell\in[L]} \log|\mathcal{U}_{j\ell}| &= \sum_{\ell\in[L]}\sum_{i:i\neq j} \log|\mathcal{W}_{ji}^{(\ell)}| \\ &= \sum_{i:i\neq j}\sum_{\ell\in[L]} \log|\mathcal{W}_{ji}^{(\ell)}| \\ &\leq n\sum_{i:i\neq j} C_{ji} \leq nC_{\mathrm{in}}^{j}.\end{aligned}$$

Similarly, we show

$$\sum_{\ell\in[L]} \log|\mathcal{V}_{j\ell}| \leq n\sum_{i:i\neq j} C_{ij} \leq nC_{\mathrm{out}}^{j}.$$

Next for every $j \in [k]$ and $\ell \in [L]$, define the mapping

$$\begin{aligned}\varphi_{j\ell} &: [2^{nR_j}] \times \mathcal{V}_j^{\ell-1} \to \mathcal{U}_{j\ell} \\ &\left(m_j, \left(w_{ij}^{\ell-1}\right)_{i:i\neq j}\right) \mapsto \left(w_{ji}^{(\ell)}\right)_{i:i\neq j}.\end{aligned}$$

Similarly, define

$$\begin{aligned}\psi_{j\ell} &: \prod_{i\in[k]} \mathcal{U}_i^{\ell} \to \mathcal{V}_{j\ell} \\ &\left(w_{ij'}^{\ell}\right)_{i,j'} \mapsto \left(w_{ij}^{(\ell)}\right)_{i:i\neq j}.\end{aligned}$$

This completes the proof of the first part.

For the second part, we show that the capacity region of a MAC with a single-round $(C_{ij})_{i,j}$-conference contains the outer bound given in Proposition 3.2.5 if $C_{ij} \geq C_{\mathrm{in}}^{i}$ for all $i, j \in [k]$. The coding strategy is simple. For each $j \in [k]$, encoder j sends the first nC_{in}^{j} bits of its message to all other encoders. The encoders then form a "common message" that contains the initial nC_{in}^{j} bits of message j for all $j \in [k]$. The rest of the proof follows by applying the forwarding inner bound (Corollary 3.2.2) with $C_{j0} = C_{\mathrm{in}}^{j}$ for all $j \in [k]$. □

Combining the first part of Proposition 3.6.1 with the outer bound from Proposition 3.2.5 results in the next corollary, which holds regardless of the number of conferencing rounds.

Corollary 3.6.2 (Conferencing Outer Bound)**.** *The capacity region of a MAC with a* $(C_{ij})_{i,j=1}^k$*-conference is a subset of the closure of the set of all rate vectors* $(R_1, \ldots, R_k)$ *that for some distribution* $p(u_0)p(x_1|u_0)\ldots p(x_k|u_0)$ *satisfy*

$$\sum_{j\in S} R_j \leq I\big(X_S; Y|U_0, X_{S^c}\big) + \sum_{j\in S}\sum_{i\neq j} C_{ji} \qquad \forall\, \varnothing \neq S \subseteq [k]$$

$$\sum_{j\in[k]} R_j \leq I(X_{[k]}; Y).$$

Chapter 4

THE ROLE OF STATE INFORMATION

In this chapter, we extend the exploration of cooperation beyond the networks of Chapters 2 and 3 to examine the cost-benefit tradeoff of cooperation in networks where state information is present at some nodes.

4.1 Channel State Information

Networks where state information is available at some nodes appear in many applications, including wireless channels with fading [9], [47], cognitive radios [48], and computer memory with defects [49]. Depending on the application at hand, the state information may be either fully available at all network nodes or available in a distributed manner; in the latter case, each node has access to a component or a function of the state sequence. Furthermore, the state information may be available non-causally, or alternatively, may be subject to causality constraints. For example, when state information models fading effects in wireless communication [9], the transmitters' knowledge of state information is strictly causal or causal. On the other hand, when the state sequence models a signal that the transmitter sends to another receiver, then the state sequence is available non-causally at the transmitter [50].

In this chapter, we study the advantage of *encoder cooperation* in the setting of networks with state information. In this context, network nodes work together to increase transmission rates—not only by sharing message information, but also by sharing state information (Figure 4.1). As an example of message and state cooperation, Permuter, Shamai, and Somekh-Baruch [11] find the capacity region of the MAC with encoder cooperation under the assumption that distributed, non-causal state information is available at the encoders and full state information is available at the decoder. As their cooperation model, the authors use a special case of the Willems conferencing model [5], originally defined for MACs in the absence of state information.

Indirect forms of cooperation, in the presence of state information, are also considered in the literature. Cemal and Steinberg [51] study a model where a

This material is based upon work supported by the National Science Foundation under Grant Numbers 1527524 and 1526771. It originally appears in [46].

central state-encoder sends rate limited versions of non-causal state information to each encoder, while the decoder has access to full state information.

Here we study cooperation under the CF model. Specifically, we characterize channels for which the cooperation gain has an infinite slope in the presence of state information (Section 4.4); interestingly, this includes channels for which the infinite slope phenomenon did not arise in the absence of state information.[1]

For state information at the encoders we consider four cases: (i) no state information, (ii) strictly causal state information, (iii) causal state information, and (iv) non-causal state information. In case (i), the CF is used for sharing message information (a strategy here called "message cooperation") since no state information is available at the encoders. In cases (ii)-(iv), the CF enables both message and state cooperation. Here we study message and state cooperation only in case (iv); in this case we show that the use of joint message and state cooperation leads to a weaker sufficient condition for an infinite-slope gain compared to the sole use of message cooperation. Whether in cases (ii) and (iii), the use of joint message and state cooperation likewise leads to a weaker sufficient condition for an infinite-slope gain compared to message cooperation alone, remains an open problem.

Throughout, we assume that any state information available at the encoders is distributed; that is, we assume $S = (S_1, S_2)$, where for $i \in \{1, 2\}$, S_i is available at encoder i. As we do not make any assumptions regarding the dependence between S_1 and S_2, our results apply to the limiting cases of independent states (i.e., independent S_1 and S_2) and common state (i.e., $S_1 = S_2$).

Since the decoder starts the decoding process only after receiving all the output symbols in a given transmission block, causality constraints at the decoder do not impose limitations on the availability of state information. Thus we may assume that the decoder either has full state information or no state information. Here we focus on the former scenario. Jafar [10] provides the capacity region of the MAC with distributed independent (causal or non-causal) state information at the encoders and full state information at the decoder. The

[1]As an example, consider the MAC $Y = X_1 + X_2 + S \pmod 3$, where S is uniform on $\{0, 1, 2\}$, X_1 and X_2 are binary, and Y is ternary. The infinite slope sum-capacity gain is achievable when the decoder has full knowledge of S, but no sum-capacity gain is possible when it does not have access to S.

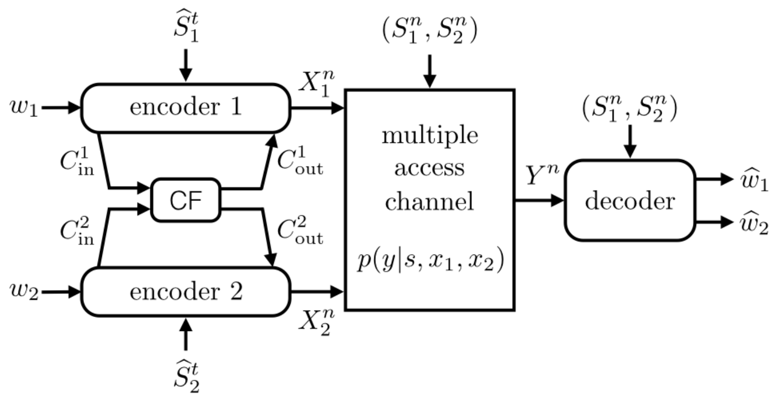

Figure 4.1: The network studied here consists of a pair of encoders communicating, with the help of a CF, to a decoder through a state-dependent MAC. Full state information is available at the decoder. Partial state information $\hat{S}_i^t$ is available to encoder $i \in \{1, 2\}$ at time $t \in [n]$.

capacity region is unknown when the encoders have access to state information but the decoder does not [52], [53].

4.2 Model

4.2.1 Preliminaries

Let $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ be discrete or continuous alphabets. A MAC with input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, output alphabet $\mathcal{Y}$, and state alphabet $\mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2$ is given by the sequence

$$\Big\{p(s^n)p(y^n|s^n, x_1^n, x_2^n)\Big\}_{n=1}^{\infty}.$$

The MAC is said to be memoryless and stationary if for some $p(s)p(y|s, x_1, x_2)$ and all positive integers n,

$$p(s^n)p(y^n|s^n, x_1^n, x_2^n) = \prod_{t=1}^{n} p(s_t)p(y_t|s_t, x_{1t}, x_{2t}).$$

4.2.2 Message Cooperation

In this subsection, we define the capacity region of a MAC with a CF that enables message cooperation. We include four scenarios in our definition based on the availability of state information at the encoders: no state, strictly causal, causal, and non-causal. We assume full state information is available at the decoder.

We start by defining a $(2^{nR_1}, 2^{nR_2}, n)$-code for the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF, cost functions $b_i : \mathcal{X}_i \to \mathbb{R}_{\geq 0}$ for $i \in \{1,2\}$, and cost constraints $B_1, B_2 \geq 0$. The pairs $\mathbf{C}_{\text{in}} = (C_{\text{in}}^1, C_{\text{in}}^2)$ and $\mathbf{C}_{\text{out}} = (C_{\text{out}}^1, C_{\text{out}}^2)$ denote the CF input and output edge capacities, respectively. Encoder i, for $i \in \{1,2\}$, is represented by $(\varphi_{\text{in}}^i, (f_{it})_{t=1}^n)$, the CF is represented by $(\varphi_{\text{out}}^1, \varphi_{\text{out}}^2)$, and the decoder is represented by g. These mappings are defined in the order of their use below. For $i \in \{1,2\}$, the transmission from encoder i to the CF is represented by the mapping

$$\varphi_{\text{in}}^i : [2^{nR_i}] \to [2^{nC_{\text{in}}^i}] \tag{4.1}$$

and the transmission from the CF to encoder i is represented by

$$\varphi_{\text{out}}^i : [2^{nC_{\text{in}}^1}] \times [2^{nC_{\text{in}}^2}] \to [2^{nC_{\text{out}}^i}].$$

For simplicity, the transmissions to and from the CF occur prior to the transmission of codewords over the channel.

At time $t \in [n]$, for $i \in \{1,2\}$, the transmission of encoder i over the channel is represented by the mapping

$$f_{it} : [2^{nR_i}] \times [2^{nC_{\text{out}}^i}] \times \hat{\mathcal{S}}_i^t \to \mathcal{X}_i. \tag{4.2}$$

Here $\hat{S}_i^t$ represents any knowledge about the state gathered by encoder i in times $\{1, \ldots, t\}$. Let $*$ be a symbol not in $\mathcal{S}_1 \cup \mathcal{S}_2$. For $t \in [n]$, we have

$$\hat{S}_{it} = \begin{cases} * & \text{no state information} \\ S_{i(t-1)} & \text{strictly causal} \\ S_{it} & \text{causal} \\ S_i^n & \text{non-causal.} \end{cases}$$

For every message pair (w_1, w_2), the codeword of encoder i is required to satisfy the cost constraint

$$\sum_{t=1}^n \mathbb{E} b_i \Big[f_{it}\big(w_i, \varphi_{\text{out}}^i(\varphi_{\text{in}}^1(w_1), \varphi_{\text{in}}^2(w_2)), \hat{S}_i^t\big) \Big] \leq B_i. \tag{4.3}$$

The decoder has full state information and is represented by the mapping

$$g : \mathcal{S}^n \times \mathcal{Y}^n \to [2^{nR_1}] \times [2^{nR_2}].$$

The average probability of error is given by

$$P_e^{(n)} = \Pr\Big\{ g(S^n, Y^n) \neq (W_1, W_2) \Big\},$$

Table 4.1: Parameter τ designates the type of state information available at the encoders.

τ	encoder state information
0	none
$T-1$	strictly causal
T	causal
∞	non-causal

where (W_1, W_2) is uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}]$. A rate pair (R_1, R_2) is achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$-codes with $P_e^{(n)} \to 0$ as $n \to \infty$. We use subscript $\tau \in \{0, T-1, T, \infty\}$ to specify the dependence of the capacity region and sum-capacity on the availability of state information at the encoders. Table 4.1 makes this dependence clear. The capacity region $\mathscr{C}_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is given by the closure of all achievable rate pairs. The sum-capacity, denoted by $C_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, is defined as

$$C_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) := \max_{\mathscr{C}_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})} (R_1 + R_2). \tag{4.4}$$

For example, $\mathscr{C}_T(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ and $C_T(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ denote the capacity region and sum-capacity of a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF and distributed causal state information available at the encoders, respectively.

4.2.3 Message and State Cooperation

In the scenario where non-causal state information is available at the encoders, we also study the benefit of joint message and state cooperation. In the definition of a code for the case where non-causal state information is available at the encoders (Subsection 4.2.2), for $i \in \{1, 2\}$, replace (4.1) and (4.3) with

$$\varphi_{\text{in}}^i : [2^{nR_i}] \times \mathcal{S}_i^n \to [2^{nC_{\text{in}}^i}], \text{ and}$$

$$\sum_{t=1}^n \mathbb{E} b_i \Big[f_{it}\big(w_i, \varphi_{\text{out}}^i(\varphi_{\text{in}}^1(w_1, S_1^n), \varphi_{\text{in}}^2(w_2, S_2^n)), S_i^n\big)\Big] \leq B_i.$$

We denote the capacity region and sum-capacity with $\mathscr{C}_{\infty,s}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ and $C_{\infty,s}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, respectively. The subscript "s" indicates the dependence of the cooperation strategy on the channel state information.

4.3 Coding Strategy

Here we describe our coding strategies, which are based on random coding arguments. Since our aim is to determine conditions sufficient for an infinite slope cooperation gain, we specifically focus on coding strategies that lead to large gains for small cooperation rates such as the coordination strategy (Chapter 3). In particular, in the coding strategies below, the CF does not use its rate for forwarding message or state information, since in the cases studied in the literature [5], [11], the gain of such a strategy is at most linear in the cooperation rate. We start with message cooperation and conclude with message and state cooperation.

4.3.1 Inner Bound for Message Cooperation

For simplicity, we assume the CF has access to both messages by setting $\mathbf{C}_{\text{in}} = \mathbf{C}^*_{\text{in}} = (C^{*1}_{\text{in}}, C^{*2}_{\text{in}})$, where C^{*1}_{in} and C^{*2}_{in} are sufficiently large. Despite this assumption, our main result regarding sum-capacity gain, Theorem 4.4.1, holds for any $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$. This is due to the fact that using time-sharing, as stated in the lemma below, we can use the inner bounds for $\mathbf{C}^*_{\text{in}}$ to obtain inner bounds for any $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$.

Lemma 4.3.1. *Fix a memoryless stationary MAC. For any* $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{\geq 0}$*, there exists* $\mu > 0$*, depending only on* $\mathbf{C}_{\text{in}}$*, such that for all* $\tau \in \{0, T-1, T, \infty\}$,

$$C_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_\tau(\mathbf{C}_{\text{in}}, \mathbf{0}) \geq \mu\Big(C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) - C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{0})\Big).$$

The proof appears in Subsection 4.6.1.

We first describe our inner bound for the case where the encoders do not have access to state information. In this case, even though the decoder has access to full state information, we can obtain a suitable inner bound by applying results where state information is absent at both the encoders and the decoder to a modified channel. Specifically, applying Corollary 3.4.1 to the channel

$$\Big(\mathcal{X}_1 \times \mathcal{X}_2, p(y, s|x_1, x_2), \mathcal{Y} \times \mathcal{S}\Big),$$

where

$$p(y, s|x_1, x_2) = p(s)p(y|s, x_1, x_2),$$

gives an inner bound for the channel $p(y|s, x_1, x_2)$ when full state information is available at the decoder.

Lemma 4.3.2. *The set of all rate pairs* (R_1, R_2) *satisfying*

$$\begin{aligned} R_1 &\leq I(X_1; Y|S_1, S_2, X_2) \\ R_2 &\leq I(X_2; Y|S_1, S_2, X_1) \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S_1, S_2) \end{aligned}$$

for some distribution $p(x_1)p(x_2)$ *with*

$$I(X_1; X_2) \leq C^1_{\text{out}} + C^2_{\text{out}}$$

and $\mathbb{E}[b_i(X_i)] \leq B_i$ *for* $i \in \{1, 2\}$*, is contained in* $\mathscr{C}_0(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}})$.

In the case where the encoders have access to causal state information, the codeword transmitted by an encoder can depend both on the encoder's message and the present state information available at the encoder. Lemma 4.3.3 provides an inner bound for the capacity region in this scenario. In the inner bound, for $i \in \{1, 2\}$, U_i encodes the message of encoder i in addition to the information it receives from the CF. The proof is given in Appendix 4.6.2.

Lemma 4.3.3. *The set of all rate pairs satisfying*

$$\begin{aligned} R_1 &\leq I(U_1; Y|S_1, S_2, U_2) \\ R_2 &\leq I(U_2; Y|S_1, S_2, U_1) \\ R_1 + R_2 &\leq I(U_1, U_2; Y|S_1, S_2) \end{aligned}$$

for some distribution $p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2)$ *with*

$$I(U_1; U_2) \leq C^1_{\text{out}} + C^2_{\text{out}}$$

and $\mathbb{E}[b_i(X_i)] \leq B_i$ *for* $i \in \{1, 2\}$*, is contained in* $\mathscr{C}_T(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}})$.

4.3.2 Inner Bound for Message and State Cooperation

As discussed in Subsection 4.2.3, we only consider message and state cooperation in the scenario where non-causal state information is available at the encoders.

Here we assume that the state alphabet $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ is discrete and $H(S_1, S_2)$ is finite. Furthermore, we assume the CF not only has access to both messages, but also knows the state sequences S_1^n and S_2^n; equivalently, we set $\mathbf{C}_{\text{in}} = \bar{\mathbf{C}}_{\text{in}} = (\bar{C}^1_{\text{in}}, \bar{C}^2_{\text{in}})$, where $\bar{C}^1_{\text{in}}$ and $\bar{C}^2_{\text{in}}$ are sufficiently large. A lemma similar to Lemma 4.3.1 holds in this case.

Lemma 4.3.4. *Fix a memoryless stationary MAC. For any* $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{\geq 0}$*, there exists* $\mu > 0$*, depending only on* $\mathbf{C}_{\text{in}}$*, such that*

$$C_{(\infty,s)}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{(\infty,s)}(\mathbf{C}_{\text{in}}, \mathbf{0}) \geq \mu\Big(C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{0})\Big),$$

Codebook Generation. Choose a distribution $p(x_1, x_2|s_1, s_2)$. For $i \in \{1, 2\}$, $w_i \in [2^{nR_i}]$, $z_i \in [2^{nC^i_{\text{out}}}]$, $s_i^n \in \mathcal{S}_i^n$, generate $X_i^n(w_i, z_i|s_i^n)$ i.i.d. according to the distribution

$$\Pr\left\{X_i^n(w_i, z_i|s_i^n) = x_i^n \middle| S_i^n = s_i^n\right\} = \prod_{t=1}^n p(x_{it}|s_{it}).$$

Encoding. The CF, having access to (w_1, w_2) and (S_1^n, S_2^n), looks for a pair $(Z_1, Z_2) \in [2^{nC^1_{\text{out}}}] \times [2^{nC^2_{\text{out}}}]$ satisfying

$$\big(S_1^n, S_2^n, X_1^n(w_1, Z_1|S_1^n), X_2^n(w_2, Z_2|S_2^n)\big) \in A_\delta^{(n)}, \tag{4.5}$$

where $A_\delta^{(n)}$ is the weakly typical set with respect to the distribution

$$p(s_1, s_2)p(x_1, x_2|s_1, s_2).$$

If there is more than one such pair, the CF chooses the smallest pair according to the lexicographical order. If there is no such pair, it sets $(Z_1, Z_2) = (1, 1)$. The CF sends Z_i to encoder i for $i \in \{1, 2\}$. Encoder i transmits $X_i^n(w_i, Z_i|S_i^n)$ over n uses of the channel.

By the multivariate covering lemma (Appendix A), the probability that a pair (Z_1, Z_2) satisfying (4.5) exists goes to one as n goes to infinity provided that

$$\begin{aligned} C^1_{\text{out}} &> H(X_1|S_1) - H(X_1|S_1, S_2) + 24\delta \\ C^2_{\text{out}} &> H(X_2|S_2) - H(X_2|S_1, S_2) + 24\delta \\ C^1_{\text{out}} + C^2_{\text{out}} &> H(X_1|S_1) + H(X_2|S_2) - H(X_1, X_2, S_1, S_2) + 4\delta. \end{aligned}$$

Decoding. Once the decoder receives Y^n, using (S_1^n, S_2^n), it looks for a pair $(\hat{w}_1, \hat{w}_2)$ that satisfies

$$\big(S_1^n, S_2^n, X_1^n(\hat{w}_1, \hat{Z}_1|S_1^n), X_2^n(\hat{w}_2, \hat{Z}_2|S_2^n), Y^n\big) \in A_\epsilon^{(n)}.$$

Here $A_\epsilon^{(n)}$ is the weakly typical set with respect to the distribution

$$p(s_1, s_2)p(x_1, x_2|s_1, s_2)p(y|s_1, s_2, x_1, x_2).$$

If there is no such pair, or there is such a pair but it is not unique, the decoder sets $(\hat{w}_1, \hat{w}_2) = (1, 1)$.

The error analysis of the above coding scheme leads to the following lemma, which provides an inner bound for $\mathscr{C}_{\infty,s}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{C}_{\text{out}})$.

Lemma 4.3.5. *The set of all rate pairs satisfying*

$$\begin{aligned} R_1 &\leq I(X_1; Y|S_1, S_2, X_2) \\ R_2 &\leq I(X_2; Y|S_1, S_2, X_1) \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S_1, S_2) \end{aligned}$$

for some distribution $p(x_1, x_2|s_1, s_2)$ *with*

$$\begin{aligned} C^1_{\text{out}} &\geq I(X_1; S_2|S_1) \\ C^2_{\text{out}} &\geq I(X_2; S_1|S_2) \\ C^1_{\text{out}} + C^2_{\text{out}} &\geq I(X_1; S_2|S_1) + I(X_2; S_1|S_2) + I(X_1; X_2|S_1, S_2) \end{aligned}$$

and $\mathbb{E}[b_i(X_i)] \leq B_i$ *for* $i \in \{1, 2\}$*, is contained in* $\mathscr{C}_{\infty,s}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{C}_{\text{out}})$.

4.4 Main Result

Our main result describes conditions on a MAC that, if satisfied, guarantee for every fixed $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$, an infinite slope in sum-capacity as a function of $\mathbf{C}_{\text{out}}$. As sum-capacity depends on the availability of state information at the encoders, so do our conditions. The proof appears in Subsection 4.6.3.

Theorem 4.4.1. *Let* $\mathcal{S}$*,* $\mathcal{X}_1$*,* $\mathcal{X}_2$*, and* $\mathcal{Y}$ *be finite sets. For any* $\tau \in \{0, T-1, T, \infty, (\infty, s)\}$*, any MAC in* $\mathcal{C}_\tau(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$*, and any* $(\mathbf{C}_{\text{in}}, \mathbf{v}) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{>0}$*,*

$$\lim_{h \to 0^+} \frac{C_\tau(\mathbf{C}_{\text{in}}, h\mathbf{v}) - C_\tau(\mathbf{C}_{\text{in}}, \mathbf{0})}{h} = \infty.$$

We next specifically define $\mathcal{C}_\tau(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ for each subscript $\tau \in \{0, T-1, T, \infty, (\infty, s)\}$; as defined previously, τ specifies the availability of state information at the encoders. Note that the definition of C_τ provides a sufficient condition for a large cooperation gain; the given condition may not be necessary.

In our descriptions below, all mentioned distributions satisfy

$$\mathbb{E}b_i(X_i) \leq B_i \text{ for } i \in \{1, 2\}.$$

No state information. A MAC is in $\mathcal{C}_0(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ if

(i) for some $p_0(x_1)p_0(x_2)$ that satisfies

$$I_0(X_1, X_2; Y|S) = \max_{p(x_1)p(x_2)} I(X_1, X_2; Y|S),$$

there exists $p_1(x_1, x_2)$ that satisfies

$$\begin{aligned} &I_1(X_1, X_2; Y|S) + \mathbb{E}\Big[D\big(p_1(y|S) \| p_0(y|S)\big)\Big] \\ &\qquad > I_0(X_1, X_2; Y|S), \text{ and} \end{aligned}$$

(ii) $\text{supp}(p_1(x_1, x_2)) \subseteq \text{supp}(p_0(x_1, x_2))$, where "supp" denotes the support.

Intuitively, condition (i) ensures that our channel has the property that dependence created through message cooperation increases sum-capacity. Condition (ii) allows the CF to use a small rate (i.e., small $\mathbf{C}_{\text{out}}$) to help the encoders, whose codewords are generated according to $p_0(x_1)p_0(x_2)$, to transmit codewords whose distribution is sufficiently close to $p_1(x_1, x_2)$ to achieve a large gain in sum-capacity.

Strictly causal state information. The availability of strictly causal state information at the encoders of a MAC without cooperation does not enlarge the capacity region, thus we set $\mathcal{C}_{T-1}(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}) = \mathcal{C}_0(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$.

Causal state information. A MAC is in $\mathcal{C}_T(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ if

(i) for some $p_0(x_1|s_1)p_0(x_2|s_2)$ that satisfies

$$I_0(X_1, X_2; Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1, X_2; Y|S),$$

there exist alphabets $\mathcal{U}_1$, $\mathcal{U}_2$, distributions $p_0(u_1)p_0(u_2)$ and $p_1(u_1, u_2)$, and conditional distributions $p_*(x_1|u_1, s_1)$ and $p_*(x_2|u_2, s_2)$ such that

$$\begin{aligned} &p_0(x_1|s_1)p_0(x_2|s_2) = \\ &\qquad \sum_{u_1, u_2} p_0(u_1)p_0(u_2)p_*(x_1|u_1, s_1)p_*(x_2|u_2, s_2), \\ &I_1(U_1, U_2; Y|S) + \mathbb{E}\Big[D\big(p_1(y|S) \| p_0(y|S)\big)\Big] \\ &\qquad > I_0(U_1, U_2; Y|S), \end{aligned} \tag{4.6}$$

(ii) $\text{supp}(p_1(u_1, u_2)) \subseteq \text{supp}(p_0(u_1)p_0(u_2))$.

In (4.6), the expressions are calculated with respect to the input distributions

$$\sum_{u_1,u_2} p_0(u_1)p_0(u_2)p_*(x_1|u_1,s_1)p_*(x_2|u_2,s_2), \text{ and}$$

$$\sum_{u_1,u_2} p_1(u_1,u_2)p_*(x_1|u_1,s_1)p_*(x_2|u_2,s_2).$$

Non-causal state information (message cooperation). In the case with no cooperation ($\mathbf{C}_{\text{out}} = \mathbf{0}$), the capacity region is not dependent on whether the state information available at the encoders is causal or non-causal. This follows from the converse argument in [11] and relies on the fact that S^n is an i.i.d. sequence. Thus, similar to the strictly causal case above, we set $\mathcal{C}_\infty(\mathcal{S},\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}) = \mathcal{C}_T(\mathcal{S},\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y})$.

Non-causal state information (message and state cooperation). Finally, we say a MAC is in $\mathcal{C}_{\infty,s}(\mathcal{S},\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y})$ if there exists

(i) $p_0(x_1|s_1)p_0(x_2|s_2)$ that satisfies

$$I_0(X_1,X_2;Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1,X_2;Y|S),$$

(ii) $p_1(x_1,x_2|s_1,s_2)$ that satisfies

$$\begin{aligned} I_1(X_1,X_2;Y|S) + \mathbb{E}\Big[D\big(p_1(y|S)\|p_0(y|S)\big)\Big] \\ > I_0(X_1,X_2;Y|S), \text{ and} \end{aligned}$$

(iii) for all $s \in \mathcal{S}$, $\text{supp}(p_1(\cdot|s)) \subseteq \text{supp}(p_0(\cdot|s))$.

4.5 Example: Gaussian MAC with Binary Fading

While we prove Theorem 4.4.1 only for finite alphabet MACs, the result is not limited to such MACs. Specifically, for a given MAC, we can use our inner bounds described in Section 4.3 to calculate an inner bound for sum-capacity and verify the result of Theorem 4.4.1 directly. We next describe an example of such a MAC.

Consider a MAC that models the wireless communication between two transmitters and a receiver in the presence of binary fading. The input-output relationship of our MAC is given by

$$Y = S_1X_1 + S_2X_2 + Z,$$

where S_1 and S_2 are independent Bernoulli(1/2) random variables, and Z is a Gaussian random variable with mean zero and variance N. In addition, we set the cost functions $b_1(x) = b_2(x) = x^2$ and cost constraints $B_i = P_i$ for $i \in \{1,2\}$, so that the cost constraints correspond to the usual power constraints of the Gaussian MAC.

Proposition 4.5.1. *Consider the Gaussian MAC with binary fading. Fix $(\mathbf{C}_{\text{in}}, \mathbf{v}) \in \mathbb{R}^2_{>0} \times \mathbb{R}^2_{>0}$. Then for all $\tau \in \{0, T-1, T, \infty, (\infty, s)\}$,*

$$\lim_{h \to 0^+} \frac{C_\tau(\mathbf{C}_{\text{in}}, h\mathbf{v}) - C_\tau(\mathbf{C}_{\text{in}}, \mathbf{0})}{h} = \infty.$$

The proof is given in Appendix 4.6.4.

4.6 Proofs

4.6.1 Proof of Lemma 4.3.1

Since $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$, there exists $\mu \in (0,1)$ such that for $i \in \{1,2\}$,

$$\mathbf{C}^i_{\text{in}} \geq \mu \mathbf{C}^{*i}_{\text{in}}.$$

Then for each $\tau \in \{0, T-1, T, \infty\}$, a time-sharing argument shows that

$$\begin{aligned}\mathscr{C}_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) &\supseteq \mu \mathscr{C}_\tau(\mathbf{C}_{\text{in}}/\mu, \mathbf{C}_{\text{out}}) + (1-\mu)\mathscr{C}_\tau(\mathbf{0}, \mathbf{C}_{\text{out}}) \\ &\supseteq \mu \mathscr{C}_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) + (1-\mu)\mathscr{C}_\tau(\mathbf{0}, \mathbf{C}_{\text{out}}).\end{aligned}$$

Thus

$$C_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \geq \mu C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) + (1-\mu) C_\tau(\mathbf{0}, \mathbf{C}_{\text{out}}),$$

which implies

$$C_\tau(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_\tau(\mathbf{C}_{\text{in}}, \mathbf{0}) \geq \mu\Big(C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) - C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{0})\Big)$$

since

$$C_\tau(\mathbf{0}, \mathbf{C}_{\text{out}}) = C_\tau(\mathbf{0}, \mathbf{0}) = C_\tau(\mathbf{C}_{\text{in}}, \mathbf{0}) = C_\tau(\mathbf{C}^*_{\text{in}}, \mathbf{0}).$$

4.6.2 Proof of Lemma 4.3.3

Fix alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$, and mappings

$$f_i : \mathcal{U}_i \times \mathcal{S}_i \to \mathcal{X}_i \text{ for } i \in \{1,2\}.$$

Applying Lemma 4.3.2, where state information is only available at the decoder, to the channel

$$p(y|s, u_1, u_2) = \sum_{x_1, x_2} p(y|s, x_1, x_2) \mathbf{1}\big\{x_1 = f_1(u_1, s_1)\big\} \mathbf{1}\big\{x_2 = f_2(u_2, s_2)\big\} \quad (4.7)$$

shows that the set of all rate pairs satisfying

$$R_1 \le I(U_1; Y|S, U_2)$$
$$R_2 \le I(U_2; Y|S, U_1)$$
$$R_1 + R_2 \le I(U_1, U_2; Y|S)$$

for some distribution $p(u_1, u_2)$ with

$$I(U_1; U_2) \le C^1_{\text{out}} + C^2_{\text{out}},$$

is achievable for the channel $p(y|s, u_1, u_2)$ when no state information is available at the encoders. Note that every code for this channel can be transformed into a code for the channel $p(y|s, x_1, x_2)$ with causal state information available at the encoders; for all times $t \in [n]$ and $i \in \{1, 2\}$, simply apply the mapping f_i to the pair (U_{it}, S_{it}), where U_{it} is the output symbol of encoder i and S_{it} is component i of the state at time t. Note that the new code has the same rate and by (4.7), the same average error probability as the original code. Thus $\mathscr{C}_T(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}})$ contains the set of all rate pairs (R_1, R_2) satisfying

$$R_1 \le I(U_1; Y|S, U_2)$$
$$R_2 \le I(U_2; Y|S, U_1)$$
$$R_1 + R_2 \le I(U_1, U_2; Y|S)$$

for some distribution $p(u_1, u_2)$ with

$$I(U_1; U_2) \le C^1_{\text{out}} + C^2_{\text{out}}$$

and mappings

$$f_i : \mathcal{U}_i \times \mathcal{S}_i \to \mathcal{X}_i \quad \text{for } i \in \{1, 2\}.$$

To complete the proof, we show that for every $\delta \ge 0$ and every distribution

$$p(u_1, u_2)p(s_1, s_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2)$$

satisfying $I(U_1; U_2) \le \delta$, there exist alphabets $\mathcal{U}'_1$ and $\mathcal{U}'_2$, mappings

$$f_i : \mathcal{U}'_i \times \mathcal{S}_i \to \mathcal{X}_i \quad \text{for } i \in \{1, 2\},$$

and distribution $p(u'_1, u'_2)$ such that $I(U'_1; U'_2) = I(U_1, U_2)$, and the rate region calculated with respect to

$$p(u'_1, u'_2)\mathbf{1}\big\{x_1 = f_1(u'_1, s_1)\big\}\mathbf{1}\big\{x_2 = f_2(u'_2, s_2)\big\},$$

contains the region calculated with respect to $p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2)$.

To this end, applying Lemma 4.6.1 (see end of this appendix) to $p(x_i|u_i, s_i)$ demonstrates the existence of a random variable V_i that is independent of (U_i, S_i) and a mapping

$$f_i : \mathcal{V}_i \times \mathcal{U}_i \times \mathcal{S}_i \to \mathcal{X}_i$$

that satisfies

$$p(x_i|u_i, s_i) = \sum_{v_i} p(v_i)\mathbf{1}\big\{x_i = f_i(v_i, u_i, s_i)\big\}.$$

Furthermore, without loss of generality, we may assume V_1 and V_2 are independent, and (V_1, V_2) is independent of (U_1, U_2, S_1, S_2).

Let $U_i' := (U_i, V_i)$ for $i \in \{1, 2\}$. Then

$$\begin{aligned} I(U_1'; U_2') &= I(U_1, V_1; U_2, V_2) \\ &= H(U_1, V_1) + H(U_2, V_2) - H(U_1, U_2, V_1, V_2) \\ &= I(U_1; U_2) + I(V_1; V_2) = I(U_1; U_2). \end{aligned}$$

We next show that the rate region calculated with respect to the distribution

$$p(s_1, s_2)p(u_1', u_2')\mathbf{1}\{x_1 = f_1(u_1', s_1)\}\mathbf{1}\{x_2 = f_2(u_2', s_2)\}$$

contains the rate region with respect to

$$p(s_1, s_2)p(u_1, u_2)p(x_1|u_1, s_1)p(x_2|u_2, s_2).$$

Recall that $S = (S_1, S_2)$. We have

$$\begin{aligned} I(U_1'; Y|S, U_2') &= I(U_1'; Y, U_2'|S) - I(U_1'; U_2'|S) \\ &= I(U_1'; Y, U_2'|S) - I(U_1; U_2|S) \\ &= I(U_1, V_1; Y, U_2, V_2|S) - I(U_1; U_2|S) \\ &\geq I(U_1; Y, U_2|S) - I(U_1; U_2|S) \\ &= I(U_1; Y|S, U_2). \end{aligned}$$

Similarly, we show

$$I(U_2'; Y|S, U_1') \geq I(U_2; Y|S, U_1).$$

Finally, we have

$$\begin{aligned} I(U_1', U_2'; Y|S) &= I(U_1, V_1, U_2, V_2; Y|S) \\ &\geq I(U_1, U_2; Y|S). \end{aligned}$$

This completes the proof. We next state and prove Lemma 4.6.1, which was used earlier. In Lemma 4.6.1, the scenario where $\mathcal{X}$ and $\mathcal{S}$ are finite is a special case of the functional representation lemma [38, p. 626].

Lemma 4.6.1. *Let $\{F(\cdot|s)\}_{s\in\mathcal{S}}$ be a collection of cumulative distribution functions (CDFs) on alphabet $\mathcal{X} \subseteq \mathbb{R}$ and let S be a random variable with alphabet $\mathcal{S}$. Then there exists a random variable U independent of S and a mapping*

$$g : \mathcal{S} \times \mathcal{U} \to \mathcal{X}$$

such that the conditional CDF of $g(S,U)$ given $S = s$ equals $F(\cdot|s)$. In the case where $\mathcal{X}$ and $\mathcal{S}$ are finite, we can choose $\mathcal{U}$ such that

$$|\mathcal{U}| \leq |\mathcal{S}|\big(|\mathcal{X}| - 1\big) + 1. \tag{4.8}$$

Proof. We prove the result for general alphabets $\mathcal{X} \subseteq \mathbb{R}$. Let $\mathcal{U} = [0,1]$. Define the mapping $g : \mathcal{S} \times \mathcal{U} \to \mathcal{X}$ as

$$g(s,u) = \inf\left\{x \in \mathcal{X} \middle| F(x|s) \geq u\right\}.$$

Let U be independent of S and uniformly distributed on $(0,1)$. From the quantile function theorem [54, Theorem 2], it follows that for all $s \in \mathcal{S}$, $g(s,U)$ has CDF $F(\cdot|s)$. Set $X = g(S,U)$. Then

$$\begin{aligned} F_{X|S}(x|s) &= \Pr\left\{X \leq x \middle| S = s\right\} \\ &= \Pr\left\{g(S,U) \leq x \middle| S = s\right\} \\ &= \Pr\left\{g(s,U) \leq x\right\} = F(x|s). \end{aligned}$$

□

4.6.3 Proof of Theorem 4.4.1

From the description of the set $\mathcal{C}_\tau(\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y})$ in Section 4.4, we see that it suffices to prove Theorem 4.4.1 only in the cases $\tau = 0$, $\tau = T$, and $\tau = (\infty, s)$.

The case $\tau = 0$. When no state information is available at the encoders, Theorem 4.4.1 follows immediately by applying Theorem 3.2.3 to the MAC

$$p(s, y|x_1, x_2) = p(s)p(y|s, x_1, x_2),$$

with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and output alphabet $\mathcal{S} \times \mathcal{Y}$.

The case $\tau = T$. When causal state information is available at the encoders, Theorem 4.4.1 follows by applying the case $\tau = 0$ to the MAC

$$p(y|s, u_1, u_2) = \sum_{x_1, x_2} p_*(x_1|u_1, s_1)p_*(x_2|u_2, s_2)p(y|s, x_1, x_2),$$

with input alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$, state alphabet $\mathcal{S}$, and output alphabet $\mathcal{Y}$.

The case $\tau = (\infty, s)$. In this case, we provide a self-contained proof as it is not straightforward to derive it from prior cases. This is due to the fact that in this case, as described in Lemma 4.3.5, the family of achievable distributions is constrained by three inequalities rather than one.

Let $p_0(x_1|s_1)p_0(x_2|s_2)$ be a distribution that satisfies

$$I_0(X_1, X_2; Y|S) = \max_{p(x_1|s_1)p(x_2|s_2)} I(X_1, X_2; Y|S).$$

By assumption, there exists a distribution $p_1(x_1, x_2|s_1, s_2)$ such that

$$\begin{aligned} I_1(X_1, X_2; Y|S) + \mathbb{E}\Big[D\big(p_1(y|S)\|p_0(y|S)\big)\Big] \\ > I_0(X_1, X_2; Y|S), \end{aligned} \tag{4.9}$$

and for all $s \in \mathcal{S}$,

$$\mathrm{supp}(p_1(\cdot|s)) \subseteq \mathrm{supp}(p_0(\cdot|s)). \tag{4.10}$$

For every $\lambda \in [0, 1]$, define

$$p_\lambda(x_1, x_2|s_1, s_2) = (1 - \lambda)p_0(x_1|s_1)p_0(x_2|s_2) + \lambda p_1(x_1, x_2|s_1, s_2).$$

Fix $\epsilon > 0$ and $\mathbf{v} \in \mathbb{R}^2_{>0}$. Define the mapping $h : [0, 1] \to \mathbb{R}$ as

$$h(\lambda) = \frac{1}{v_1} I_\lambda(X_1; S_2|S_1) + \frac{1}{v_2} I_\lambda(X_2; S_1|S_2) + \frac{1}{v_1 + v_2} I_\lambda(X_1; X_2|S_1, S_2).$$

Using (4.10), a direct calculation shows

$$\begin{aligned} \frac{d}{d\lambda} I_\lambda(X_1; S_2|S_1)\Big|_{\lambda=0^+} &= 0 \\ \frac{d}{d\lambda} I_\lambda(X_2; S_1|S_2)\Big|_{\lambda=0^+} &= 0 \\ \frac{d}{d\lambda} I_\lambda(X_1; X_2|S_1, S_2)\Big|_{\lambda=0^+} &= 0. \end{aligned}$$

Since h is continuously differentiable and

$$\frac{dh}{d\lambda}\Big|_{\lambda=0^+} = \epsilon > 0,$$

by the inverse function theorem, there exists $h_0 > 0$ such that h is invertible on $[0, h_0)$; that is, there exists a mapping $\lambda^* : [0, h_0) \to [0,1]$ that satisfies

$$h = \frac{1}{v_1} I_{\lambda^*(h)}(X_1; S_2|S_1) + \frac{1}{v_2} I_{\lambda^*(h)}(X_2; S_1|S_2) + \frac{1}{v_1+v_2} I_{\lambda^*(h)}(X_1; X_2|S_1, S_2), \tag{4.11}$$

and

$$\frac{d\lambda^*}{dh}\bigg|_{h=0^+} = \frac{1}{\epsilon}.$$

We henceforth write λ^* instead of $\lambda^*(h)$ when the value of h is clear from context.

From (4.11), it follows that for all $h \in [0, h_0)$,

$$\begin{aligned} hv_1 &\geq I_{\lambda^*}(X_1; S_2|S_1) \\ hv_2 &\geq I_{\lambda^*}(X_2; S_1|S_2) \\ h(v_1+v_2) &\geq I_{\lambda^*}(X_1; S_2|S_1) + I_{\lambda^*}(X_2; S_1|S_2) + I_{\lambda^*}(X_1; X_2|S_1, S_2). \end{aligned}$$

Thus, by Lemma 4.3.5,

$$C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, h\mathbf{v}) \geq I_{\lambda^*}(X_1, X_2; Y|S) - I_{\lambda^*}(X_1; X_2|S). \tag{4.12}$$

Since equality holds in (4.12) at $h = 0$, we have

$$\begin{aligned} &\liminf_{h\to 0^+} \frac{C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, h\mathbf{v}) - C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{0})}{h} && (4.13) \\ &\quad \geq \frac{1}{\epsilon}\frac{d}{d\lambda^*}\Big(I_{\lambda^*}(X_1, X_2; Y|S) - I_{\lambda^*}(X_1; X_2|S)\Big)\Big|_{\lambda^*=0^+} \\ &\quad = \frac{1}{\epsilon}\frac{d}{d\lambda^*} I_{\lambda^*}(X_1, X_2; Y|S)\Big|_{\lambda^*=0^+} \\ &\quad \geq \frac{1}{\epsilon}\Big(I_1(X_1, X_2; Y|S) + \mathbb{E}\Big[D\big(p_1(y|S)\|p_0(y|S)\big)\Big] - I_0(X_1, X_2; Y|S)\Big). && (4.14) \end{aligned}$$

The proof of (4.14) is analogous to Lemma 3.5.6 (ii) and is omitted. Since (4.14) holds for all $\epsilon > 0$, from (4.9) it follows that

$$\lim_{h\to 0^+} \frac{C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, h\mathbf{v}) - C_{(\infty,s)}(\bar{\mathbf{C}}_{\text{in}}, \mathbf{0})}{h} = \infty.$$

4.6.4 Proof of Proposition 4.5.1

Since $C_0(\mathbf{C}^*_{\text{in}}, \mathbf{0}) = C_{T-1}(\mathbf{C}^*_{\text{in}}, \mathbf{0})$ and $C_T(\mathbf{C}^*_{\text{in}}, \mathbf{0}) = C_\infty(\mathbf{C}^*_{\text{in}}, \mathbf{0}) = C_{(\infty,s)}(\mathbf{C}^*_{\text{in}}, \mathbf{0})$, it suffices to prove the result only when $\tau = 0$ or $\tau = T$.

When $\tau = 0$, from Lemma 4.3.2, it follows that any distribution $p(x_1)p(x_2)$ satisfying $\mathbb{E}[X_i^2] \leq P_i$ for $i \in \{1, 2\}$ and

$$I(X_1; X_2) \leq C_{\text{out}}^1 + C_{\text{out}}^2,$$

we have

$$C_0(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) \geq I(X_1, X_2; Y|S) - I(X_1; X_2). \tag{4.15}$$

Fix $h > 0$. Let (X_1, X_2) be jointly Gaussian with mean zero and covariance matrix

$$\Sigma := \begin{pmatrix} \sqrt{P_1} & \rho\sqrt{P_1 P_2} \\ \rho\sqrt{P_1 P_2} & \sqrt{P_2} \end{pmatrix},$$

where $\rho \in [0, 1]$ is chosen such that

$$I(X_1; X_2) = \frac{1}{2} \log \frac{1}{1 - \rho^2} := h(v_1 + v_2).$$

Then

$$\left.\frac{d\rho}{dh}\right|_{h=0^+} = \infty.$$

Using (4.15), it follows that

$$C_0(\mathbf{C}_{\text{in}}^*, h\mathbf{v}) - C_0(\mathbf{C}_{\text{in}}^*, \mathbf{0}) \geq \frac{1}{8} \log\left(1 + \frac{2\rho\sqrt{P_1 P_2}}{1 + P_1 + P_2 + N}\right) - h(v_1 + v_2),$$

from which the desired result follows.

A similar proof follows when $\tau = T$. In this case, for fixed $h > 0$, let (U_1, U_2) be jointly Gaussian with mean zero and covariance matrix

$$\Sigma := \begin{pmatrix} \sqrt{2P_1} & 2\rho\sqrt{P_1 P_2} \\ 2\rho\sqrt{P_1 P_2} & \sqrt{2P_2} \end{pmatrix},$$

where $\rho \in [0, 1]$ is chosen such that

$$I(U_1; U_2) = \frac{1}{2} \log \frac{1}{1 - \rho^2} := h(v_1 + v_2).$$

Now set $X_i := S_i U_i$ and apply Lemma 4.3.3.

4.6.5 Outer Bounds in the Absence of Cooperation

We next prove outer bounds for $\mathscr{C}_{T-1}(\mathbf{0}, \mathbf{0})$ and $\mathscr{C}_\infty(\mathbf{0}, \mathbf{0})$. Together with our inner bounds in Section 4.3, these outer bounds determine the capacity region $\mathscr{C}_\tau(\mathbf{0}, \mathbf{0})$ for all τ, and show

$$\mathscr{C}_0(\mathbf{0}, \mathbf{0}) = \mathscr{C}_{T-1}(\mathbf{0}, \mathbf{0}) \text{ and } \mathscr{C}_T(\mathbf{0}, \mathbf{0}) = \mathscr{C}_\infty(\mathbf{0}, \mathbf{0}) = \mathscr{C}_{(\infty,s)}(\mathbf{0}, \mathbf{0}).$$

The bounds presented here are well known [38, p. 175] and are included for completeness.

For convergent sequences $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$, define notation $\simeq$ and $\lesssim$ as

$$\begin{aligned} a_n \simeq b_n &:\iff \lim_{n\to\infty} \frac{1}{n}(a_n - b_n) = 0 \\ a_n \lesssim b_n &:\iff \lim_{n\to\infty} \frac{1}{n}(a_n - b_n) \leq 0. \end{aligned}$$

Consider a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$ for the MAC with full state information at the decoder. Initially, we do not make any assumptions regarding the presence of state information at the encoders.

We begin with the bound on R_1. We have

$$\begin{aligned} nR_1 &= H(W_1) \\ &= H(W_1|S^n, W_2) \\ &\simeq I(W_1; Y^n|S^n, W_2) \\ &= H(Y^n|S^n, W_2, X_2^n) - H(Y^n|S^n, W_1, W_2, X_1^n, X_2^n) \\ &= H(Y^n|S^n, X_2^n) - H(Y^n|S^n, X_1^n, X_2^n) \\ &= \sum_{t=1}^{n} \Big(H(Y_t|Y^{t-1}, S^n, X_2^n) - H(Y_t|Y^{t-1}, S^n, X_1^n, X_2^n) \Big). \end{aligned} \tag{4.16}$$

Similarly,

$$nR_2 \simeq \sum_{t=1}^{n} \Big(H(Y_t|Y^{t-1}, S^n, X_1^n) - H(Y_t|Y^{t-1}, S^n, X_1^n, X_2^n) \Big).$$

Next we bound $R_1 + R_2$. We have

$$\begin{aligned} n(R_1 + R_2) &= H(W_1, W_2) \\ &= H(W_1, W_2|S^n) \\ &\simeq I(W_1, W_2; Y^n|S^n) \\ &= H(Y^n|S^n) - H(Y^n|S^n, W_1, W_2, X_1^n, X_2^n) \\ &= H(Y^n|S^n) - H(Y^n|S^n, X_1^n, X_2^n) \\ &= \sum_{t=1}^{n} \Big(H(Y_t|Y^{t-1}, S^n) - H(Y_t|Y^{t-1}, S^n, X_1^n, X_2^n) \Big). \end{aligned}$$

Before proceeding further, we specify the nature of state information available at the encoders.

The case $\tau = T-1$. In this case, strictly causal state information is available at the encoders. That is, for $i \in \{1,2\}$ and $t \in [n]$, X_{it} is a deterministic function of (W_i, S_i^{t-1}). Continuing from (4.16), we get

$$\begin{aligned} nR_1 &\lesssim \sum_{t=1}^{n} \Big(H(Y_t|S^t, X_{2t}) - H(Y_t|S^t, X_{1t}, X_{2t}) \Big) \\ &= \sum_{t=1}^{n} I(X_{1t}; Y_t|S^t, X_{2t}). \end{aligned}$$

Similarly, we get

$$\begin{aligned} nR_2 &\lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t|S^t, X_{1t}) \\ n(R_1+R_2) &\lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t|S^t). \end{aligned}$$

Setting $Q_t := S^{t-1}$ for $t \in [n]$ gives

$$\begin{aligned} nR_1 &\lesssim \sum_{t=1}^{n} I(X_{1t}; Y_t|Q_t, S_t, X_{2t}) \\ nR_2 &\lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t|Q_t, S_t, X_{1t}) \\ n(R_1+R_2) &\lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t|Q_t, S_t). \end{aligned}$$

Thus $\mathscr{C}_{T-1}(\mathbf{0},\mathbf{0})$ is contained in the closure of the set of all rate pairs satisfying

$$\begin{aligned} R_1 &\le I(X_1; Y|Q, S, X_2) \\ R_2 &\le I(X_2; Y|Q, S, X_1) \\ R_1 + R_2 &\le I(X_1, X_2; Y|Q, S) \end{aligned}$$

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

The case $\tau = \infty$. In this case, noncausal state information is available at the encoders, meaning that for $i \in \{1,2\}$ and $t \in [n]$, X_{it} is a deterministic function of (W_i, S_i^n). From (4.16), we have

$$\begin{aligned} nR_1 &\lesssim \sum_{t=1}^{n} \Big(H(Y_t|S_1^t, S_2^{t:n}, X_{2t}) - H(Y_t|S_1^t, S_2^{t:n}, X_{1t}, X_{2t}) \Big) \\ &= \sum_{t=1}^{n} I(X_{1t}; Y_t|S_1^t, S_2^{t:n}, X_{2t}), \end{aligned}$$

where for $t \in [n]$,

$$S_2^{t:n} = \big(S_{2t}, S_{2(t+1)}, \ldots, S_{2n}\big).$$

Similarly, we have

$$\begin{aligned} nR_2 &\lesssim \sum_{t=1}^{n} I(X_{2t}; Y_t | S_1^t, S_2^{t:n}, X_{1t}) \\ n(R_1 + R_2) &\lesssim \sum_{t=1}^{n} I(X_{1t}, X_{2t}; Y_t | S_1^t, S_2^{t:n}). \end{aligned}$$

For $t \in [n]$, following [11], define

$$Q_t := (S_1^{t-1}, S_2^{t+1:n}).$$

By assumption, $(S_1^n, S_2^n) \overset{\text{iid}}{\sim} p(s_1, s_2)$. Thus

$$\begin{aligned} p(s_1^n, s_2^n | s_1^{t-1}, s_2^{t+1:n}, s_{1t}, s_{2t}) &= p(s_1^{t+1:n}, s_2^{t-1} | s_1^t, s_2^{t:n}) \\ &= p(s_1^{t+1:n} | s_2^{t+1:n}) p(s_2^{t-1} | s_1^{t-1}), \end{aligned}$$

which implies S_1^n and S_2^n are independent given (Q_t, S_{1t}, S_{2t}). Since (W_1, W_2) is independent of (S_1^n, S_2^n), it follows that for $t \in [n]$, $X_{1t}(W_1, S_1^n)$ and $X_{2t}(W_2, S_2^n)$ are independent given (Q_t, S_{1t}, S_{2t}). Thus $\mathscr{C}_\infty(\mathbf{0}, \mathbf{0})$ is contained in the set of all rate pairs satisfying

$$\begin{aligned} R_1 &\leq I(X_1; Y | Q, S, X_2) \\ R_2 &\leq I(X_2; Y | Q, S, X_1) \\ R_1 + R_2 &\leq I(X_1, X_2; Y | Q, S) \end{aligned}$$

for some distribution $p(q)p(x_1|q, s_1)p(x_2|q, s_2)$.

Chapter 5

THE RELIABILITY BENEFIT

Consider a network consisting of a two-user MAC and a CF as shown in Figure 5.1. The main result of this chapter, Theorem 5.2.1, considers the case where C_{in}^1 and C_{in}^2 are sufficiently large so that the CF has access to both source messages. In such a network, Theorem 5.2.1 shows that whenever C_{out}^1 and C_{out}^2 are positive, the maximal- and average-error capacity regions are equal. Thus, unlike the classical MAC scenario, where codes with small maximal error in general achieve lower rates than codes with small average error, when the encoders cooperate through a CF that has full access to the messages and outgoing links of positive capacity, any rate pair that is achievable with small average error is also achievable with small maximal error. Therefore, cooperation removes the tradeoff that exists between transmission rates and reliability in the classical MAC. We discuss this result in detail in Section 5.2.

In Section 5.3, we apply the equality between maximal- and average-error capacity regions in the scenario described above to Dueck's "contraction MAC," a MAC with maximal-error capacity region strictly smaller than its average-error region [15]. We use this example to prove the existence of a network where a noiseless link of "negligible capacity" has a non-trivial effect on the maximal-error capacity region. Intuitively, a link has negligible capacity if for any function $f(n) = o(n)$ and all sufficiently large n, it can reliably deliver $f(n)$ bits over the channel in n channel uses.

Proposition 5.2.4 in Section 5.2 shows that when C_{in}^1 and C_{in}^2 are small, equality between the maximal-error and average-error capacity regions is not guaranteed. This motivates the definition of a family of error probability constraints that are not as stringent as maximal-error, yet more strict than average-error (Section 5.4). In Theorem 5.4.1, we show that this family captures the reliability benefit of rate-limited cooperation under the CF model. For the proof, we use techniques from [16], in which Willems shows that the average- and maximal-error capacity regions of the discrete memoryless broadcast channel

This material is based upon work supported by the National Science Foundation under Grant Numbers 1527524, 1526771, and 1321129. It originally appears in [55].

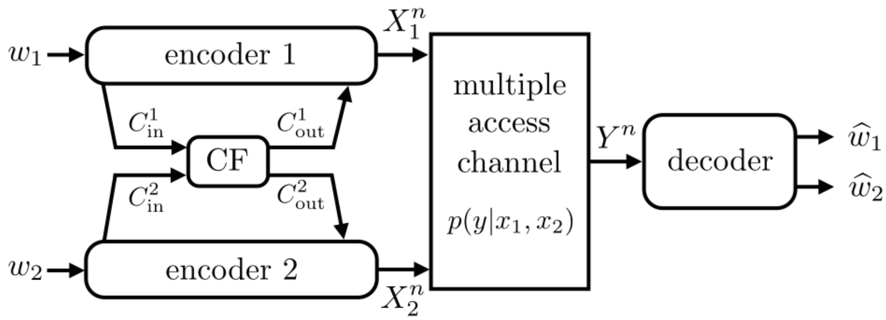

Figure 5.1: A network consisting of two encoders, a CF, a MAC, and a decoder.

are identical.

All proofs appear in Section 5.5.

5.0.1 Prior Work

The edge removal problem [12], [20] studies the change in the capacity region of a network that results from removing a point-to-point channel, here called an "edge," from the network. In this context, Proposition 5.3.1 in Section 5.3 shows that the removal of a negligible capacity edge has a non-negligible effect on the maximal-error capacity region.

The edge removal problem for edges with negligible capacity has some history in the literature. In the context of lossless source coding over networks, Gu, Effros, and Bakshi [18] state the "Vanishment Conjecture," which roughly says that in a class of network source coding problems, certain edges with negligible capacity can be removed with no consequences. In [25] and [24, p. 51], the authors study the relationship between the edge removal problem for edges with negligible capacity and a notion of strong converse. In [56], Sarwate and Gastpar show that feedback via links of negligible capacity does not affect the average-error capacity region of a memoryless MAC. In [26], Langberg and Effros demonstrate a connection between the edge removal problem for edges with negligible capacity and the equivalence between zero-error and ϵ-error capacities in network coding. In work that appeared after [55], Langberg and Effros [57] demonstrate the existence of a network where even a single bit of communication (over the entire transmission block) results in a positive maximal-error sum-capacity gain.

Given that one may view feedback as a form of cooperation, similar questions may be posed about feedback and reliability. In [15], Dueck shows that for some MACs, the maximal-error capacity region with feedback is strictly contained in the average-error region without feedback. This contrasts with our results on encoder cooperation via a CF that has access to both messages and output edges of negligible capacity. Specifically, we show in Section 5.3 that the maximal-error region of a MAC with negligible encoder cooperation of this kind contains the average-error region of the same MAC without encoder cooperation. For further discussion of results regarding feedback and the average- and maximal-error capacity regions of the MAC, we refer the reader to Cai [58].

Other networks under which maximal- and average-error capacity regions are identical include MACs where one of the MAC encoders is “stochastic,” that is, its codewords depend on some randomly generated key in addition to its message. For such codes, the definitions of the maximal- and average-error probabilities require an expectation with respect to the distribution of the random bits. Cai shows in [58] that the maximal-error capacity region of a MAC where one encoder has access to a random key of negligible rate equals the average-error capacity region of the same MAC when both encoders are deterministic. While some of the techniques we use in this chapter are conceptually similar to Cai's proof [58], the respective models are rather different. For example, it holds that stochastic encoders cannot achieve higher rates than deterministic encoders under average error, even if they have access to random keys with positive rates. The same result however, is not true of the cooperation model we study here when the cooperation rate is positive (Chapter 3). That is, at least for some MACs, a positive cooperation rate leads to a strictly positive gain. Furthermore, for a negligible cooperation rate, while we do not demonstrate a gain in the average-error capacity region, the proof that applies in the case of stochastic encoders does not rule out such a gain here.

5.1 Definitions

Consider a network comprising two encoders, a cooperation facilitator (CF), a multiple access channel (MAC)

$$\big(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y}\big),$$

and a decoder as depicted in Figure 5.1. The following definitions aid our description of an (n, M_1, M_2, J)-code with encoder cooperation for a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. Recall that for every real number $x \geq 1$, $[x]$ denotes the set $\{1, \ldots, \lfloor x \rfloor\}$, where $\lfloor x \rfloor$ is the integer part of x. For each $i \in \{1, 2\}$, fix two sequences of sets $(\mathcal{U}_{ij})_{j=1}^J$ and $(\mathcal{V}_{ij})_{j=1}^J$ such that

$$\log \left|\mathcal{U}_i^J\right| = \sum_{j=1}^J \log |\mathcal{U}_{ij}| \leq nC_{\text{in}}^i$$
$$\log \left|\mathcal{V}_i^J\right| = \sum_{j=1}^J \log |\mathcal{V}_{ij}| \leq nC_{\text{out}}^i,$$

where log denotes the base-2 logarithm and for all $j \in [J]$,

$$\mathcal{U}_i^j := \prod_{\ell=1}^j \mathcal{U}_{i\ell}$$
$$\mathcal{V}_i^j := \prod_{\ell=1}^j \mathcal{V}_{i\ell}.$$

Here $\mathcal{U}_{ij}$ represents the alphabet for the round-j transmission from encoder i to the CF while $\mathcal{V}_{ij}$ represents the alphabet for the round-j transmission from the CF to encoder i. The given alphabet size constraints are chosen to match the total rate constraints nC_{in}^i and nC_{out}^i over J rounds of communication between the encoders and the CF. For $i \in \{1, 2\}$, encoder i is represented by $((\varphi_{ij})_{j=1}^J, f_i)$, where

$$\varphi_{ij} : [M_i] \times \mathcal{V}_i^{j-1} \to \mathcal{U}_{ij}$$

captures the round-j transmission from encoder i to the CF, and

$$f_i : [M_i] \times \mathcal{V}_i^J \to \mathcal{X}_i^n$$

captures the transmission of encoder i across the channel.[1] The CF is represented by the functions $((\psi_{1j})_{j=1}^J, (\psi_{2j})_{j=1}^J)$, where for $i \in \{1, 2\}$ and $j \in [J]$,

$$\psi_{ij} : \mathcal{U}_1^j \times \mathcal{U}_2^j \to \mathcal{V}_{ij}$$

captures the round-j transmission from the CF to encoder i. For each message pair (m_1, m_2) and $i \in \{1, 2\}$, define the sequences $(u_{ij})_{j\in[J]}$ and $(v_{ij})_{j\in[J]}$ recursively as

$$u_{ij} := \varphi_{ij}(m_i, v_i^{j-1}) \tag{5.1}$$
$$v_{ij} := \psi_{ij}(u_1^j, u_2^j). \tag{5.2}$$

[1]For notational simplicity, we omit encoder cost constraints (e.g., power constraints in a Gaussian MAC). We note, however, that the same proofs apply in the presence of such constraints.

In round j, encoder i sends u_{ij} to the CF and receives v_{ij} from the CF. After the J-round communication between the encoders and the CF is completed, encoder i transmits $f_i(m_i, v_i^J)$ over the channel. The decoder is represented by the function

$$g : \mathcal{Y}^n \to [M_1] \times [M_2].$$

The collection of mappings

$$\Big((\varphi_{1j})_{j=1}^J, (\varphi_{2j})_{j=1}^J, (\psi_{1j})_{j=1}^J, (\psi_{2j})_{j=1}^J, f_1, f_2, g\Big)$$

defines an (n, M_1, M_2, J)-code for the MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF.

The probability that a message pair (m_1, m_2) is decoded incorrectly is given by

$$\lambda_n(m_1, m_2) := \sum_{y^n \notin g^{-1}(m_1,m_2)} p\big(y^n | f_1(m_1, v_1^J), f_2(m_2, v_2^J)\big), \tag{5.3}$$

where

$$g^{-1}(m_1, m_2) := \big\{y^n \big| g(y^n) = (m_1, m_2)\big\}.$$

Note that λ_n depends only on (m_1, m_2) since by (5.1) and (5.2), v_1^J and v_2^J are deterministic functions of (m_1, m_2). The average probability of error, $P_{e,\text{avg}}^{(n)}$, and the maximal probability of error, $P_{e,\max}^{(n)}$, are defined as

$$P_{e,\text{avg}}^{(n)} = \frac{1}{M_1 M_2} \sum_{m_1,m_2} \lambda_n(m_1, m_2)$$
$$P_{e,\max}^{(n)} = \max_{m_1,m_2} \lambda_n(m_1, m_2),$$

respectively.

We say that a rate pair (R_1, R_2) is average-error achievable for a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF if for all $\epsilon, \delta > 0$ and n sufficiently large, there exists an (n, M_1, M_2, J)-code such that

$$\frac{1}{n} \log M_i \geq R_i - \delta \text{ for } i \in \{1, 2\},$$

and $P_{e,\text{avg}}^{(n)} \leq \epsilon$. We define the average-error capacity region, $\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, as the set of all rates that are average-error achievable. We define the maximal-error capacity region, $\mathscr{C}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$, similarly.

5.2 Average- and Maximal-Error Capacity Regions

The main theorem of this section states that cooperation through a CF that has access to both messages results in a network whose maximal- and average-error capacity regions are identical. We address the necessity of the assumption that the CF has access to both messages in Proposition 5.2.4.

Theorem 5.2.1. *For a given MAC* $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1,x_2), \mathcal{Y})$*, let* $\mathbf{C}_{\text{in}}^* = (C_{\text{in}}^{*1}, C_{\text{in}}^{*2})$ *be any rate vector that satisfies*

$$\min\{C_{\text{in}}^{*1}, C_{\text{in}}^{*2}\} > \max_{p(x_1,x_2)} I(X_1, X_2; Y).$$

Then for every $\mathbf{C}_{\text{out}} \in \mathbb{R}_{>0}^2$*,*

$$\mathscr{C}_{\text{max}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) = \mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}).$$

We next present a sequence of lemmas that prove Theorem 5.2.1. Our first lemma shows that any positive rate, no matter how small, is sufficient for the CF to inform the encoders which message pairs result in a small error at the decoder. The proof, using ideas similar to [16] and [26], appears in Subsection 5.5.1.

Lemma 5.2.2. *For every* $\mathbf{C}_{\text{out}} = (C_{\text{out}}^1, C_{\text{out}}^2)$ *and* $\tilde{\mathbf{C}}_{\text{out}} = (\tilde{C}_{\text{out}}^1, \tilde{C}_{\text{out}}^2)$ *satisfying* $\tilde{C}_{\text{out}}^1 > C_{\text{out}}^1$ *and* $\tilde{C}_{\text{out}}^2 > C_{\text{out}}^2$*,*

$$\mathscr{C}_{\text{max}}(\mathbf{C}_{\text{in}}^*, \tilde{\mathbf{C}}_{\text{out}}) \supseteq \mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}).$$

Note that $\mathscr{C}_{\text{max}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ is also a subset of $\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$. Thus examining the continuity of $\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ in $\mathbf{C}_{\text{out}}$ may be helpful in proving equality between the average- and maximal-error capacity regions. It turns out to be simpler, however, to introduce a function of the capacity region and study the continuity of that function.

Let $\mathscr{C}$ be a compact subset of $\mathbb{R}_{\geq 0}^2$. For every $\alpha \in [0,1]$, define

$$C^\alpha(\mathscr{C}) := \max_{(x,y)\in\mathscr{C}} \big(\alpha x + (1-\alpha)y\big). \tag{5.4}$$

Thus, C^α is the value of the support function of $\mathscr{C}$ computed with respect to the vector $(\alpha, 1-\alpha)$ [59, p. 37]. When $\mathscr{C}$ is the capacity region of a network, $C^{1/2}(\mathscr{C})$ equals half the corresponding sum-capacity.

Now consider a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. For every $\alpha \in [0,1]$, define

$$C^{\alpha}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) := C^{\alpha}\big(\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})\big).$$

Define $C^{\alpha}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ similarly. For $\alpha \in [0,1]$ and pairs $\mathbf{C}_{\text{out}}$ and $\tilde{\mathbf{C}}_{\text{out}}$ satisfying the conditions of Lemma 5.2.2, we have

$$C^{\alpha}_{\max}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) \leq C^{\alpha}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) \leq C^{\alpha}_{\max}(\mathbf{C}^*_{\text{in}}, \tilde{\mathbf{C}}_{\text{out}}). \tag{5.5}$$

The next lemma, for fixed $\alpha \in [0,1]$, investigates the continuity of $C^{\alpha}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ and $C^{\alpha}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ as a function of $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$. Since the proof and thus the result apply in both the maximal- and average-error cases, we omit the "avg" and "max" subscripts in Lemma 5.2.3.

Lemma 5.2.3. *For every $\alpha \in [0,1]$, the mapping $C^{\alpha}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is concave on $\mathbb{R}^4_{\geq 0}$ and thus continuous on $\mathbb{R}^4_{>0}$.*

In Lemma 5.2.3, the proof of concavity follows from a time-sharing argument. Continuity then follows from the fact that any concave function defined on an open convex subset of $\mathbb{R}^n$ is continuous [60, pp. 22-23]. The details are omitted.

Since $C^{\alpha}_{\max}$ is continuous on $\mathbb{R}^4_{>0}$, for any $\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{>0}$, taking the limits $\tilde{C}^1_{\text{out}} \to (C^1_{\text{out}})^+$ and $\tilde{C}^2_{\text{out}} \to (C^2_{\text{out}})^+$ in (5.5) gives

$$C^{\alpha}_{\max}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) = C^{\alpha}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}})$$

for all $\alpha \in [0,1]$. By Lemma 5.6.1 in the appendix, for a given capacity region $\mathscr{C}$, the mapping $\alpha \mapsto C^{\alpha}(\mathscr{C})$ characterizes $\mathscr{C}$ precisely. Thus for every $\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{>0}$, we have

$$\mathscr{C}_{\max}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) = \mathscr{C}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}).$$

One question that arises from Theorem 5.2.1 is whether it is necessary for the CF to have access to both messages in order to guarantee identical maximal- and average-error capacity regions. Proposition 5.2.4 shows that the mentioned condition is necessary; that is, if the CF only has partial access to the messages, regardless of the capacities of the CF output links, the average- and maximal-error regions sometimes differ. The proof appears in Subsection 5.5.3.

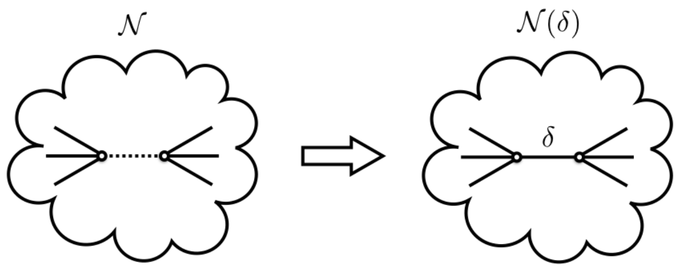

Figure 5.2: Left: A network $\mathcal{N}$ with a single edge of "negligible capacity." Right: The network $\mathcal{N}(\delta)$, where the negligible capacity edge of $\mathcal{N}$ is replaced with an edge of capacity $\delta > 0$.

Proposition 5.2.4. *There exists a MAC* $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$ *and* $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$ *such that for every* $\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{\geq 0}$*,*

$$\mathscr{C}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \neq \mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}).$$

This proposition shows that the maximal probability of error is too stringent a concept for capturing the reliability benefit of rate-limited cooperation. We address this problem in Section 5.4 by defining a continuum of error probability criteria that bridge the gap between average and maximal probability of error. Before moving to that description, we discuss the effect of negligible rate in a network in Section 5.3.

5.3 Effect of Negligible Rate

We begin by giving a rough description of the capacity region of a network containing an edge of negligible capacity. Let $\mathcal{N}$ be a network containing exactly one edge of negligible capacity and possibly other edges of positive capacity. For every $\delta > 0$, let $\mathcal{N}(\delta)$ be the same network with the difference that the edge with negligible capacity is replaced with an edge of capacity δ. (See Figure 5.2.) Then we say a rate vector is achievable over $\mathcal{N}$ if and only if for all $\delta > 0$, that rate vector is achievable over $\mathcal{N}(\delta)$. Formally, if we denote the capacity regions of $\mathcal{N}$ and $\mathcal{N}(\delta)$ with $\mathscr{C}(\mathcal{N})$ and $\mathscr{C}(\mathcal{N}(\delta))$, respectively, then

$$\mathscr{C}(\mathcal{N}) := \bigcap_{\delta > 0} \mathscr{C}(\mathcal{N}(\delta)).$$

We define achievability over networks with multiple edges of negligible capacity inductively.

Based on the above discussion, the capacity region of a network $\mathcal{N}$ comprised of a MAC with a CF that has complete access to both messages and output edges of negligible capacity is given by

$$\mathscr{C}(\mathcal{N}) = \bigcap_{\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{>0}} \mathscr{C}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}),$$

where we again drop the subscript "avg" or "max" since the definition applies in both cases. From Theorem 5.2.1 it follows that for every MAC,

$$\bigcap_{\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{>0}} \mathscr{C}_{\text{max}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) = \bigcap_{\mathbf{C}_{\text{out}} \in \mathbb{R}^2_{>0}} \mathscr{C}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}) \supseteq \mathscr{C}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{0}) = \mathscr{C}_{\text{avg}}(\mathbf{0}, \mathbf{0}), \tag{5.6}$$

where $\mathbf{0} = (0, 0)$. Thus even a negligible cooperation rate suffices to guarantee a small maximal probability of error for rate pairs that without cooperation can only be achieved with small average probability of error.

The reliability gain of negligible cooperation is closely related to the question of the continuity of the capacity region of a network with respect to its edges. Using the ideas discussed above, Proposition 5.3.1 provides conditions under which $C^{\alpha}_{\text{max}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is not continuous with respect to $\mathbf{C}_{\text{out}}$. The proof appears in Subsection 5.5.4.

Proposition 5.3.1. *Fix $\alpha \in (0, 1)$ and $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$. Given any MAC for which*

$$C^{\alpha}_{\text{avg}}(\mathbf{0}, \mathbf{0}) > C^{\alpha}_{\text{max}}(\mathbf{0}, \mathbf{0}), \tag{5.7}$$

$C^{\alpha}_{\text{max}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is not continuous with respect to $\mathbf{C}_{\text{out}}$ at $\mathbf{C}_{\text{out}} = \mathbf{0}$.

In Subsection 5.5.5, we show that Dueck's contraction MAC [15] is an example of a MAC that satisfies (5.7) for *every* $\alpha \in (0, 1)$. This results in the next corollary.

Corollary 5.3.2. *There exists a MAC where for all $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$ and $\alpha \in (0, 1)$, $C^{\alpha}_{\text{max}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is not continuous with respect to $\mathbf{C}_{\text{out}}$ at $\mathbf{C}_{\text{out}} = \mathbf{0}$.*

Note that Corollary 5.3.2 provides an example of a network that does not satisfy the weak edge removal property with respect to the maximal-error capacity region (Chapter 1).

5.4 Cooperation and Reliability

As observed in Section 5.2, to quantify the reliability benefit of cooperation when the capacity of the CF input links are limited, we require a more general notion of probability of error. We next describe this concept.

For $r_1, r_2 \geq 0$, the (r_1, r_2)-error probability $P_e^{(n)}(r_1, r_2)$ is a compromise between average and maximal error probability. The pair (r_1, r_2) provides a measure of reliability, where larger values of r_1 and r_2 correspond to higher reliability. For a given (n, M_1, M_2, J)-code, define the probability of error matrix as

$$\Lambda_n := \big(\lambda_n(m_1, m_2)\big)_{\substack{m_1 \in [M_1] \\ m_2 \in [M_2]}}, \tag{5.8}$$

where $\lambda_n(m_1, m_2)$, given by (5.3), equals the probability of error at the decoder when the pair (m_1, m_2) is transmitted. To compute $P_e^{(n)}(r_1, r_2)$, we partition Λ_n into K_1K_2 blocks of size $L_1 \times L_2$, where for $i \in \{1, 2\}$,

$$\begin{aligned} K_i &:= \min\left\{\lfloor 2^{nr_i} \rfloor, M_i\right\} \\ L_i &:= \lfloor M_i / K_i \rfloor, \end{aligned}$$

and a single block containing the remaining $M_1M_2 - K_1K_2L_1L_2$ entries. We begin by calculating the *average* of the entries within each $L_1 \times L_2$ block and obtain the K_1K_2 values

$$\left\{\frac{1}{L_1L_2} \sum_{\substack{m_1 \in S_{1,k_1} \\ m_2 \in S_{2,k_2}}} \lambda_n(m_1, m_2)\right\}_{\substack{k_1 \in [K_1] \\ k_2 \in [K_2]}}, \tag{5.9}$$

where for $i \in \{1, 2\}$ and $k_i \in [K_i]$, the set $S_{i,k_i} \subseteq [M_i]$ is defined as

$$S_{i,k_i} = \Big\{(k_i - 1)L_i + 1, \ldots, k_iL_i\Big\}. \tag{5.10}$$

Next we find the *maximum* of the K_1K_2 average values, namely

$$\max_{k_1,k_2} \frac{1}{L_1L_2} \sum_{\substack{m_1 \in S_{1,k_1} \\ m_2 \in S_{2,k_2}}} \lambda_n(m_1, m_2). \tag{5.11}$$

The averages in (5.9) and thus the maximum in (5.11) depend on the labeling of the messages, which is not desirable. To avoid this issue, we choose the labeling that minimizes (5.11) over all permutations of the rows and columns of Λ_n. This results in the definition

$$P_e^{(n)}(r_1, r_2) := \min_{\pi_1, \pi_2} \max_{k_1, k_2} \frac{1}{L_1L_2} \sum_{\substack{m_1 \in S_{1,k_1} \\ m_2 \in S_{2,k_2}}} \lambda_n(\pi_1(m_1), \pi_2(m_2)),$$

where the minimum is over all permutations π_1 and π_2 of the sets $[M_1]$ and $[M_2]$, respectively. Note that

$$P_e^{(n)}(0,0) = P_{e,\text{avg}}^{(n)},$$

and for sufficiently large values of r_1 and r_2,

$$P_e^{(n)}(r_1,r_2) = P_{e,\max}^{(n)}.$$

Thus $P_e^{(n)}(r_1,r_2)$ gives a continuum of error probabilities between $P_{e,\text{avg}}^{(n)}$ and $P_{e,\max}^{(n)}$.

We say a rate pair (R_1,R_2) is (r_1,r_2)-error achievable for a MAC with a $(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$-CF if for all $\epsilon,\delta > 0$ and all sufficiently large n, there exists an (n,M_1,M_2,J)-code such that

$$\frac{1}{n}\log(K_iL_i) \geq R_i - \delta \text{ for } i \in \{1,2\}, \tag{5.12}$$

and $P_e^{(n)}(r_1,r_2) \leq \epsilon$. In (5.12), we use K_iL_i instead of M_i since only K_iL_i elements of $[M_i]$ are used in calculating $P_e^{(n)}(r_1,r_2)$. We define the (r_1,r_2)-error capacity region, $\mathscr{C}_{(r_1,r_2)}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$, as the set of all rate pairs that are (r_1,r_2)-error achievable.

The main result of this section, Theorem 5.4.1, says that if a rate pair is achievable for a MAC with a CF under the average error criterion, then sufficiently increasing the capacities of the CF links ensures that the same rate pair is also achievable under a stricter notion of error. This result applies to any MAC whose average-error capacity region is bounded. Prior to stating this result, we introduce notation used in Theorem 5.4.1.

Define R_1^* and R_2^* to be the maximum of R_1 and R_2, respectively, over the average-error capacity region of a MAC with a $(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$-CF. Each rate is maximized when the other rate is set to zero. When one encoder transmits at rate zero, cooperation through a CF is no more powerful than a simple forwarding strategy (Chapter 3). Thus R_1^* and R_2^* equal the corresponding maximal rates in the capacity region of the MAC with conferencing encoders [5]. Thus,

$$\begin{aligned} R_1^* &= \max_{X_1-U-X_2} \min\big\{I(X_1;Y|U,X_2) + C_{12}, I(X_1,X_2;Y)\big\} \\ R_2^* &= \max_{X_1-U-X_2} \min\big\{I(X_2;Y|U,X_1) + C_{21}, I(X_1,X_2;Y)\big\}, \end{aligned}$$

where $C_{12} = \min\{C_{\text{in}}^1, C_{\text{out}}^2\}$ and $C_{21} = \min\{C_{\text{in}}^2, C_{\text{out}}^1\}$.

Theorem 5.4.1 (Reliability under the CF model)**.** *If for* $i \in \{1, 2\}$*,*

$$\tilde{C}^i_{\text{in}} > \min\{C^i_{\text{in}} + \tilde{r}_i, R^*_i\}$$
$$\tilde{C}^i_{\text{out}} > C^i_{\text{out}},$$

then

$$\mathscr{C}_{(r_1,r_2)}(\tilde{\mathbf{C}}_{\text{in}}, \tilde{\mathbf{C}}_{\text{out}}) \supseteq \mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}).$$

A detailed proof of Theorem 5.4.1 appears in Subsection 5.5.6.

Remark. We note that Lemma 5.2.2 follows as a corollary of Theorem 5.4.1. To see this, let $\mathbf{C}_{\text{out}} = (C^1_{\text{out}}, C^2_{\text{out}})$ and $\tilde{\mathbf{C}}_{\text{out}} = (\tilde{C}^1_{\text{out}}, \tilde{C}^2_{\text{out}})$ be elements of $\mathbb{R}^2_{>0}$ such that for $i \in \{1, 2\}$, $\tilde{C}^i_{\text{out}} > C^i_{\text{out}}$. In Theorem 5.4.1, for $i \in \{1, 2\}$, set $\tilde{C}^i_{\text{in}} = C^i_{\text{in}} = C^{*i}_{\text{in}} > R^*_i$ and $\tilde{r}_i > R^*_i$. Then Theorem 5.4.1 implies

$$\mathscr{C}_{\max}(\mathbf{C}^*_{\text{in}}, \tilde{\mathbf{C}}_{\text{out}}) \supseteq \mathscr{C}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}}).$$

5.5 Proofs

This section contains proofs for preceding results. Throughout, proof outlines are followed by formal arguments.

5.5.1 Proof of Lemma 5.2.2 (Reliability under CF model with high capacity CF input links)

The assumption $\mathbf{C}_{\text{in}} = \mathbf{C}^*_{\text{in}}$ implies that the CF has access to both messages. Thus one round of cooperation suffices to achieve any rate pair in the maximal- or average-error capacity region. That is, any function that the CF can compute in J rounds of communication with the encoders, it can also compute in a single round. We therefore set $J = 1$.

Now suppose $(R_1, R_2) \in \mathscr{C}_{\text{avg}}(\mathbf{C}^*_{\text{in}}, \mathbf{C}_{\text{out}})$. Fix $\epsilon, \delta > 0$. Then there exists an $(n, M_1, M_2, 1)$-code such that

$$\frac{1}{n} \log M_i \geq R_i - \delta$$

and $P^{(n)}_{e,\text{avg}} \leq \epsilon$. Note that for this code, the CF can be represented by a pair of functions (ψ_1, ψ_2), where for $i \in \{1, 2\}$,

$$\psi_i : [M_1] \times [M_2] \to [2^{nC^i_{\text{out}}}].$$

To keep track of the message pairs that have small probability of error, we define the $(0,1)$-matrix $A_n = (a_n(m_1, m_2))_{m_1,m_2}$ as

$$a_n(m_1, m_2) := \begin{cases} 1 & \text{if } \lambda_n(m_1, m_2) > e^3\epsilon \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda_n(m_1, m_2)$ is the probability of error when the pair (m_1, m_2) is transmitted. We next (deterministically) partition this matrix into blocks of size $M_* \times M_*$, where

$$M_* = \left\lceil n\Big(\max_{p(x_1,x_2)} I(X_1, X_2; Y) + 2\delta \Big) \right\rceil. \tag{5.13}$$

Suppose that for $i \in \{1, 2\}$, encoder i wants to transmit message m_i. If there is at least one zero entry in the $M_* \times M_*$ block containing (m_1, m_2), then the CF, using $\log M_*$ bits, sends to the encoders the location of that zero entry. This is possible since the CF has access to both messages and the fixed partition, and thus knows which block contains the message pair. Assuming the zero entry is located at (m_1^*, m_2^*), the CF also sends $\psi_i(m_1^*, m_2^*)$ to encoder i for $i \in \{1, 2\}$. This is possible since for sufficiently large n,

$$\tilde{C}^i_{\text{out}} > C^i_{\text{out}} + \frac{1}{n} \log M_*.$$

Then encoder i transmits the codeword corresponding to

$$\big(m_i^*, \psi_i(m_1^*, m_2^*)\big).$$

Note that the code described above has a maximal probability of error of at most $e^3\epsilon$. Furthermore, the transmission rate of encoder i, for $i \in \{1, 2\}$, is at least as large as

$$\frac{1}{n} \log \frac{M_i}{M_*} \geq R_i - \delta - \frac{1}{n} \log M_*,$$

which exceeds $R_i - 2\delta$ for sufficiently large n, since $M_* = O(n)$.

In the scenario where not every $M_* \times M_*$ block in A_n contains a zero entry, we permute the rows and columns such that at least one zero entry ends up in each block. To prove the existence of such permutations, we use the next lemma.

Lemma 5.5.1 (Permutation Lemma). *Let $A = (a_{ij})_{i,j=1}^{m,n}$ be a $(0,1)$-matrix and let $N(A)$ denote the number of ones in A. Suppose k is a positive integer* *smaller than or equal to* $\min\{m, n\}$. *For any pair of permutations* (π_1, π_2), *where* π_1 *is a permutation on* $[m]$ *and* π_2 *is a permutation on* $[n]$, *and every* $(s,t) \in [\frac{m}{k}] \times [\frac{n}{k}]$, *define the* $k \times k$ *matrix* $B_{st}(\pi_1, \pi_2)$ *as*

$$B_{st}(\pi_1, \pi_2) := \big(a_{\pi_1(i)\pi_2(j)}\big),$$

where $i \in \{(s-1)k+1, \ldots, sk\}$ *and* $j \in \{(t-1)k+1, \ldots, tk\}$. *If*

$$\frac{mn}{k^2}\left[\frac{N(A)e^2}{mn}\right]^k < 1, \tag{5.14}$$

then there exists a pair of permutations (π_1, π_2) *such that for every* (s,t), $B_{st}(\pi_1, \pi_2)$ *contains at least one zero entry.*

The proof of Lemma 5.5.1 appears in Subsection 5.5.2, below.

To apply Lemma 5.5.1 to our matrix A_n, we first bound the number of ones in A_n as

$$\begin{aligned} N(A_n) &= \sum_{m_1,m_2} a_n(m_1, m_2) \\ &\leq \frac{1}{e^3\epsilon}\sum_{m_1,m_2} \lambda_n(m_1, m_2) \\ &= \frac{1}{e^3\epsilon} M_1 M_2 P_{e,\text{avg}}^{(n)} \leq M_1 M_2 e^{-3}. \end{aligned}$$

Thus

$$\begin{aligned} \frac{M_1 M_2}{M_*^2}\left[\frac{N(A_n)e^2}{M_1 M_2}\right]^{M_*} &= \frac{M_1 M_2}{M_*^2} e^{-M_*} \\ &= 2^{\log M_1 M_2 - M_* \log e - 2\log M_*}, \end{aligned}$$

which is less than one for sufficiently large n by the definition of M_* (Equation (5.13)). Applying Lemma 5.5.1 completes the proof.

5.5.2 Proof of Lemma 5.5.1 (Existence of good permutations)

Let $A = (a_{ij})_{i,j=1}^{m,n}$ be a $(0,1)$-matrix. We apply the probabilistic method. Let Π_1 and Π_2 be independent and uniformly distributed random variables on the set of all permutations of $[m]$ and $[n]$, respectively. Since $N(A)$ denotes the number of ones in A, we have

$$N(A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}.$$

For $(s,t) \in [\frac{m}{k}] \times [\frac{n}{k}]$, define the $k \times k$ matrix $B_{st}(\Pi_1, \Pi_2)$ as

$$B_{st}(\Pi_1, \Pi_2) = \big(a_{\Pi_1(i)\Pi_2(j)}\big),$$

where $i \in \{(s-1)k+1, \ldots, sk\}$ and $j \in \{(t-1)k+1, \ldots, tk\}$. Let J_k denote the $k \times k$ matrix consisting of all ones. By the union bound,

$$\Pr\Big\{\exists (s,t) : B_{st}(\Pi_1, \Pi_2) = J_k\Big\} \le \frac{mn}{k^2} \Pr\big\{B_{11}(\Pi_1, \Pi_2) = J_k\big\}. \tag{5.15}$$

We next find an upper bound for the right hand side of (5.15). Consider the pairs (S_1, S_2) and (τ_1, τ_2), where $S_1 \subseteq [m]$, $S_2 \subseteq [n]$, $|S_1| = |S_2| = k$, and τ_1 and τ_2 are permutations on $[k]$. Denote the elements of S_1 and S_2 by

$$S_1 = \{i_1, \ldots, i_k\}$$
$$S_2 = \{j_1, \ldots, j_k\}.$$

Define $E_{S_1S_2}^{\tau_1\tau_2}$ as the event where for all $\ell \in [k]$, $\Pi_1(\ell) = i_{\tau_1(\ell)}$ and $\Pi_2(\ell) = j_{\tau_2(\ell)}$. In other words, when $E_{S_1S_2}^{\tau_1\tau_2}$ occurs, $B_{11}(\Pi_1, \Pi_2)$ is a (permuted) submatrix of A with row indices $(i_{\tau_1(\ell)})_{\ell\in[k]}$ and column indices $(j_{\tau_2(\ell)})_{\ell\in[k]}$. Then

$$\begin{aligned}
&\Pr\Big\{B_{11}(\Pi_1, \Pi_2) = J_k\Big\} \\
&\quad \le \Pr\Big\{\forall \ell \in [k] : a_{\Pi_1(\ell)\Pi_2(\ell)} = 1\Big\} \\
&\quad = \sum_{S_1, S_2, \tau_1, \tau_2} \Pr\big(E_{S_1S_2}^{\tau_1\tau_2}\big) \Pr\Big\{\forall \ell \in [k] : a_{\Pi_1(\ell)\Pi_2(\ell)} = 1 \Big| E_{S_1S_2}^{\tau_1\tau_2}\Big\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\Pr\big(E_{S_1S_2}^{\tau_1\tau_2}\big) &= \Pr\Big\{\forall \ell \in [k] : \Pi_1(\ell) = i_{\tau_1(\ell)}, \Pi_2(\ell) = j_{\tau_2(\ell)}\Big\} \\
&\overset{(a)}{=} \Pr\Big\{\forall \ell \in [k] : \Pi_1(\ell) = i_{\tau_1(\ell)}\Big\} \times \Pr\Big\{\forall \ell \in [k] : \Pi_2(\ell) = j_{\tau_2(\ell)}\Big\} \\
&\overset{(b)}{=} \frac{(m-k)!}{m!} \times \frac{(n-k)!}{n!} = \frac{1}{(k!)^2\binom{m}{k}\binom{n}{k}},
\end{aligned}$$

where (a) follows from the independence of Π_1 and Π_2, and (b) follows from the fact that Π_1 and Π_2 are uniformly distributed. Furthermore,

$$\Pr\Big\{\forall \ell \in [k] : a_{\Pi_1(\ell)\Pi_2(\ell)} = 1 \Big| E_{S_1S_2}^{\tau_1\tau_2}\Big\} = \mathbf{1}\big\{\forall \ell \in [k] : a_{i_{\tau_1(\ell)} j_{\tau_2(\ell)}} = 1\big\}.$$

Thus

$$\Pr\Big\{B_{11}(\Pi_1, \Pi_2) = J_k\Big\} \le \frac{1}{(k!)^2\binom{m}{k}\binom{n}{k}} \sum_{S_1, S_2} \sum_{\tau_1, \tau_2} \mathbf{1}\big\{\forall \ell \in [k] : a_{i_{\tau_1(\ell)} j_{\tau_2(\ell)}} = 1\big\}.$$

For a fixed pair (S_1, S_2), we have

$$\begin{aligned}
&\sum_{\tau_1}\sum_{\tau_2} \mathbf{1}\big\{\forall \ell \in [k] : a_{i_{\tau_1(\ell)} j_{\tau_2(\ell)}} = 1\big\} \\
&\quad = \sum_{\tau_1}\sum_{\tau_2} \mathbf{1}\big\{\forall \ell \in [k] : a_{i_{(\tau_1 \circ \tau_2^{-1})(\ell)} j_\ell} = 1\big\} \\
&\quad = k! \sum_{\tau} \mathbf{1}\big\{\forall \ell \in [k] : a_{i_{\tau(\ell)} j_\ell} = 1\big\}. \qquad (5.16)
\end{aligned}$$

Note that the expression in (5.16) equals $k!$ times the number of k-subsets of $S_1 \times S_2$ that consist only of ones and have exactly one entry in each row and each column. Summing over all S_1 and S_2, we see that that the total number of such subsets is bounded from above by $\binom{N(A)}{k}$. Thus

$$\Pr\Big\{B_{11}(\Pi_1, \Pi_2) = J_k\Big\} \leq \frac{k!\binom{N(A)}{k}}{(k!)^2\binom{m}{k}\binom{n}{k}} = \frac{\binom{N(A)}{k}}{k!\binom{m}{k}\binom{n}{k}}. \qquad (5.17)$$

Therefore,

$$\begin{aligned}
\Pr\Big\{\exists (s,t) : B_{st}(\Pi_1, \Pi_2) = J_k\Big\} &\overset{(a)}{\leq} \frac{mn}{k^2} \times \frac{\binom{N(A)}{k}}{k!\binom{m}{k}\binom{n}{k}} \\
&\overset{(b)}{\leq} \frac{mn}{k^2} \times \frac{\left(\frac{N(A)e}{k}\right)^k}{\left(\frac{m}{e}\right)^k\left(\frac{n}{k}\right)^k} \\
&= \frac{mn}{k^2}\left(\frac{N(A)e^2}{mn}\right)^k,
\end{aligned}$$

where (a) follows from combining (5.15) and (5.17), and (b) follows from Lemma 5.5.2 [61, Appendix C.1], which is stated below.

Lemma 5.5.2. *For integers k and n that satisfy $1 \leq k \leq n$, we have*

$$\left(\frac{n}{k}\right)^k \leq \frac{1}{k!}\left(\frac{n}{e}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k.$$

5.5.3 Proof of Proposition 5.2.4 (Necessity of high capacity CF input links)

We begin by providing an upper bound on the maximal-error sum-capacity gain.

Lemma 5.5.3. *For any MAC and any $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \in \mathbb{R}^4_{\geq 0}$,*

$$C^{1/2}_{\text{max}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \leq C^{1/2}_{\text{max}}(\mathbf{0}, \mathbf{0}) + \frac{C^1_{\text{in}} + C^2_{\text{in}}}{2}.$$

Proof. For every (n, M_1, M_2, J)-code for the MAC with a CF, the set of all messages that lead to the same CF inputs is of the form $\mathcal{A}_1 \times \mathcal{A}_2$ for some $\mathcal{A}_i \subseteq [M_i]$ for $i \in \{1, 2\}$. This is demonstrated in the proof of Proposition 3.2.5. Now fix a sequence of (n, M_1, M_2, J)-codes that achieve the rate pair (R_1^*, R_2^*), where

$$R_1^* + R_2^* = 2C_{\max}^{1/2}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}).$$

Since there are at most $2^{n(C_{\text{in}}^1 + C_{\text{in}}^2)}$ possible CF inputs, the pigeonhole principle implies that there exist sets $\mathcal{A}_1^* \subseteq [M_1]$ and $\mathcal{A}_2^* \subseteq [M_2]$ such that

$$|\mathcal{A}_1^*| \times |\mathcal{A}_2^*| \geq M_1 M_2 2^{-n(C_{\text{in}}^1 + C_{\text{in}}^2)},$$

and the set of all message pairs in $\mathcal{A}_1^* \times \mathcal{A}_2^*$ lead to the same CF inputs. For $i \in \{1, 2\}$ and sufficiently large n, we have

$$\frac{1}{n} \log M_i \geq R_i^* - \delta,$$

thus

$$\begin{aligned} \frac{1}{n} \log \Big(|\mathcal{A}_1^*| |\mathcal{A}_2^*| \Big) &\geq R_1^* + R_2^* - C_{\text{in}}^1 - C_{\text{in}}^2 - 2\delta \\ &= 2C_{\max}^{1/2}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) - C_{\text{in}}^1 - C_{\text{in}}^2 - 2\delta. \end{aligned} \tag{5.18}$$

Now consider the code where for $i \in \{1, 2\}$, encoder i transmits codewords from the original code that correspond to messages in $\mathcal{A}_i^*$. This code has small maximal error probability since the maximal error probability over $\mathcal{A}_1^* \times \mathcal{A}_2^*$ is at most as large as the maximal error probability of the original code over $[M_1] \times [M_2]$. This new code achieves a sum-rate of at least

$$\frac{1}{n} \log \Big(|\mathcal{A}_1^*| |\mathcal{A}_2^*| \Big).$$

Thus by (5.18),

$$C_{\max}^{1/2}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}) \leq C_{\max}^{1/2}(\mathbf{0}, \mathbf{0}) + \frac{C_{\text{in}}^1 + C_{\text{in}}^2}{2}.$$

□

We now prove Proposition 5.2.4 by showing that for Dueck's contraction MAC [15], there exists $\mathbf{C}_{\text{in}} \in \mathbb{R}_{>0}^2$ such that for every $\mathbf{C}_{\text{out}} \in \mathbb{R}_{\geq 0}^2$, $\mathscr{C}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is a proper subset of $\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$. In Subsection 5.5.5, we show that for the contraction MAC,

$$C_{\text{avg}}^{1/2}(\mathbf{0}, \mathbf{0}) > C_{\max}^{1/2}(\mathbf{0}, \mathbf{0}).$$

Thus it is possible to choose $\mathbf{C}_{\text{in}} = (C_{\text{in}}^1, C_{\text{in}}^2) \in \mathbb{R}_{>0}^2$ such that

$$C_{\text{avg}}^{1/2}(\mathbf{0},\mathbf{0}) - C_{\text{max}}^{1/2}(\mathbf{0},\mathbf{0}) > \frac{C_{\text{in}}^1 + C_{\text{in}}^2}{2}.$$

For this $\mathbf{C}_{\text{in}}$ and every $\mathbf{C}_{\text{out}} \in \mathbb{R}_{\geq 0}^2$, we have

$$\begin{aligned} C_{\text{max}}^{1/2}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}}) &\leq C_{\text{max}}^{1/2}(\mathbf{0},\mathbf{0}) + \frac{C_{\text{in}}^1 + C_{\text{in}}^2}{2} \qquad (5.19) \\ &< C_{\text{avg}}^{1/2}(\mathbf{0},\mathbf{0}) \\ &\leq C_{\text{avg}}^{1/2}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}}), \end{aligned}$$

where (5.19) follows from Lemma 5.5.3. This completes the proof.

5.5.4 Proposition 5.3.1 (Discontinuity of C^α under CF model)

We first use the condition

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) > C_{\text{max}}^{\alpha}(\mathbf{0},\mathbf{0}),$$

together with Theorem 5.2.1 to show that for sufficiently large C_{in}^1 and C_{in}^2, $C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$ is not continuous at $\mathbf{C}_{\text{out}} = \mathbf{0}$. The proposition then follows by using the concavity of C_{max}^{α} (Lemma 5.2.3).

Choose $\lambda \in (0,1)$ such that

$$\min\{C_{\text{in}}^1, C_{\text{in}}^2\} > \lambda \max_{p(x_1,x_2)} I(X_1,X_2;Y),$$

and define $\mathbf{C}_{\text{in}}^* = (C_{\text{in}}^{*1}, C_{\text{in}}^{*2})$, where $C_{\text{in}}^{*i} = C_{\text{in}}^i/\lambda$ for $i \in \{1,2\}$. Then by Theorem 5.2.1,

$$\begin{aligned} \lim_{C_{\text{out}}\to 0^+} C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, (C_{\text{out}}, C_{\text{out}})) &= \lim_{C_{\text{out}}\to 0^+} C_{\text{avg}}^{\alpha}(\mathbf{C}_{\text{in}}^*, (C_{\text{out}}, C_{\text{out}})) \\ &\geq C_{\text{avg}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{0}) = C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) \\ &> C_{\text{max}}^{\alpha}(\mathbf{0},\mathbf{0}) = C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{0}). \end{aligned}$$

Thus $C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ is not continuous in $\mathbf{C}_{\text{out}}$. Now by the concavity of $C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}})$ (Lemma 5.2.3),

$$C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}}) \geq \lambda C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) + (1-\lambda) C_{\text{max}}^{\alpha}(\mathbf{0}, \mathbf{C}_{\text{out}}),$$

which can be rearranged as

$$C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}},\mathbf{C}_{\text{out}}) - C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}},\mathbf{0}) \geq \lambda\Big(C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) - C_{\text{max}}^{\alpha}(\mathbf{C}_{\text{in}}^*, \mathbf{0})\Big),$$

since

$$C^{\alpha}_{\max}(\mathbf{0}, \mathbf{C}_{\text{out}}) = C^{\alpha}_{\max}(\mathbf{0}, \mathbf{0}) = C^{\alpha}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{0}) = C^{\alpha}_{\max}(\mathbf{C}^{*}_{\text{in}}, \mathbf{0}).$$

The fact that $\lambda > 0$, together with the discontinuity of $C^{\alpha}_{\max}(\mathbf{C}^{*}_{\text{in}}, \mathbf{C}_{\text{out}})$ at $\mathbf{C}_{\text{out}} = \mathbf{0}$, implies that $C^{\alpha}_{\max}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$ is not continuous in $\mathbf{C}_{\text{out}}$ for arbitrary $\mathbf{C}_{\text{in}} \in \mathbb{R}^2_{>0}$.

5.5.5 Proof of Corollary 5.3.2 (Dueck's Contraction MAC)

Dueck's introduction of the "contraction MAC" in [15] proves the existence of multiterminal networks where the maximal-error capacity region is a strict subset of the average-error capacity region. The input and output alphabets of the contraction MAC are given by

$$\begin{aligned}
\mathcal{X}_1 &= \{A, B, a, b\} \\
\mathcal{X}_2 &= \{0, 1\} \\
\mathcal{Y} &= \{A, B, C, a, b, c\} \times \{0, 1\}.
\end{aligned}$$

The channel is deterministic and defined by the function $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$, where

$$\begin{aligned}
f(a, 0) = f(b, 0) &= (c, 0) \\
f(A, 1) = f(B, 1) &= (C, 1),
\end{aligned}$$

and $f(x_1, x_2) = (x_1, x_2)$ for all other (x_1, x_2). Dueck [15] shows that the maximal-error capacity region of this channel is contained in the set of all rate pairs (R_1, R_2) that satisfy

$$\begin{aligned}
R_1 &\leq \log 3 - p \\
R_2 &\leq h(p)
\end{aligned}$$

for some $0 \leq p \leq 1/2$, where $h(p)$ denotes the binary entropy function

$$h(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}.$$

Thus for every $\alpha \in [0, 1]$,

$$\begin{aligned}
C^{\alpha}_{\max}(\mathbf{0}, \mathbf{0}) &\leq \max_{p \in [0, 1/2]} \Big[\alpha(\log 3 - p) + (1 - \alpha) h(p)\Big] \\
&= \alpha(\log 3 - 1) + (1 - \alpha) \log \left(1 + 2^{\frac{\alpha}{1 - \alpha}}\right);
\end{aligned}$$

the maximum is achieved by

$$p^* = \frac{1}{1 + 2^{\frac{\alpha}{1-\alpha}}}.$$

To show that

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) > C_{\max}^{\alpha}(\mathbf{0},\mathbf{0}) \text{ for all } \alpha \in (0,1),$$

we find a lower bound for $C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0})$. From the average-error capacity region of the MAC [2]–[4], it follows that for $\alpha \in [0, 1/2]$,

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) = \max_{p(x_1)p(x_2)} \Big(\alpha I(X_1;Y) + (1-\alpha) I(X_2;Y|X_1)\Big)$$

and for $\alpha \in [1/2, 1]$,

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) = \max_{p(x_1)p(x_2)} \Big(\alpha I(X_1;Y|X_2) + (1-\alpha) I(X_2;Y)\Big).$$

Since the contraction MAC is deterministic, the above equations simplify to

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) = \max_{p(x_1)p(x_2)} \Big(\alpha H(Y) + (1-2\alpha) H(Y|X_1)\Big) \tag{5.20}$$

for $\alpha \in [0, 1/2]$ and

$$C_{\text{avg}}^{\alpha}(\mathbf{0},\mathbf{0}) = \max_{p(x_1)p(x_2)} \Big((1-\alpha) H(Y) + (2\alpha-1) H(Y|X_2)\Big) \tag{5.21}$$

for $\alpha \in [1/2, 1]$. Let the input distribution of the first transmitter be given by

$$p_{X_1}(A) = p_A, p_{X_1}(B) = p_B, p_{X_1}(a) = p_a, p_{X_1}(b) = p_b,$$

and the input distribution of the second transmitter be given by $p_{X_2}(1) = q$ and $p_{X_2}(0) = 1-q$. In addition, let Y_1 and Y_2 denote the components of Y so that $Y = (Y_1, Y_2)$. Note that $Y_2 = X_2$, and

$$\begin{aligned} H(Y) &= H(Y_1, Y_2) \\ &= H(Y_2) + H(Y_1|Y_2) \\ &= h(q) + qH(p_a, p_b, p_A + p_B) + (1-q)H(p_A, p_B, p_a + p_b). \end{aligned}$$

Furthermore,

$$\begin{aligned} H(Y|X_1) &= H(Y_1, Y_2|X_1) \\ &= H(Y_2|X_1) = h(q), \end{aligned}$$

and

$$\begin{aligned} H(Y|X_2) &= H(Y_1, Y_2|X_2) \\ &= H(Y_1, Y_2, X_2) - H(X_2) \\ &= H(Y) - h(q). \end{aligned}$$

From (5.20) and (5.21) it follows for all $\alpha \in [0, 1]$,

$$\begin{aligned} C^{\alpha}_{\text{avg}}(\mathbf{0}, \mathbf{0}) &\geq \alpha H(Y) + (1 - 2\alpha)H(q) \\ &= (1-\alpha)h(q) + \alpha\big[qH(p_a, p_b, p_A + p_B) + (1-q)H(p_A, p_B, p_a + p_b)\big]. \end{aligned}$$

If we set $q = p^*$, $p_A = p_B = 1/3$, and $p_a = p_b = 1/6$, we get

$$C^{\alpha}_{\text{avg}}(\mathbf{0}, \mathbf{0}) \geq (1-\alpha)h(p^*) + \alpha(\log 3 - p^*/3).$$

Recall that

$$C^{\alpha}_{\max}(\mathbf{0}, \mathbf{0}) \leq (1-\alpha)h(p^*) + \alpha(\log 3 - p^*).$$

Thus $C^{\alpha}_{\text{avg}}(\mathbf{0}, \mathbf{0}) > C^{\alpha}_{\max}(\mathbf{0}, \mathbf{0})$, unless $\alpha = 0$ or $p^* = 0$ (which occurs if and only if $\alpha = 1$).

5.5.6 Proof of Theorem 5.4.1 (Reliability under the CF model)

The argument that follows involves modifying an (n, M_1, M_2, J) average-error code for a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF to get an $(n, \tilde{M}_1, \tilde{M}_2, \tilde{J})$ code for the same MAC with a $(\tilde{\mathbf{C}}_{\text{in}}, \tilde{\mathbf{C}}_{\text{out}})$-CF. Our aim is to obtain small $(\tilde{r}_1, \tilde{r}_2)$ probability of error and $\frac{1}{n}\log \tilde{M}_i$ only slightly smaller than $\frac{1}{n}\log M_i$ for $i \in \{1, 2\}$. To achieve this goal, we first partition Λ_n, as given by (5.8), into $2^{n\tilde{r}_1} \times 2^{n\tilde{r}_2}$ blocks. For $m_i \in [M_i]$, let k_i denote the first $n\tilde{r}_i$ bits of m_i. We next construct a $2^{n\tilde{r}_1} \times 2^{n\tilde{r}_2}$ $(0, 1)$-matrix, where entry (k_1, k_2) equals zero if the average of the $\lambda_n(m_1, m_2)$ entries in the corresponding (k_1, k_2)-block of Λ_n is small, and equals one otherwise. (See Figure 5.3.)

Next, we partition our $(0, 1)$-matrix into blocks of size roughly $n \times n$. In the first cooperation round, encoder i sends the first $n\tilde{r}_i$ bits of m_i to the CF so that the CF knows the block in the $(0, 1)$-matrix that contains (m_1, m_2). If there is at least one zero entry in that block, the CF sends the location of that entry back to each encoder using $\log n$ bits. Then encoder i modifies the first $n\tilde{r}_i$ bits of its message and communicates with the CF over J rounds using the original average-error code. As a result of transmitting (m_1, m_2) pairs that

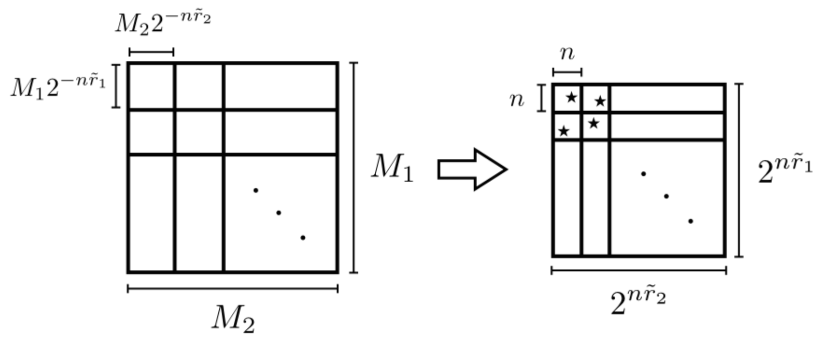

Figure 5.3: Left: The $M_1 \times M_2$ matrix Λ_n with entries $\lambda_n(m_1, m_2)$. Right: The $(0,1)$-matrix constructed from Λ_n. The stars indicate the location of the zeros.

correspond to zeros in our $(0,1)$-matrix, the encoders ensure a small (r_1, r_2)-probability of error.

It may be the case that not every block contains a zero entry. Lemma 5.5.1 shows that if there is a sufficiently large number of zeros in the $(0,1)$-matrix, then there exists a permutation of the rows and a permutation of the columns such that each block of the permuted matrix contains at least one zero entry. Since the original code has a small average error, it follows that our $(0,1)$-matrix has a large number of zeros.

For the formal proof, we wish to show that if

$$\begin{aligned} \tilde{C}^i_{\text{in}} &> \min\{C^i_{\text{in}} + \tilde{r}_i, R^*_i\} \\ \tilde{C}^i_{\text{out}} &> C^i_{\text{out}} \end{aligned}$$

for $i \in \{1, 2\}$, then

$$\mathscr{C}_{(\tilde{r}_1,\tilde{r}_2)}(\tilde{\mathbf{C}}_{\text{in}}, \tilde{\mathbf{C}}_{\text{out}}) \supseteq \mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}}).$$

Recall that R^*_1 and R^*_2 are defined as the maximum of R_1 and R_2, respectively, over the capacity region of a MAC with a $(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$-CF. Our proof follows [16], where Willems proves that the maximal- and average-error capacity regions of the broadcast channel are identical.

Suppose (R_1, R_2) is in $\mathscr{C}_{\text{avg}}(\mathbf{C}_{\text{in}}, \mathbf{C}_{\text{out}})$. Assume $\tilde{r}_1, \tilde{r}_2, R_1, R_2$ are all positive. We discuss the case where some of these quantities are zero at the end of this subsection. Fix $\epsilon, \delta > 0$. Then for sufficiently large N and any $n > N$, there exists an (n, M_1, M_2, J)-code such that for $i = 1, 2$,

$$\log \left|\mathcal{U}_i^J\right| \leq nC_{\text{in}}^i \tag{5.22}$$

$$\log \left|\mathcal{V}_i^J\right| \leq nC_{\text{out}}^i \tag{5.23}$$

$$\frac{1}{n}\log M_i \geq R_i - \delta \tag{5.24}$$

and $P_{e,\text{avg}}^{(n)} \leq \epsilon$. In addition, from Fano's inequality it follows that for sufficiently large n,

$$\frac{1}{n}\log M_i \leq R_i^* + \delta. \tag{5.25}$$

Let $K_* = \lceil n(R_1^* + R_2^* + 2\delta)\rceil$. For $i \in \{1, 2\}$, define $K_i = \min\{K_*\lfloor 2^{n\tilde{r}_i}\rfloor, M_i\}$ and $L_i = \lfloor M_i/K_i\rfloor$. From the set $[M_1]$ choose the K_1L_1 messages that have the smallest

$$\sum_{m_2=1}^{M_2} \lambda_n(m_1, m_2),$$

and renumber them as $\{1, \ldots, K_1L_1\}$. Similarly, from the set $[M_2]$ choose K_2L_2 messages that have the smallest

$$\sum_{m_1=1}^{K_1L_1} \lambda_n(m_1, m_2)$$

and renumber them as $\{1, \ldots, K_2L_2\}$. Then

$$\begin{aligned}\frac{1}{K_1L_1K_2L_2}&\sum_{m_1=1}^{K_1L_1}\sum_{m_2=1}^{K_2L_2} \lambda_n(m_1, m_2)\\ &\leq \frac{1}{M_1M_2}\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2} \lambda_n(m_1, m_2) \leq \epsilon.\end{aligned} \tag{5.26}$$

Next, for every $(k_1, k_2) \in [K_1] \times [K_2]$, define $a_n(k_1, k_2)$ as

$$a_n(k_1, k_2) := \begin{cases} 1 & \text{if } \sum_{S_{1,k_1}\times S_{2,k_2}} \lambda_n(m_1, m_2) > L_1L_2e^3\epsilon \\ 0 & \text{otherwise,} \end{cases}$$

where S_{1,k_1} and S_{2,k_2} are defined by (5.10). Set $A_n := (a_n(k_1, k_2))_{k_1,k_2}$, and let

$N(A_n)$ denote the number of ones in A_n. Then

$$\begin{aligned} N(A_n) &= \sum_{k_1,k_2} a_n(k_1,k_2) \\ &\le \frac{1}{L_1L_2e^3\epsilon}\sum_{k_1,k_2}\sum_{S_{1,k_1}\times S_{2,k_2}} \lambda_n(m_1,m_2) \\ &= \frac{1}{L_1L_2e^3\epsilon}\sum_{m_1=1}^{K_1L_1}\sum_{m_2=1}^{K_2L_2}\lambda_n(m_1,m_2) \\ &\le K_1K_2e^{-3}, \end{aligned} \tag{5.27}$$

where the last inequality follows from (5.26).

Let α_n denote the quantity on the left hand side of (5.14) in Lemma 5.5.1. Then

$$\alpha_n = \frac{K_1K_2}{K_*^2}\Big(\frac{N(A)e^2}{K_1K_2}\Big)^{K_*}.$$

To apply Lemma 5.5.1, we must first prove $\alpha_n < 1$. We have

$$\begin{aligned} \alpha_n &\overset{(a)}{\le} \frac{K_1K_2}{K_*^2e^{K_*}} \\ &\overset{(b)}{\le} 2^{n(R_1^*+R_2^*+2\delta)-K_*\log e-2\log K_*} \overset{(c)}{<} 1, \end{aligned}$$

where (a) follows from (5.27), (b) follows from (5.25) and the fact that $K_i \le M_i$, and (c) follows from the fact that $K_* = \lceil n(R_1^*+R_2^*+2\delta)\rceil$. Thus by Lemma 5.5.1, there exist permutations π_1 and π_2 on the sets $[K_1]$ and $[K_2]$, respectively, such that if we partition the matrix $(a_{\pi_1(k_1)\pi_2(k_2)})_{k_1,k_2}$ into blocks of size $K_*\times K_*$, then there is at least one zero in each block. For $i\in\{1,2\}$, let

$$K_i^* := \lfloor K_i/K_*\rfloor.$$

Note that the partition of the matrix $(a_{\pi_1(k_1)\pi_2(k_2)})_{k_1,k_2}$ contains at least $K_1^*\times K_2^*$ blocks.

Next we use the partition defined above to construct a code that achieves a rate pair sufficiently close to (R_1,R_2) under $(\tilde r_1,\tilde r_2)$-error. For $i\in\{1,2\}$, encoder i splits its message as $m_i=(k_i,\ell_i)\in[K_i]\times[L_i]$ and sends k_i to the CF. Let $(\pi_1(k_1^*),\pi_2(k_2^*))$ be the good entry in the $K_*\times K_*$ block containing the pair $(\pi_1(k_1),\pi_2(k_2))$. For $i\in\{1,2\}$, the CF sends the difference $\pi_i(k_i^*)-\pi_i(k_i)$ (mod K_*) back to encoder i. Encoder 1 and encoder 2 then use the original average-error code with J rounds of cooperation to transmit the message pair

(m_1^*, m_2^*) where for $i \in \{1,2\}$, $m_i^* = (\pi_i(k_i^*), \ell_i)$. By combining (5.22), (5.23), and the fact that $K_i \le K_* 2^{n\tilde{r}_i}$, we see that for sufficiently large n,

$$\frac{1}{n}\log|\mathcal{U}_i^J|K_i \le C_{\text{in}}^i + \tilde{r}_i + \frac{1}{n}\log(1+n(R_1^*+R_2^*+2\delta)) < \tilde{C}_{\text{in}}^i$$
$$\frac{1}{n}\log|\mathcal{V}_i^J|K_* \le C_{\text{out}}^i + \frac{1}{n}\log(1+n(R_1^*+R_2^*+2\delta)) < \tilde{C}_{\text{out}}^i.$$

Thus the rate achieved by encoder i under an $(\tilde{r}_1, \tilde{r}_2)$ notion of error is at least as large as

$$\frac{1}{n}\log K_i^* L_i = \frac{1}{n}\log\left\lfloor\frac{K_i}{K_*}\right\rfloor\left\lfloor\frac{M_i}{K_i}\right\rfloor.$$

We next find a lower bound for the above expression. If $\tilde{r}_i < R_i$, then for sufficiently large n, $K_i = K_*\lfloor 2^{n\tilde{r}_i}\rfloor$, and the above quantity is at least as large as

$$\begin{aligned}
&\frac{1}{n}\log\left(2^{n\tilde{r}_i}-1\right)\left(\frac{1}{K_*}2^{n(R_i-\delta-\tilde{r}_i)}-1\right)\\
&\qquad\ge R_i - \delta + \frac{1}{n}\log\left(1-2^{-n\tilde{r}_i}\right)\left(\frac{1}{n(R_1^*+R_2^*+2\delta)+1} - 2^{-n(R_i-\delta-\tilde{r}_i)}\right)\\
&\qquad> R_i - 2\delta.
\end{aligned}$$

On the other hand, if $\tilde{r}_i \ge R_i$, then for sufficiently large n, $K_i \ge 2^{n(R_i-\delta)}$ for $i \in \{1,2\}$. Thus

$$\begin{aligned}
\frac{1}{n}\log\left\lfloor\frac{K_i}{K_*}\right\rfloor\left\lfloor\frac{M_i}{K_i}\right\rfloor &\ge \frac{1}{n}\log\left\lfloor\frac{K_i}{K_*}\right\rfloor\\
&\ge \frac{1}{n}\log\left(\frac{2^{n(R_i-\delta)}}{1+n(R_1^*+R_2^*+2\delta)} - 1\right)\\
&= R_i - \delta + \frac{1}{n}\log\left(\frac{1}{1+n(R_1^*+R_2^*+2\delta)} - 2^{-n(R_i-\delta)}\right)\\
&> R_i - 2\delta.
\end{aligned}$$

When either $\min\{\tilde{r}_1, \tilde{r}_2\} = 0$ or $\min\{R_1, R_2\} = 0$, we apply a similar argument, but instead of using Lemma 5.5.1, we use its corresponding vector version, which we state below.

Lemma 5.5.4 (Permutation Lemma – Vector Version)**.** *Let $A = (a_i)_{i=1}^m$ be a $(0,1)$-vector and let $N(A)$ denote the number of ones in A; that is,*

$$N(A) = \sum_{i=1}^m a_i.$$

Suppose k is a positive integer smaller or equal to m. For any permutation π on $[m]$ and $s \in [\frac{m}{k}]$, let $B_s(\pi)$ denote the vector

$$B_s(\pi) = \left(a_{\pi(i)}\right)_{i=(s-1)k+1}^{sk}.$$

If

$$N(A) < m\Big(1 - \frac{1}{k}\Big),$$

then there exists a permutation π such that for every $s \in [\frac{m}{k}]$, the vector $B_s(\pi)$ contains at least one zero.

5.6 Appendix: Characterization of Special Regions of $\mathbb{R}^k_{\geq 0}$

In Section 5.2, we require the following well-known result.

Lemma 5.6.1. *Let $\mathscr{C} \subseteq \mathbb{R}^2_{\geq 0}$ be non-empty, compact, convex, and closed under projections onto the axes, that is, if (x, y) is in $\mathscr{C}$, then so are $(x, 0)$ and $(0, y)$. Then*

$$\mathscr{C} = \Big\{(x, y) \in \mathbb{R}^2_{\geq 0} \Big| \forall \alpha \in [0, 1] : \alpha x + (1-\alpha) y \leq C^{\alpha}\Big\}.$$

For completeness, we state and prove an extension of this result to arbitrary dimensions here.

Let k be a positive integer and $\mathscr{C}$ be a compact subset of $\mathbb{R}^k_{\geq 0}$. In addition, let $\Delta_k \subseteq \mathbb{R}^k_{\geq 0}$ denote the k-dimensional probability simplex, that is, the set of all $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ in $\mathbb{R}^k_{\geq 0}$ such that $\sum_{j=1}^k \alpha_j = 1$. For every $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k) \in \Delta_k$, define $C^{\boldsymbol{\alpha}} \in \mathbb{R}_{\geq 0}$ as

$$C^{\boldsymbol{\alpha}} := \max_{\boldsymbol{x} \in \mathscr{C}} \boldsymbol{\alpha}^T \boldsymbol{x}.$$

For $j \in [k]$, define the projection $P_j : \mathbb{R}^k \to \mathbb{R}^k$ as

$$P_j(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_k) := (x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots, x_k).$$

In words, P_j sets the jth coordinate of its input to zero and leaves the other coordinates unchanged. We say a set $\mathscr{C} \subseteq \mathbb{R}^k$ is closed under projection P_j if and only if $P_j(\mathscr{C}) \subseteq \mathscr{C}$.

Lemma 5.6.2. *Let $\mathscr{C} \subseteq \mathbb{R}^k_{\geq 0}$ be non-empty, compact, convex, and closed under the projections $\{P_j\}_{j=1}^k$. Then*

$$\mathscr{C} = \Big\{\mathbf{x} \in \mathbb{R}^k_{\geq 0} \Big| \forall\, \boldsymbol{\alpha} \in \Delta_k \colon \boldsymbol{\alpha}^T \boldsymbol{x} \leq C^{\boldsymbol{\alpha}}\Big\}. \tag{5.28}$$

Proof. Let $\mathscr{C}'$ denote the set on the right hand side of (5.28). From the definition of $C^{\boldsymbol{\alpha}}$, it follows $\mathscr{C} \subseteq \mathscr{C}'$. Thus it suffices to show $\mathscr{C}' \subseteq \mathscr{C}$.

Every hyperplane in $\mathbb{R}^k$ divides $\mathbb{R}^k$ into two sets, each of which is referred to as a half-space. Since $\mathscr{C}$ is closed and convex, it equals the intersection of all the half-spaces containing it [62, p. 36]. Thus it suffices to show that if for some $\boldsymbol{\beta} = (\beta_j)_{j=1}^k \in \mathbb{R}^k$ and $\gamma \in \mathbb{R}$, $\mathscr{C}$ is a subset of the half-space

$$H := \left\{ \mathbf{x} \in \mathbb{R}^k \middle| \boldsymbol{\beta}^T \mathbf{x} \leq \gamma \right\},$$

then $\mathscr{C}'$ is also a subset of H. Suppose H contains $\mathscr{C}$. Since $\mathscr{C}$ is nonempty and closed under the projections $\{P_j\}_{j=1}^k$, $\mathscr{C}$ contains the origin. But $\mathscr{C} \subseteq H$, thus H contains the origin as well. This implies $\gamma \geq 0$. Let S be the set of all $j \in [k]$ such that $\beta_j > 0$. If S is empty, then H contains $\mathbb{R}^k_{\geq 0}$ and therefore, $\mathscr{C}'$. If S is nonempty, define $\boldsymbol{\alpha} = (\alpha_j)_{j \in [k]} \in \Delta_k$ as

$$\alpha_j := \begin{cases} \beta_j / \beta_S & \text{if } j \in S \\ 0 & \text{otherwise,} \end{cases}$$

where $\beta_S = \sum_{j \in S} \beta_j > 0$. From the definition of $C^{\boldsymbol{\alpha}}$, it follows that there exists $\mathbf{x} \in \mathscr{C}$ such that $\boldsymbol{\alpha}^T \mathbf{x} = C^{\boldsymbol{\alpha}}$, or equivalently,

$$\sum_{j \in S} \beta_j x_j = \beta_S C^{\boldsymbol{\alpha}}. \tag{5.29}$$

Since $\mathscr{C}$ is closed under the projections $\{P_j\}_{j=1}^k$, the vector $\mathbf{x}^* = (x_j^*)_{j \in [k]}$ is also in $\mathscr{C}$, where

$$x_j^* := \begin{cases} x_j & \text{if } j \in S \\ 0 & \text{otherwise.} \end{cases}$$

Using (5.29) and the fact that $\mathbf{x}^* \in \mathscr{C} \subseteq H$, we get

$$\beta_S C^{\boldsymbol{\alpha}} = \boldsymbol{\beta}^T \mathbf{x}^* \leq \gamma.$$

Now for every $\mathbf{x}' \in \mathscr{C}'$, we have

$$\boldsymbol{\beta}^T \mathbf{x}' = \sum_{j=1}^k \beta_j x_j' \leq \sum_{j \in S} \beta_j x_j' = \beta_S \boldsymbol{\alpha}^T \mathbf{x}' \leq \beta_S C^{\boldsymbol{\alpha}} \leq \gamma.$$

Thus $\mathscr{C}' \subseteq H$. Since H was an arbitrary half-space containing $\mathscr{C}$, it follows that $\mathscr{C}' \subseteq \mathscr{C}$. ∎

Chapter 6

CONCLUSION

When small networks are not studied in isolation, but rather viewed as part of larger networks, many opportunities arise for a more efficient utilization of resources. Consider the classical MAC; the presence of a single additional node significantly expands the coding strategies available to the encoders, which results in major improvements in rate and reliability. This simple example demonstrates not only the complexity but also the opportunity that lies ahead in the study of communication networks.

We conclude this work by discussing a number of problems that both extend the theoretical reach of this work and bring it closer to implementation.

Low-Complexity Coding for Cooperation. While random coding arguments are very useful for obtaining inner bounds on the capacity region of a network, their direct application is not possible in practice. The reason is that randomly-generated codes have a high computational cost: both in their construction and in their encoding and decoding processes. Thus, to ensure that the cooperative gains discussed here are attainable in practice, we need low-complexity codes. Polar codes [63] may be well-suited for this purpose, for reasons which we next discuss.

Recall that our coding strategy in Chapter 3 combines three schemes: forwarding, coordination (Marton coding), and classical MAC coding. Due to the simplicity of the forwarding strategy, the main problem lies in obtaining a low-complexity implementation of Marton coding and classical MAC coding. Both of these strategies, however, are implemented using polar codes in [64], [65], and [66], respectively. It remains to be seen whether these strategies can be successfully combined to obtain the large sum-capacity gains described in Chapter 3.

Ongoing Cooperation in the Presence of Causal State Information. In the absence of cooperation, assuming that the state process is i.i.d. and the decoder has full state information, the MAC capacity region is the same regardless of whether the state information at the encoders is causal or non-causal. The open question here is whether the same result holds in the presence of cooperation. Note that progress on this problem requires an extension of our cooperation model; this is due to the fact that when causal state information is available at the encoders, the communication between the CF and the encoders has to occur during the same time that the encoders transmit their codewords over the channel. We refer to this extended cooperation model as "ongoing cooperation."

The Weak Edge Removal Property in Noisy Networks. In Chapter 5, we demonstrate the existence of a network that does not satisfy the weak edge removal property with respect to the maximal-error capacity region. Whether such a network exists with respect to the average-error capacity region remains open. The result of Sarwate and Gastpar [56] regarding negligible feedback and our result in Appendix C provide evidence against the existence of such a network.

Appendix A

THE MULTIVARIATE COVERING LEMMA

The covering lemma and its extensions play a crucial role in achievability results in network information theory. Covering lemmas are useful for enabling network nodes to transmit codewords that "look like" they are generated from a dependent distribution, whereas in reality, they are carefully selected from sufficiently large codebooks that are independently generated. This allows nodes to obtain the benefits of both independent and dependent codewords: like independent codewords, such codewords can be decoded in different locations; like dependent codewords, they have the potential to achieve rates higher than those achieved by independent codewords. This benefit, however, comes at a cost in rate. Thus this strategy is useful when the benefit of transmitting dependent codewords exceeds its cost.

In the context of the covering lemma, the concept of "looking like" dependent codewords is captured by the notion of being jointly typical with respect to a dependent distribution. As there are various ways to define the typical set (here we specifically focus on weakly typical [40] and strongly typical sets [38]), one may ask whether a specific version of the covering lemma holds for a given definition of the typical set. The weakly typical set has two advantages over the strongly typical set. First, it is easily defined for continuous (e.g., Gaussian) distributions. Second, the weakly typical set has a simple one-shot counterpart, which allows proofs using the weakly typical set to be written in the one-shot framework in a simple manner. On the other hand, the strongly typical set has properties that the weakly typical set does not possess. Thus it is helpful to review the covering lemma and its extensions and see for which definitions of the typical set each result is currently known to hold.

The simplest case of the covering lemma is the situation where, given a random vector and an independently generated codebook, a node looks for a codeword in the codebook that is jointly typical (with respect to a dependent distribution) with the given random vector. The result obtained in this case, simply referred to as the "covering lemma," appears in the achievability proof of the rate distortion theorem using weakly typical sets [40]. The second case, called the “mutual covering lemma,” treats the case where given two independently generated codebooks, a node looks for a jointly typical pair of codewords, where each codeword is from one of the codebooks. This result is used in Marton’s inner bound for the two-user broadcast channel and is proved for strongly typical sets [8], [37]. Recently, by extending the proof of [40], the authors of [67], [68] prove a one-shot version of the mutual covering lemma. This proof can be used to show the validity of the mutual covering lemma for weakly typical sets in the asymptotic setting. The proof in [67], [68], however, requires stronger independence assumptions on the codebooks than the proof using strongly typical sets in [37], [38]. Finally, the “multivariate covering lemma” is the extension of the mutual covering lemma to k independently generated codebooks for arbitrary $k \geq 2$, and can be used to obtain an inner bound on the broadcast channel with k users [38]. As stated in [38], one can show this result holds for strongly typical sets by extending the proof of the mutual covering lemma [37].

In this work, using the general strategy of El Gamal and Van der Meulen [37] and some ideas regarding weakly typical sets from Koetter, Effros, and Médard [39], we prove an extension of the multivariate covering lemma for weakly typical sets. This extension is motivated by cooperative strategies where nodes share both message and state information as in Chapter 4.

We also provide a converse, a special case of which is usually referred to as the “packing lemma” [38]. We remark that while similar to [37], we use Chebyshev’s inequality to prove the direct result in Section A.3, it is also possible to use the Cauchy-Schwarz inequality, which leads to a more accurate upper bound in this case (Section A.5).

A.1 Problem Statement

Let $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$ be a finite collection of sets. For every nonempty $S \subseteq \mathcal{I}$, define $\mathcal{A}_S$ as

$$\mathcal{A}_S := \prod_{i\in S} \mathcal{A}_i.$$

An element of $\mathcal{A}_S$ is denoted with $a_S := (a_i)_{i\in S}$.

For every positive integer n, define $[n] := \{1, \dots, n\}$. Fix a positive integer k. Consider the sets $(\mathcal{U}_i)_{i\in[k+1]}$ and $(\mathcal{V}_j)_{j\in[k]}$. We study discrete and continuous settings. In the discrete case, these sets are all finite or countably infinite, and $p(u_{[k+1]}, v_{[k]})$ is a probability mass function on $\mathcal{U}_{[k+1]} \times \mathcal{V}_{[k]}$. In the continuous case, $p(u_{[k+1]}, v_{[k]})$ is a probability density function on $\mathbb{R}^{2k+1}$, and the sets $\mathcal{U}_i$ and $\mathcal{V}_j$ represent the support of the marginals $p(u_i)$ and $p(v_j)$, respectively.

For every $j \in [k]$, let M_j be a nonnegative integer. For nonempty $S \subseteq [k]$, define the set $\mathcal{M}_S$ as

$$\mathcal{M}_S := \prod_{j \in S} [M_j],$$

and let $\mathcal{M} := \mathcal{M}_{[k]}$. For every $m_{[k]} = (m_1, \ldots, m_k) \in \mathcal{M}$, let the random vector

$$\big(U_{[k+1]}, V_1(m_1), \ldots, V_k(m_k)\big)$$

have distribution $p_{\text{ind}}(u_{[k+1]}, v_{[k]})$, where

$$p_{\text{ind}}(u_{[k+1]}, v_{[k]}) := p(u_{[k+1]}) \prod_{j \in [k]} p(v_j | u_j). \tag{A.1}$$

In (A.1), $p(u_{[k+1]})$ and each $p(v_j|u_j)$ are calculated from our original distribution $p(u_{[k+1]}, v_{[k]})$.

Let $\mathcal{F}$ be an arbitrary subset of $\mathcal{U}_{[k+1]} \times \mathcal{V}_{[k]}$. For every $S \subseteq [k]$, define $\mathcal{F}_S$ as the projection of $\mathcal{F}$ on $\mathcal{U}_{[k+1]} \times \mathcal{V}_S$. For $(u_{[k+1]}, v_S) \in \mathcal{F}_S$, let $\mathcal{F}(u_{[k+1]}, v_S)$ be the set of all v_{S^c} such that $(u_{[k+1]}, v_{[k]}) \in \mathcal{F}$, where $S^c := [k] \setminus S$. Furthermore, assume that for every nonempty $S \subseteq [k]$, there exist real numbers α_S and β_S such that

$$\forall\, (u_{[k+1]}, v_S) \in \mathcal{F}_S \colon \alpha_S \leq \log \frac{p(v_S | u_{[k+1]})}{p_{\text{ind}}(v_S | u_S)} \tag{A.2}$$

$$\forall\, (u_{[k+1]}, v_{[k]}) \in \mathcal{F} \colon \beta_S \leq \log \frac{p(v_S | u_{[k+1]}, v_{S^c})}{p_{\text{ind}}(v_S | u_S)}. \tag{A.3}$$

Furthermore, assume that there exists a constant γ such that

$$\forall\, (u_{[k+1]}, v_{[k]}) \in \mathcal{F} \colon \gamma \geq \log \frac{p(v_{[k]} | u_{[k+1]})}{p_{\text{ind}}(v_{[k]} | u_{[k]})}. \tag{A.4}$$

Our aim is to find upper and lower bounds on the probability

$$\Pr\Big\{\forall\, m_{[k]} \in \mathcal{M} \colon \big(U_{[k+1]}, V_1(m_1), \ldots, V_k(m_k)\big) \notin \mathcal{F}\Big\}.$$

We derive the lower bound in Section A.2 using the union bound, which does not depend on the statistical dependencies of the vectors

$$\big(U_{[k+1]}, V_1(m_1), \ldots, V_k(m_k)\big)$$

for different values of $m_{[k]}$. For the upper bound, given in Subsection A.3, which leads to the multivariate covering lemma, we require a stronger assumption, which we next describe.

Let $m_{[k]}$ and $m'_{[k]}$ be in $\mathcal{M}$. Define the set $S(m_{[k]}, m'_{[k]})$ as

$$S(m_{[k]}, m'_{[k]}) := \left\{ j \in [k] \middle| m_j = m'_j \right\}.$$

When $m_{[k]}$ and $m'_{[k]}$ are clear from context, we denote $S(m_{[k]}, m'_{[k]})$ with S. Recall that $S^c = [k] \setminus S$.

In the proof of the upper bound, we require

Assumption I. For all $(m_{[k]}, m'_{[k]})$ and $(u_{[k+1]}, v_{[k]}, v'_{S^c})$,

$$\begin{aligned} &\Pr\left\{ \forall\, j \in [k] \colon V_j(m_j) = v_j \text{ and } V_j(m'_j) = v'_j \middle| U_{[k+1]} = u_{[k+1]} \right\} \\ &\quad = \prod_{j \in [k]} p(v_j|u_j) \times \prod_{j \in S^c} p(v'_j|u_j). \end{aligned}$$

In the corresponding asymptotic problem, which we study in Section A.4, we apply our bounds to

$$\Pr\left\{ \forall\, m_{[k]} \colon \left(U^n_{[k+1]}, V^n_1(m_1), \ldots, V^n_k(m_k)\right) \notin A^{(n)}_\delta \right\},$$

where the probability is calculated according to the distribution defined by the assumption below.

Assumption II. For every $m_{[k]} \in \mathcal{M}$,

$$\left(U^n_{[k+1]}, V^n_1(m_1), \ldots, V^n_k(m_k)\right) \overset{\text{iid}}{\sim} p_{\text{ind}}(u_{[k+1]}, v_{[k]}).$$

Finally, we need $A^{(n)}_\delta$, the weakly typical set [40, p. 521] defined with respect to the distribution $p(u_{[k+1]}, v_{[k]})$, to satisfy the requirements posed by (A.2), (A.3), and (A.4). This is guaranteed by the next assumption.

Assumption III. The distribution $p(u_{[k+1]}, v_{[k]})$ has the property that for all nonempty $S \subseteq [k]$, the conditional marginal distributions $p(v_S|u_{[k+1]})$, $p(v_S|u_{[k+1]}, v_{S^c})$, and $\{p(v_j|u_j)\}_{j\in[k]}$ are well defined and have finite entropy.

Note that this assumption holds trivially for probability mass functions with finite support.

The multivariate covering lemma follows.

Lemma A.1.1 (Multivariate Covering Lemma)**.** *Suppose assumptions (I-III) hold for the joint distribution of*

$$U^n_{[k+1]}, \{V^n_1(m_1), \ldots, V^n_k(m_k)\}_{m_{[k]}}.$$

For the direct part, suppose for all $j \in [k]$, $M_j \geq 2^{nR_j}$. If for all nonempty $S \subseteq [k]$,

$$\sum_{j\in S} R_j > \sum_{j \in S} H(V_j|U_j) - H(V_S|U_{[k+1]}) + (8k - 2|S| + 10)\delta, \tag{A.5}$$

then

$$\lim_{n\to\infty} \Pr\left\{\exists\, m_{[k]}\colon \left(U^n_{[k+1]}, V^n_1(m_1), \ldots, V^n_k(m_k)\right) \in A^{(n)}_\delta\right\} = 1. \tag{A.6}$$

For the converse, assume for all $j \in [k]$, $M_j \leq 2^{nR_j}$. If (A.6) holds, then

$$\sum_{j\in S} R_j \geq \sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]}) - 2(|S| + 1)\delta,$$

for all nonempty $S \subseteq [k]$.

Remark. In the direct part of Lemma A.1.1, we can weaken the lower bound on $\sum_{j\in S} R_j$ when $S = [k]$. Specifically, for $S = [k]$, we can replace (A.5) with

$$\sum_{j\in[k]} R_j > \sum_{j\in[k]} H(V_j|U_j) - H(V_{[k]}|U_{[k+1]}) + 2(k + 1)\delta.$$

A.2 Lower Bound

For every $m_{[k]} = (m_1, \ldots, m_k) \in \mathcal{M}$, define the random variable $Z(m_{[k]})$ as

$$Z(m_{[k]}) := \mathbf{1}\Big\{\big(U_{[k+1]}, V_1(m_1), \ldots, V_k(m_k)\big) \in \mathcal{F}\Big\}$$

and set

$$Z := \sum_{m_{[k]}\in\mathcal{M}} Z(m_{[k]}).$$

Our aim is to find a lower bound for $\Pr\{Z = 0\}$. Note that for every nonempty $S \subseteq [k]$,

$$\begin{aligned}
\Pr\left\{\exists\, m_{[k]}\colon Z(m_{[k]}) = 1\right\} &= \Pr\left\{\exists\, m_{[k]}\colon \left(U_{[k+1]}, V_1(m_1), \ldots, V_k(m_k)\right) \in \mathcal{F}\right\} \\
&\leq \Pr\left\{\exists\, m_S\colon \left(U_{[k+1]}, \left(V_j(m_j)\right)_{j\in S}\right) \in \mathcal{F}_S\right\} \\
&\leq |\mathcal{M}_S| \sum_{\mathcal{F}_S} p(u_{[k+1]}) p_{\text{ind}}(v_S|u_S) \\
&\leq |\mathcal{M}_S| 2^{-\alpha_S} \sum_{\mathcal{F}_S} p(u_{[k+1]}, v_S) \\
&\leq |\mathcal{M}_S| 2^{-\alpha_S}.
\end{aligned}$$

Thus

$$\Pr\{Z=0\} = 1 - \Pr\left\{\exists\, m_{[k]}\colon Z(m_{[k]}) = 1\right\}$$
$$\geq 1 - \min_{|S|\neq\emptyset} |\mathcal{M}_S| 2^{-\alpha_S}. \tag{A.7}$$

A.3 Upper Bound

In deriving our upper bound on $\Pr\{Z=0\}$, we apply conditioning and Chebyshev's inequality, which lead to the appearance of the factor

$$\frac{1}{\left(\Pr\{\mathcal{F}(u_{[k+1]})\}\right)^2},$$

where

$$\Pr\left\{\mathcal{F}(u_{[k+1]})\right\} := \Pr\left\{V_{[k]} \in \mathcal{F}(u_{[k+1]}) \middle| U_{[k+1]} = u_{[k+1]}\right\}$$
$$= \sum_{v_{[k]}\in\mathcal{F}(u_{[k+1]})} p(v_{[k]}|u_{[k+1]}).$$

Recall that $\mathcal{F}(u_{[k+1]})$, defined in Section A.2, is the set of all $v_{[k]} \in \mathcal{V}_{[k]}$ such that $(u_{[k+1]}, v_{[k]}) \in \mathcal{F}$. Thus to get an accurate upper bound, we require $\Pr\{\mathcal{F}(u_{[k+1]})\}$ to be large. However, as this cannot be guaranteed for all $u_{[k+1]}$, we partition $\mathcal{U}_{[k+1]}$ into "good" and "bad" sets, corresponding to large and small values of $\Pr\{\mathcal{F}(u_{[k+1]})\}$, respectively. We show that the probability of the good set is large whenever $\Pr\{(U_{[k+1]}, V_{[k]}) \in \mathcal{F}\}$ is large. Following [39, Appendix III], fix $\epsilon > 0$ and define $\mathcal{G}_\epsilon \subseteq \mathcal{U}_{[k+1]}$ as

$$\mathcal{G}_\epsilon := \left\{u_{[k+1]} \middle| \Pr\{\mathcal{F}(u_{[k+1]})\} \geq 1-\epsilon\right\}.$$

Note that $\mathcal{G}_\epsilon$ represents the set of all good $u_{[k+1]}$ as described above. We have

$$\Pr\left\{(U_{[k+1]}, V_{[k]}) \in \mathcal{F}\right\} = \sum_{u_{[k+1]}} p(u_{[k+1]}) \Pr\{\mathcal{F}(u_{[k+1]})\}$$
$$\leq (1-\epsilon)\Pr\{U_{[k+1]} \notin \mathcal{G}_\epsilon\} + \Pr\{U_{[k+1]} \in \mathcal{G}_\epsilon\}$$
$$= 1 - \epsilon \Pr\{U_{[k+1]} \notin \mathcal{G}_\epsilon\}.$$

Thus

$$\Pr\{U_{[k+1]} \notin \mathcal{G}_\epsilon\} \leq \frac{1}{\epsilon} \Pr\left\{(U_{[k+1]}, V_{[k]}) \notin \mathcal{F}\right\}. \tag{A.8}$$

Our aim is to find an upper bound for $\Pr\{Z=0\}$. To do this, we write

$$\begin{aligned}\Pr\{Z=0\} &= \sum_{u_{[k+1]}} p(u_{[k+1]}) \Pr\left\{Z=0\middle|u_{[k+1]}\right\} \\ &\le \frac{1}{\epsilon}\Pr\left\{(U_{[k+1]},V_{[k]})\notin\mathcal{F}\right\} + \sum_{u_{[k+1]}\in\mathcal{G}_\epsilon} p(u_{[k+1]}) \Pr\left\{Z=0\middle|u_{[k+1]}\right\},\end{aligned} \tag{A.9}$$

where the inequality follows from (A.8). Therefore, to find an upper bound on $\Pr\{Z=0\}$, it suffices to find an upper bound on $\Pr\{Z=0|u_{[k+1]}\}$ for all $u_{[k+1]}\in\mathcal{G}_\epsilon$.

Fix $u_{[k+1]}\in\mathcal{G}_\epsilon$. We use Chebyshev's inequality to find an upper bound on $\Pr\{Z=0|u_{[k+1]}\}$. Thus we need to calculate $\mathbb{E}[Z|u_{[k+1]}]$ and $\mathbb{E}[Z^2|u_{[k+1]}]$. For a given $m_{[k]}$, from (A.4), it follows that

$$\begin{aligned}\mathbb{E}\big[Z(m_{[k]})\big|u_{[k+1]}\big] &= \Pr\left\{\big(V_1(m_1),\dots,V_k(m_k)\big)\in\mathcal{F}(u_{[k+1]})\middle|u_{[k+1]}\right\} \\ &= \sum_{\mathcal{F}(u_{[k+1]})} p_{\text{ind}}(v_{[k]}|u_{[k+1]}) \\ &\ge \sum_{\mathcal{F}(u_{[k+1]})} 2^{-\gamma} p(v_{[k]}|u_{[k+1]}) \\ &= 2^{-\gamma}\Pr\{\mathcal{F}(u_{[k+1]})\} \ge (1-\epsilon)2^{-\gamma},\end{aligned}$$

where the last inequality follows from the fact that $u_{[k+1]}\in\mathcal{G}_\epsilon$. Thus, by linearity of expectation,

$$\mathbb{E}[Z|u_{[k+1]}] \ge |\mathcal{M}|2^{-\gamma}(1-\epsilon). \tag{A.10}$$

Next, we find an upper bound on $\mathbb{E}[Z^2|u_{[k+1]}]$. We have

$$Z^2 = \sum_{m_{[k]}}[Z(m_{[k]})]^2 + \sum_{m_{[k]}\ne m'_{[k]}} Z(m_{[k]})Z(m'_{[k]}) = Z + \sum_{m_{[k]}\ne m'_{[k]}} Z(m_{[k]})Z(m'_{[k]}),$$

since $[Z(m_{[k]})]^2 = Z(m_{[k]})$ and $Z=\sum_{m_{[k]}} Z(m_{[k]})$. Thus

$$\mathbb{E}[Z^2|u_{[k+1]}] = \mathbb{E}[Z|u_{[k+1]}] + \mathbb{E}\Big[\sum_{m_{[k]}\ne m'_{[k]}} Z(m_{[k]})Z(m'_{[k]})\Big|u_{[k+1]}\Big].$$

For any pair of distinct $m_{[k]}$ and $m'_{[k]}$ with nonempty $S := S(m_{[k]}, m'_{[k]})$, we have

$$\begin{aligned}
&\mathbb{E}\Big[Z(m_{[k]})Z(m'_{[k]})\Big|u_{[k+1]}\Big] \\
&\quad= \sum_{\mathcal{F}_S(u_{[k+1]})} p_{\text{ind}}(v_S|u_S)\Bigg[\sum_{v_{S^c}\in\mathcal{F}(u_{[k+1]},v_S)} p_{\text{ind}}(v_{S^c}|u_{S^c})\Bigg]^2 \\
&\quad\le 2^{-\alpha_S-2\beta_{S^c}} \sum_{\mathcal{F}_S(u_{[k+1]})} p(v_S|u_{[k+1]})\Bigg[\sum_{v_{S^c}\in\mathcal{F}(u_{[k+1]},v_S)} p(v_{S^c}|u_{[k+1]},v_S)\Bigg]^2 \\
&\quad\le 2^{-\alpha_S-2\beta_{S^c}},
\end{aligned}$$

where $\mathcal{F}_S(u_{[k+1]})$ is the set of all v_S where $(u_{[k+1]}, v_S) \in \mathcal{F}_S$. On the other hand, if S is empty, then $Z(m_{[k]})$ and $Z(m'_{[k]})$ are independent given $U_{[k+1]} = u_{[k+1]}$, and

$$\mathbb{E}\Big[Z(m_{[k]})Z(m'_{[k]})\Big|u_{[k+1]}\Big] = \Big(\mathbb{E}\big[Z(m_{[k]})\big|u_{[k+1]}\big]\Big)^2.$$

Fix $m_{[k]} \in \mathcal{M}$. For every nonempty $S \subseteq [k]$, let $m_{[k]}(S)$ be an element of $\mathcal{M}$ such that $m_j = m_j(S)$ if and only if $j \in S$. Thus

$$\begin{aligned}
\mathbb{E}\big[Z^2\big|u_{[k+1]}\big] &= \mathbb{E}\big[Z\big|u_{[k+1]}\big] + \Big[\prod_{j\in[k]}\big(|\mathcal{M}_j|^2 - |\mathcal{M}_j|\big)\Big]\Big(\mathbb{E}\big[Z(m_{[k]})\big|u_{[k+1]}\big]\Big)^2 \\
&\quad+ \sum_{\emptyset\subsetneq S\subsetneq[k]} |\mathcal{M}_S|\Big[\prod_{j\in S^c}\big(|\mathcal{M}_j|^2 - |\mathcal{M}_j|\big)\Big]\mathbb{E}\Big[Z(m_{[k]})Z(m_{[k]}(S))\Big|u_{[k+1]}\Big] \\
&\le \mathbb{E}\big[Z\big|u_{[k+1]}\big] + \Big(\mathbb{E}\big[Z\big|u_{[k+1]}\big]\Big)^2 + \sum_{\emptyset\subsetneq S\subsetneq[k]} |\mathcal{M}_S||\mathcal{M}_{S^c}|^2 2^{-\alpha_S-2\beta_{S^c}}.
\end{aligned} \tag{A.11}$$

Thus for all $u_{[k+1]} \in \mathcal{G}_\epsilon$, we have

$$\begin{aligned}
\Pr\big\{Z = 0\big|u_{[k+1]}\big\} &\le \Pr\Big\{\big|Z - \mathbb{E}[Z|u_{[k+1]}]\big| \ge \mathbb{E}[Z|u_{[k+1]}]\Big|u_{[k+1]}\Big\} \\
&\overset{(a)}{\le} \frac{\text{Var}(Z|u_{[k+1]})}{\big(\mathbb{E}[Z|u_{[k+1]}]\big)^2} = \frac{\mathbb{E}[Z^2|u_{[k+1]}]}{\big(\mathbb{E}[Z|u_{[k+1]}]\big)^2} - 1 \\
&\overset{(b)}{\le} \frac{1}{1-\epsilon}|\mathcal{M}|^{-1}2^{\gamma} + \frac{1}{(1-\epsilon)^2}\sum_{\emptyset\subset S\subset[k]} |\mathcal{M}_S|^{-1}2^{-\alpha_S-2\beta_{S^c}+2\gamma},
\end{aligned}$$

where (a) follows from Chebyshev's inequality and (b) follows from (A.10) and (A.11). Now using (A.9), we get

$$\Pr\{Z = 0\} \le \frac{1}{\epsilon}\Pr\{\mathcal{F}^c\} + \frac{1}{1-\epsilon}|\mathcal{M}|^{-1}2^{\gamma} + \frac{1}{(1-\epsilon)^2}\sum_{\emptyset\subset S\subset[k]} |\mathcal{M}_S|^{-1}2^{-\alpha_S-2\beta_{S^c}+2\gamma}, \tag{A.12}$$

where

$$\Pr\{\mathcal{F}^c\} := \Pr\left\{(U_{[k+1]}, V_{[k]}) \notin \mathcal{F}\right\}.$$

A.4 Asymptotic Result

In this section, using our lower and upper bounds, we prove Lemma A.1.1. We first prove the direct part using our upper bound from Section A.3. For every positive integer n, set $\mathcal{F}^{(n)} = A_\delta^{(n)}(U_{[k+1]}, V_{[k]})$, and for every $j \in [k]$, choose an integer $M_j \geq 2^{nR_j}$. Furthermore, fix $\epsilon \in (0, 1)$.

For every nonempty $S \subseteq [k]$, notice that if $\left(u_{[k+1]}^n, v_S^n\right) \in \mathcal{F}_S^{(n)}$, then

$$\Big|\log \frac{p(v_S^n|u_{[k+1]}^n)}{\prod_{j\in S} p(v_j^n|u_j^n)} - n\Big(\sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]})\Big)\Big| \leq 2n(|S|+1)\delta.$$

Thus we can choose

$$\alpha_S^{(n)} := n\Big(\sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]}) - 2(|S|+1)\delta\Big)$$
$$\gamma^{(n)} := n\Big(\sum_{j\in [k]} H(V_j|U_j) - H(V_{[k]}|U_{[k+1]}) + 2(k+1)\delta\Big).$$

Similarly, for every nonempty $S \subseteq [k]$, we can set

$$\beta_S^{(n)} := n\Big(\sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]}, V_{S^c}) - 2(|S|+1)\delta)\Big),$$

since for every $\left(u_{[k+1]}^n, v_S^n, v_{S^c}^n\right) \in \mathcal{F}^{(n)}$,

$$\Big|\log \frac{p(v_S^n|u_{[k+1]}^n, v_{S^c}^n)}{\prod_{j\in S} p(v_j^n|u_j^n)} - n\Big(\sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]}, V_{S^c}])\Big)\Big| \leq 2n(|S|+1)\delta.$$

From our upper bound, Equation (A.12), it now follows that if for all nonempty $S \subsetneq [k]$,

$$\begin{aligned}
\sum_{j\in S} R_j &> \frac{1}{n}(2\gamma - \alpha_S - 2\beta_{S^c}) \\
&= 2\sum_{j\in[k]} H(V_j|U_j) - 2H(V_{[k]}|U_{[k+1]}) - \sum_{j\in S} H(V_j|U_j) + H(V_S|U_{[k+1]}) \\
&\quad - 2\sum_{j\in S^c} H(V_j|U_j) + 2H(V_{S^c}|U_{[k+1]}, V_S) + (8k - 2|S| + 10)\delta \\
&= \sum_{j\in S} H(V_j|U_j) - H(V_S|U_{[k+1]}) + (8k - 2|S| + 10)\delta,
\end{aligned}$$

and for $S = [k]$,

$$\sum_{j \in [k]} R_j > \frac{1}{n}\gamma = \sum_{j \in [k]} H(V_j|U_j) - H(V_{[k]}|U_{[k+1]}) + 2(k+1)\delta,$$

then

$$\lim_{n\to\infty} \Pr\left\{\exists m_{[k]} : \left(U^n_{[k+1]}, V^n_1(m_1), \ldots, V^n_k(m_k)\right) \in A^{(n)}_\delta\right\} = 1. \tag{A.13}$$

Next we prove the converse. Suppose for each $j \in [k]$, $M_j \le 2^{nR_j}$ and (A.13) holds. Then from (A.7), it follows that

$$\sum_{j \in S} R_j \ge \frac{1}{n}\alpha_S = \sum_{j \in S} H(V_j|U_j) - H(V_S|U_{[k+1]}) - 2(|S|+1)\delta,$$

for all nonempty $S \subseteq [k]$.

A.5 Cauchy-Schwarz versus Chebyshev

Let Z be any random variable that is nonnegative with probability one and has positive first and second moments. Then

$$Z = Z\mathbf{1}\{Z > 0\}$$

almost surely. Thus by the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}\big[Z\mathbf{1}\{Z>0\}\big] \\ &\le \sqrt{\mathbb{E}[Z^2] \times \Pr\{Z>0\}}.\end{aligned}$$

Hence

$$\Pr\{Z > 0\} \ge \frac{\big(\mathbb{E}[Z]\big)^2}{\mathbb{E}[Z^2]}$$

and

$$\Pr\{Z = 0\} \le 1 - \frac{\big(\mathbb{E}[Z]\big)^2}{\mathbb{E}[Z^2]}.$$

On the other hand, using Chebyshev's inequality we get

$$\begin{aligned}\Pr\{Z = 0\} &= \Pr\big\{|Z - \mathbb{E}[Z]| \ge \mathbb{E}[Z]\big\} \\ &\le \frac{\mathrm{Var}(Z)}{\big(\mathbb{E}[Z]\big)^2} = \frac{\mathbb{E}[Z^2]}{\big(\mathbb{E}[Z]\big)^2} - 1.\end{aligned}$$

Note that the bound resulting from Cauchy-Schwarz is stronger, since for any $t > 0$,

$$1 - t \le \frac{1}{t} - 1.$$

Appendix B

LARGE DEVIATIONS

In this appendix, we prove that the probability that an i.i.d. sequence is weakly typical converges to one exponentially fast. It is proved for a modified version of weakly typical sets in [31, p. 991] and for strongly typical sets in [23, p. 117].

B.1 Result

Lemma B.1.1. *Choose a distribution $p(u_{[k]})$ on the alphabet $\mathcal{U}_{[k]}$, which may be continuous or discrete. Suppose there exists $t_0 > 0$ such that for all nonempty $S \subseteq [k]$ and $t \in (-t_0, t_0)$,*

$$\mathbb{E}\big[p(U_S)^{-t}\big] < \infty.$$

Then there exists a nondecreasing function $\underline{\theta}(U_{[k]}, \cdot) : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that for all sufficiently large n,

$$\Pr\left\{A_\epsilon^{(n)}(U_{[k]})\right\} \geq 1 - 2^{-n\underline{\theta}(U_{[k]},\epsilon)}.$$

Proof. The base-2 moment generating function of a random variable X is defined as

$$M(t) = \mathbb{E}[2^{tX}]$$

for all real t for which the expectation on the right hand side is finite. If M is defined on a neighborhood of 0, say $(-t_1, t_1)$ for some $t_1 > 0$, then it has a Taylor series expansion with a positive radius of convergence [69, pp. 278-280]. In particular,

$$\frac{d}{dt}M(t)\Big|_{t=0} = \frac{1}{\log e}\mathbb{E}[X].$$

We next find an upper bound for $\Pr\{X \geq a\}$ for any $a \in \mathbb{R}$. Choose $t \in (0, t_1)$. Using Markov's inequality, we get

$$\begin{aligned}\Pr\{X \geq a\} &= \Pr\{tX \geq ta\} \\ &= \Pr\{2^{tX} \geq 2^{ta}\} \\ &\leq 2^{-ta}\mathbb{E}[2^{tX}] \\ &= 2^{\log M(t) - ta}.\end{aligned}$$

Since $t \in (0, t_1)$ was arbitrary, we get

$$\Pr\{X \geq a\} \leq 2^{\inf_{t \in (0,t_1)} (\log M(t) - ta)}. \tag{B.1}$$

For all $\epsilon \in \mathbb{R}$, define $\theta(X, \epsilon)$ as

$$\theta(X, \epsilon) := \sup_{t \in (0,t_1)} \Big[t\big(\mathbb{E}[X] + \epsilon\big) - \log M(t) \Big].$$

Define the function f as

$$f(t) = \log M(t) - ta.$$

We have $f(0) = 0$ and $f'(0) = \mathbb{E}[X] - a$. Thus if $a > \mathbb{E}[X]$,

$$\inf_{t \in (0,t_1)} \big(\log M(t) - ta \big) < 0. \tag{B.2}$$

Similarly, if $\epsilon > 0$, then $\theta(X, \epsilon) > 0$. If we apply (B.1) to the random variable

$$\frac{1}{n} \sum_{i=1}^{n} X_i,$$

where the X_i's are i.i.d. copies of X, we get

$$\Pr\Big\{ \sum_{i=1}^{n} X_i \geq na \Big\} \leq 2^{-n\theta(X, a - \mathbb{E}[X])}. \tag{B.3}$$

Now consider a random vector $(U_1, \ldots, U_k)$ with distribution $p(u_1, \ldots, u_k)$. For every nonempty $S \subseteq [k]$, let U_S denote the random vector $(U_j)_{j \in S}$. Let $(U_1^n, \ldots, U_k^n)$ be n i.i.d. copies of $(U_1, \ldots, U_k)$. By applying inequality (B.3) to the random variables

$$\Big\{ \log \frac{1}{p(U_{Si})} \Big\}_{i=1}^{n}$$

and setting $a = H(U_S) + \epsilon$ for some $\epsilon > 0$, we get

$$\Pr\left\{ \sum_{i=1}^{n} \log \frac{1}{p(U_{Si})} \geq n\big[H(U_S) + \epsilon\big] \right\} \leq 2^{-n\theta(U_S, \epsilon)}. \tag{B.4}$$

Let

$$\underline{\theta}(U_{[k]}, \epsilon) = \frac{1}{2} \min_{\varnothing \subsetneq S \subseteq [k]} \theta(U_S, \epsilon).$$

By the union bound,

$$\begin{aligned}\Pr\left\{(U_1^n,\ldots,U_k^n)\notin A_\epsilon^{(n)}(U_1,\ldots,U_k)\right\} &\le 2\sum_{\varnothing\subsetneq S\subseteq[k]} 2^{-n\theta(U_S,\epsilon)}\\ &\le 2(2^k-1)2^{-n\min_{\emptyset\subsetneq S\subsetneq[k]}\theta(U_S,\epsilon)}\\ &\le 2^{-n\underline{\theta}(U_{[k]},\epsilon)},\end{aligned}$$

where the last inequality holds for all sufficiently large n. Finally, note that since by (B.2), each $\theta(U_S,\epsilon)$ is positive and nondecreasing, so is $\underline{\theta}(U_{[k]},\epsilon)$. □

Appendix C

CONTINUITY OF AVERAGE-ERROR SUM-CAPACITY

In this appendix, we show that the average-error sum-capacity of a network consisting of a MAC and a CF that has access to both messages is continuous with respect to the CF output edge capacities.

C.1 Model and Result

Consider a MAC

$$\big(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1,x_2), \mathcal{Y}\big),$$

where $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ are finite sets. We assume this MAC is memoryless and stationary so that its nth extension is given by

$$p(y^n|x_1^n,x_2^n) := \prod_{t\in[n]} p(y_t|x_{1t},x_{2t}).$$

We next state the main result of this appendix. In this theorem, for a given MAC, $\mathbf{C}_{\text{in}}^*$ is any pair with sufficiently large components so that any $(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$-CF has access to both messages. Our proof relies on ideas developed by Dueck [70].

Theorem C.1.1. *For any finite alphabet MAC,* $C_{\text{sum,avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ *is continuous on* $\mathbb{R}^2_{\geq 0}$.

Note that by Lemma 5.2.3, $C_{\text{sum,avg}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ is continuous on $\mathbb{R}^2_{>0}$. Thus it suffices to prove Theorem C.1.1 on the boundary of $\mathbb{R}^2_{\geq 0}$.[1]

As we are not concerned with maximal-error sum-capacity in this appendix, we drop the "avg" subscript and henceforth write the average-error sum-capacity as $C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$.

We begin with a number of definitions that are useful for the description of our lower and upper bounds for $C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$.

For every finite alphabet $\mathcal{U}$ and all $\delta \geq 0$, define

$$\mathcal{P}_{\mathcal{U}}^{(n)}(\delta) := \Big\{ p(u,x_1^n,x_2^n) \Big| I(X_1^n;X_2^n|U) \leq n\delta \Big\}.$$

[1]Note that the boundary of $\mathbb{R}^2_{\geq 0}$ is given by $\mathbb{R}^2_{\geq 0} \setminus \mathbb{R}^2_{>0}$.

For every n, define the function $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as[2]

$$f_n(\delta) := \sup_{\mathcal{U}} \max_{p \in \mathcal{P}_{\mathcal{U}}^{(n)}(\delta)} \frac{1}{n} I(X_1^n, X_2^n; Y^n|U), \tag{C.1}$$

where the supremum is over all finite sets $\mathcal{U}$. For every $\delta \geq 0$, $(f_n(\delta))_{n=1}^{\infty}$ satisfies the following superadditivity property.[3]

Lemma C.1.2. *For all $m, n \geq 1$ and all $\delta \geq 0$,*

$$f_{m+n}(\delta) \geq \frac{m}{m+n} f_m(\delta) + \frac{n}{m+n} f_n(\delta).$$

Thus by [13, Appendix 4A, Lemma 2], the sequence $(f_n(\delta))_{n=1}^{\infty}$ converges for every $\delta \geq 0$, and

$$\lim_{n\to\infty} f_n(\delta) = \sup_n f_n(\delta).$$

Therefore, we can define the function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as

$$f(\delta) := \lim_{n\to\infty} f_n(\delta). \tag{C.2}$$

We next state our inner and outer bounds for $C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$.

Lemma C.1.3. *For any MAC, we have*

$$f(C_{\text{out}}^1 + C_{\text{out}}^2) - \min\{C_{\text{out}}^1, C_{\text{out}}^2\} \leq C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) \leq f(C_{\text{out}}^1 + C_{\text{out}}^2).$$

Thus to prove that $C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$ is continuous on the boundary of $\mathbb{R}_{\geq 0}^2$, it suffices to show that f is continuous on $\mathbb{R}_{\geq 0}$. This is stated in the next lemma.

Lemma C.1.4. *For any finite alphabet MAC, the function f, defined by (C.2), is continuous on $\mathbb{R}_{\geq 0}$.*

C.2 Proofs

C.2.1 Proof of Lemma C.1.2

By definition, for all $\epsilon > 0$, there exist finite alphabets $\mathcal{U}_0$ and $\mathcal{U}_1$ and distributions $p_n \in \mathcal{P}_{\mathcal{U}_0}^{(n)}(\delta)$ and $p_m \in \mathcal{P}_{\mathcal{U}_1}^{(m)}(\delta)$ such that

$$I_n(X_1^n, X_2^n; Y^n|U_0) \geq n f_n(\delta) - n\epsilon$$
$$I_m(X_1^m, X_2^m; Y^m|U_1) \geq m f_m(\delta) - m\epsilon.$$

[2]For $n = 1$, this function also appears in the study of the MAC with negligible feedback [56].

[3]The proofs of all our results are given in Section C.2.

Consider the distribution

$$p_{n+m}(u_0, u_1, x_1^{n+m}, x_2^{n+m}) = p_n(u_0, x_1^n, x_2^n) p_m(u_1, x_1^{n+1:n+m}, x_2^{n+1:n+m}).$$

For $\mathcal{U} = \mathcal{U}_0 \times \mathcal{U}_1$, it is simple to show that $p_{n+m} \in \mathcal{P}_{\mathcal{U}}^{(n+m)}(\delta)$ and

$$I_{n+m}(X_1^{n+m}, X_2^{n+m}; Y^{n+m} | U_0, U_1) \geq n f_n(\delta) + m f_m(\delta) - (n+m)\epsilon,$$

which implies the desired result.

C.2.2 Proof of Lemma C.1.3

We first prove the lower bound. For $i \in \{1, 2\}$, choose C_{id} such that

$$0 \leq C_{id} \leq C_{\text{out}}^i.$$

Let $p(x_1, x_2)$ be any distribution satisfying

$$I(X_1; X_2) = C_{1d} + C_{2d}.$$

Then Corollary 3.4.1 implies that

$$C_{\text{sum}}(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}}) \geq I(X_1, X_2; Y) - \min\{C_{1d}, C_{2d}\}.$$

Applying the same corollary to the MAC

$$p(y^n | x_1^n, x_2^n) = \prod_{t \in [n]} p(y_t | x_{1t}, x_{2t}),$$

proves our lower bound.

For the upper bound, consider a sequence of $(2^{nR_1}, 2^{nR_2}, n)$-codes for the MAC with a $(\mathbf{C}_{\text{in}}^*, \mathbf{C}_{\text{out}})$-CF. By the data processing inequality,

$$I(X_1^n; X_2^n) \leq n(C_{\text{out}}^1 + C_{\text{out}}^2).$$

In addition, from Fano's inequality it follows that there exists a sequence $(\epsilon_n)_{n=1}^{\infty}$ such that

$$H(W_1, W_2 | Y^n) \leq n\epsilon_n,$$

and $\epsilon_n \to 0$ as $n \to \infty$. We have

$$\begin{aligned} n(R_1 + R_2) &\leq H(W_1, W_2) \\ &= I(W_1, W_2; Y^n) + H(W_1, W_2 | Y^n) \\ &= I(X_1^n, X_2^n; Y^n) + n\epsilon_n \\ &\leq n f(C_{\text{out}}^1 + C_{\text{out}}^2) + n\epsilon_n. \end{aligned}$$

Dividing by n and taking the limit $n \to \infty$ completes the proof.

C.2.3 Proof of Lemma C.1.4

The proof is via a sequence of lemmas, which we next describe.

We begin by showing that for all $n \geq 1$, f_n is concave and continuous.

Lemma C.2.1 (Concavity of f_n)**.** *For all $n \geq 1$, f_n is concave on $\mathbb{R}_{\geq 0}$.*

Proof. It suffices to prove the result for $n = 1$. We apply the technique from [71]. Note that

$$f_1(\delta) = \sup_{\mathcal{U}} \max_{p \in \mathcal{P}^{(1)}_{\mathcal{U}}(\delta)} I(X_1, X_2; Y|U).$$

Fix $a, b \geq 0$, $\lambda \in (0, 1)$, and $\epsilon > 0$. Then there exist finite sets $\mathcal{U}_0$ and $\mathcal{U}_1$ and distributions $p_0 \in \mathcal{P}^{(1)}_{\mathcal{U}_0}(a)$ and $p_1 \in \mathcal{P}^{(1)}_{\mathcal{U}_1}(b)$ satisfying

$$\begin{aligned} I_0(X_1, X_2; Y|U_0) &\geq f_1(a) - \epsilon \\ I_1(X_1, X_2; Y|U_1) &\geq f_1(b) - \epsilon, \end{aligned}$$

respectively. Define the alphabet $\mathcal{V}$ as

$$\mathcal{V} := \{0\} \times \mathcal{U}_0 \cup \{1\} \times \mathcal{U}_1.$$

We denote an element of $\mathcal{V}$ by $v = (v_1, v_2)$. Define the distribution $p_\lambda(v, x_1, x_2)$ as

$$p_\lambda(v, x_1, x_2) = p_\lambda(v_1) p_{v_1}(v_2, x_1, x_2),$$

where

$$p_\lambda(v_1) = \begin{cases} 1 - \lambda & \text{if } v_1 = 0 \\ \lambda & \text{if } v_1 = 1. \end{cases}$$

Then

$$\begin{aligned} I_\lambda(X_1; X_2|V) &= I_\lambda(X_1, X_2|V_1, V_2) \\ &= (1-\lambda) I(X_1; X_2|V_1 = 0, V_2) + \lambda I(X_1; X_2|V_1 = 1, V_2) \\ &= (1-\lambda) I_0(X_1; X_2|U_0) + \lambda I_1(X_1; X_2|U_1) \\ &\leq (1-\lambda) a + \lambda b, \end{aligned}$$

which implies $p_\lambda \in \mathcal{P}^{(1)}_{\mathcal{V}}((1-\lambda)a + \lambda b)$. Similarly,

$$\begin{aligned} I_\lambda(X_1, X_2; Y|V) &= I_\lambda(X_1, X_2; Y|V_1, V_2) \\ &= (1-\lambda) I(X_1, X_2; Y|V_1 = 0, V_2) + \lambda I(X_1, X_2; Y|V_1 = 1, V_2) \\ &= (1-\lambda) I_0(X_1, X_2; Y|U_0) + \lambda I_1(X_1, X_2; Y|U_1) \\ &\geq (1-\lambda) f_1(a) + \lambda f_1(b) - \epsilon. \end{aligned}$$

Therefore,

$$f_1\big((1-\lambda)a+\lambda b\big) \geq (1-\lambda)f_1(a)+\lambda f_1(b)-\epsilon.$$

The result now follows from the fact that the above equation holds for all $\epsilon > 0$. $\square$

Lemma C.2.2 (Cardinality of $\mathcal{U}$)**.** *In the definition of $f_n(\delta)$, namely*

$$f_n(\delta) = \sup_{\mathcal{U}} \max_{p\in\mathcal{P}_{\mathcal{U}}^{(n)}(\delta)} \frac{1}{n} I(X_1^n, X_2^n; Y^n|U),$$

it suffices to calculate the supremum over all sets $\mathcal{U}$ with $|\mathcal{U}| \leq |\mathcal{X}_1|^n|\mathcal{X}_2|^n+1$.

Proof. We prove the result for $n = 1$. A similar argument applies for any positive integer n. Let $\mathcal{U}$ be a finite set with

$$|\mathcal{U}| > |\mathcal{X}_1||\mathcal{X}_2|+1,$$

and let $p^* \in \mathcal{P}_{\mathcal{U}}^{(1)}(\delta)$ be the distribution that achieves

$$\max_{p\in\mathcal{P}_{\mathcal{U}}^{(1)}(\delta)} I(X_1, X_2; Y|U).$$

Define the set $\mathcal{Q}$ as

$$\begin{aligned}\mathcal{Q} := \Big\{q\in\mathbb{R}^{|\mathcal{U}|}\,\Big|\, &\forall u\in\mathcal{U}: q(u)\geq 0,\\ &\forall(x_1,x_2): \sum_{u\in\mathcal{U}} q(u)p^*(x_1,x_2|u) = p^*(x_1,x_2),\\ &\sum_{u\in\mathcal{U}} q(u)I^*(X_1;X_2|U=u) = I^*(X_1;X_2|U)\Big\}.\end{aligned}$$

Note that $\mathcal{Q}$ is a convex polytope in $\mathbb{R}^{|\mathcal{U}|}$ and the mapping $F : \mathcal{Q} \to \mathbb{R}_{\geq 0}$ defined as

$$F(q) := \sum_{u\in\mathcal{U}} q(u)I^*(X_1, X_2; Y|U=u)$$

is convex. Thus there exists an extremal point $q^* \in \mathcal{Q}$ at which F obtains its maximum. Any extremal point of $\mathcal{Q}$ must satisfy at least $|\mathcal{U}|$ constraints in the definition of $\mathcal{Q}$ with equality; thus $q^*(u) = 0$ for at least

$$|\mathcal{U}| - \big(|\mathcal{X}_1||\mathcal{X}_2|+1\big)$$

values in $\mathcal{U}$. This completes the proof. $\square$

Lemma C.2.3 (Continuity of f_n)**.** *For all $n \geq 1$, f_n is continuous on $\mathbb{R}_{\geq 0}$.*

Proof. Similar to prior lemmas, it suffices to prove the result for $n = 1$. By Lemma C.2.1, f_1 is concave on $\mathbb{R}_{\geq 0}$; thus, it is continuous on $\mathbb{R}_{>0}$. Therefore, we only need to prove f_1 is continuous at $\delta = 0^+$. Let $\mathcal{U}$ be a set of cardinality $1 + |\mathcal{X}_1||\mathcal{X}_2|$. By Lemma C.2.2, for all $\delta \geq 0$, we have

$$f_1(\delta) = \max_{p \in \mathcal{P}_{\mathcal{U}}^{(1)}(\delta)} I(X_1, X_2; Y|U).$$

Fix $\delta > 0$. Let $p^*(u, x_1, x_2)$ be a distribution in $\mathcal{P}_{\mathcal{U}}^{(1)}(\delta)$ achieving the maximum above, and define

$$p^*_{\text{ind}}(x_1, x_2|u) := p^*(x_1|u)p^*(x_2|u).$$

Since

$$\sum_{u \in \mathcal{U}} p^*(u) D\big(p^*(x_1, x_2|u) \| p^*_{\text{ind}}(x_1, x_2|u)\big) = I^*(X_1; X_2|U) \leq \delta,$$

by [40, Lemma 11.6.1],

$$\sum_{u \in \mathcal{U}} p^*(u) \big\|p^*(x_1, x_2|u) - p^*_{\text{ind}}(x_1, x_2|u)\big\|_{L^1}^2 \leq 2\delta \ln 2. \tag{C.3}$$

Furthermore,

$$\begin{aligned}
&\sum_{u \in \mathcal{U}} p^*(u) \big\|p^*(y|u) - p^*_{\text{ind}}(y|u)\big\|_{L^1} \\
&\quad \leq \sum_{u \in \mathcal{U}} p^*(u) \sum_{x_1, x_2} p(y|x_1, x_2) \big|p^*(x_1, x_2|u) - p^*_{\text{ind}}(x_1, x_2|u)\big| \\
&\quad \leq \sum_{u \in \mathcal{U}} p^*(u) \big\|p^*(x_1, x_2|u) - p^*_{\text{ind}}(x_1, x_2|u)\big\|_{L^1} \\
&\quad \leq \sqrt{2\delta \ln 2},
\end{aligned} \tag{C.4}$$

where (C.4) follows from (C.3) and the Cauchy-Schwarz inequality. Define the subset $\mathcal{U}_0 \subseteq \mathcal{U}$ as

$$\mathcal{U}_0 = \Big\{u \in \mathcal{U} : \big\|p^*(y|u) - p^*_{\text{ind}}(y|u)\big\|_{L^1} \leq 1/2\Big\}.$$

Clearly,

$$\sum_{u \notin \mathcal{U}_0} p^*(u) \leq 2\sqrt{2\delta \ln 2}.$$

Thus

$$
\begin{aligned}
&\left|H^*(Y|U) - H^*_{\text{ind}}(Y|U)\right| \\
&\quad \le \sum_{u\in\mathcal{U}} p^*(u)\left|H^*(Y|U=u) - H^*_{\text{ind}}(Y|U=u)\right| \\
&\quad \overset{(a)}{\le} 2\sqrt{2\delta\ln 2}\log|\mathcal{Y}| \\
&\qquad - \sum_{u\in\mathcal{U}_0} p^*(u)\left\|p^*(y|u) - p^*_{\text{ind}}(y|u)\right\|_{L^1} \log\frac{\left\|p^*(y|u) - p^*_{\text{ind}}(y|u)\right\|_{L^1}}{|\mathcal{Y}|} \\
&\quad \overset{(b)}{\le} 2\sqrt{2\delta\ln 2}\log|\mathcal{Y}| - \sqrt{2\delta\ln 2}\log\Big(\frac{1}{|\mathcal{Y}|}\sqrt{2\delta\ln 2}\Big) \\
&\quad = \sqrt{2\delta\ln 2}\log\frac{|\mathcal{Y}|^3}{\sqrt{2\delta\ln 2}},
\end{aligned}
$$

where (a) follows from [40, Theorem 17.3.3] and (b) follows from the fact that the mapping $t \mapsto -t\log(t/|\mathcal{Y}|)$ is concave and increasing for sufficiently small t. In addition, by (C.4),

$$
\begin{aligned}
&\left|H^*(Y|U,X_1,X_2) - H^*_{\text{ind}}(Y|U,X_1,X_2)\right| \\
&\quad \le \sum_{u,x_1,x_2} \left|p^*(u,x_1,x_2) - p^*_{\text{ind}}(u,x_1,x_2)\right| H(Y|X_1=x_1,X_2=x_2) \\
&\quad \le \Big(\log|\mathcal{Y}|\Big)\sqrt{2\delta\ln 2}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
f_1(\delta) &= I^*(X_1,X_2;Y|U) = H^*(Y|U) - H^*(Y|U,X_1,X_2) \\
&\le \left|H^*(Y|U) - H^*_{\text{ind}}(Y|U)\right| + \left|H^*(Y|U,X_1,X_2) - H^*_{\text{ind}}(Y|U,X_1,X_2)\right| \\
&\quad + I^*_{\text{ind}}(X_1,X_2;Y|U) \\
&\le \sqrt{2\delta\ln 2}\log\frac{|\mathcal{Y}|^3}{\sqrt{2\delta\ln 2}} + \Big(\log|\mathcal{Y}|\Big)\sqrt{2\delta\ln 2} + f_1(0).
\end{aligned}
$$

Since $f_1(0) \le f_1(\delta)$ for all $\delta \ge 0$, the continuity of f_1 at $\delta = 0^+$ follows. □

The next lemma is proved by Dueck [70]. We include the proof for completeness.

Lemma C.2.4. *Fix $\epsilon, \delta > 0$, positive integer n, and finite alphabet $\mathcal{U}$. If $p \in \mathcal{P}^{(n)}_{\mathcal{U}}(\delta)$, then there exists a set $T \subseteq [n]$ satisfying*

$$|T| \le \frac{n\delta}{\epsilon},$$

and

$$\forall t \notin T\colon I(X_{1t}; X_{2t}|U, X_1^T, X_2^T) \leq \epsilon,$$

where for $i \in \{1,2\}$, $X_i^T := (X_{it})_{t\in T}$.

Proof. If for all $t \in [n]$, we have

$$I(X_{1t}; X_{2t}|U) \leq \epsilon,$$

then we define $T := \varnothing$. Otherwise, there exists $t_1 \in [n]$ such that

$$I(X_{1t_1}; X_{2t_1}|U) > \epsilon.$$

Let $S_1 := [n] \setminus \{t_1\}$. Then

$$\begin{aligned} I(X_1^n; X_2^n|U) &= I(X_1^n; X_{2t_1}|U) + I(X_1^n; X_2^{S_1}|X_{2t_1}) \\ &= I(X_{1t_1}; X_{2t_1}|U) + I(X_1^{S_1}; X_{2t_1}|U, X_{1t_1}) \\ &\quad + I(X_{1t_1}; X_2^{S_1}|U, X_{2t_1}) + I(X_1^{S_1}; X_2^{S_1}|U, X_{1t_1}, X_{2t_1}) \\ &\geq I(X_{1t_1}; X_{2t_1}|U) + I(X_1^{S_1}; X_2^{S_1}|U, X_{1t_1}, X_{2t_1}). \end{aligned}$$

Therefore, since $p \in \mathcal{P}_{\mathcal{U}}^{(n)}(\delta)$,

$$I(X_1^{S_1}; X_2^{S_1}|U, X_{1t_1}, X_{2t_1}) \leq n\delta - \epsilon.$$

Now if for all $t \in S_1$,

$$I(X_{1t}; X_{2t}|U, X_{1t_1}, X_{2t_1}) \leq \epsilon,$$

then we define $T := \{t_1\}$. Otherwise, there exists $t_2 \in [n]$ such that

$$I(X_{1t_2}; X_{2t_2}|U, X_{1t_1}, X_{2t_1}) > \epsilon.$$

Similar to the above argument, if we define $S_2 := [n] \setminus \{t_1, t_2\}$, then

$$I(X_1^{S_2}; X_2^{S_2}|U, X_{1t_1}, X_{1t_2}, X_{2t_1}, X_{2t_2}) \leq n\delta - 2\epsilon.$$

If we continue this process, we eventually get a set $T := \{t_1, \ldots, t_k\}$ such that

$$I(X_1^{T^c}; X_2^{T^c}|U, X_1^T, X_2^T) \leq n\delta - |T|\epsilon, \tag{C.5}$$

and for all $t \in [n] \setminus T$,

$$I(X_{1t}; X_{2t}|U, X_1^T, X_2^T) \leq \epsilon.$$

In addition, from (C.5) it follows that

$$|T| \leq \frac{n\delta}{\epsilon}. \tag{C.6}$$

□

The next corollary combines bounds given in [15] with ideas developed here.

Corollary C.2.5. *For all $\epsilon, \delta > 0$, we have*

$$f(\delta) \le \frac{\delta}{\epsilon} \log |\mathcal{X}_1||\mathcal{X}_2| + f_1(\epsilon).$$

Proof. Fix a positive integer n. Let $\mathcal{U}$ be a set with cardinality $|\mathcal{X}_1|^n|\mathcal{X}_2|^n + 1$ and choose a distribution $p \in \mathcal{P}_{\mathcal{U}}^{(n)}(\delta)$. From Lemma C.2.4, it follows that there exists a set $T \subseteq [n]$ such that

$$0 \le |T| \le \frac{n\delta}{\epsilon},$$

and

$$\forall t \notin T \colon I(X_{1t}; X_{2t}|U, X_1^T, X_2^T) \le \epsilon.$$

Thus

$$\begin{aligned} I(X_1^n, X_2^n; Y^n|U) &= I(X_1^T, X_2^T; Y^n|U) + I(X_1^{T^c}, X_2^{T^c}; Y^n|U, X_1^T, X_2^T) \\ &\le |T| \log |\mathcal{X}_1||\mathcal{X}_2| + I(X_1^{T^c}, X_2^{T^c}; Y^n|U, X_1^T, X_2^T). \end{aligned}$$

We further bound the second term on the right hand side by

$$\begin{aligned} &I(X_1^{T^c}, X_2^{T^c}; Y^n|U, X_1^T, X_2^T) \\ &\quad = I(X_1^{T^c}, X_2^{T^c}; Y^{T^c}|U, X_1^T, X_2^T) + I(X_1^{T^c}, X_2^{T^c}; Y^T|U, X_1^T, X_2^T, Y^{T^c}) \\ &\quad \le \sum_{t \notin T} I(X_{1t}, X_{2t}; Y_t|U, X_1^T, X_2^T) \\ &\quad \le n \max_{p \in \mathcal{P}_{\mathcal{V}}^{(1)}(\epsilon)} I(X_1, X_2; Y|V) \le n f_1(\epsilon), \end{aligned}$$

where

$$\mathcal{V} := \mathcal{U} \times \mathcal{X}_1^{|T|} \times \mathcal{X}_2^{|T|}.$$

Therefore,

$$\frac{1}{n} I(X_1^n, X_2^n; Y^n|U) \le \frac{\delta}{\epsilon} \log |\mathcal{X}_1||\mathcal{X}_2| + f_1(\epsilon).$$

Our result now follows from Lemma C.2.2. □

By Corollary C.2.5, for all $\epsilon, \delta > 0$,

$$f(\delta) \le \frac{\delta}{\epsilon} \log |\mathcal{X}_1||\mathcal{X}_2| + f_1(\epsilon).$$

Thus

$$f(0) \le \lim_{\delta \to 0^+} f(\delta) \le f_1(\epsilon).$$

Now take the limit $\epsilon \to 0^+$ and note that f_1 is continuous by Lemma C.2.3, and $f(0) = f_1(0)$.

Appendix D

EDGE REMOVAL AT THE RECEIVERS

In this appendix, we study the effect of removing an ingoing edge of a node that has no outgoing edges. Specifically, we show that for memoryless stationary networks, such an edge always satisfies the universal edge removal property. We remark that this fact is stated in [12], where a proof sketch is provided. Here we present a more detailed proof.

An important corollary of this result is that unlike encoder cooperation, we cannot use decoder cooperation to construct networks that do not satisfy the universal edge removal property.

D.1 Memoryless Stationary Networks

Consider a memoryless stationary noisy network[1] with K source messages, M transmitters, and L receivers. The source messages are uniformly distributed and independent. For every $m \in [M]$, transmitter m has access to a subset $T(m) \subseteq [K]$ of source messages. For $\ell \in [L]$, receiver ℓ demands subset $D(\ell) \subseteq [K]$ of source messages with small average probability. In addition, receiver ℓ has access to side-information Z_ℓ of rate δ_ℓ.[2] We denote this network by $\mathcal{N}(\boldsymbol{\delta})$, where $\boldsymbol{\delta} := (\delta_\ell)_{\ell \in [L]}$. (See Figure D.1.) For $k \in [K]$, define

$$\bar{\delta}_k := \max_{\ell:\, k \in D(\ell)} \delta_\ell.$$

The main result of this appendix follows.

Proposition D.1.1. *If the rate vector $(R_k)_{k\in[K]}$ is in the average-error capacity region of $\mathcal{N}(\boldsymbol{\delta})$, then the rate vector*

$$\left(\left(R_k - \bar{\delta}_k\right)^+\right)_{k\in[K]}$$

is in the average-error capacity region of $\mathcal{N}(\mathbf{0})$.

[1] As described in the Introduction, this is a network consisting of both memoryless stationary multi-terminal components and noiseless point-to-point links. An example is the network consisting of a memoryless stationary MAC and a CF.

[2] The only constraint on the side-information is rate; that is, for a code of blocklength n, receiver ℓ receives at most $n\delta_\ell$ bits of side-information. Thus the random variable Z_ℓ can depend on the inputs of any network node.

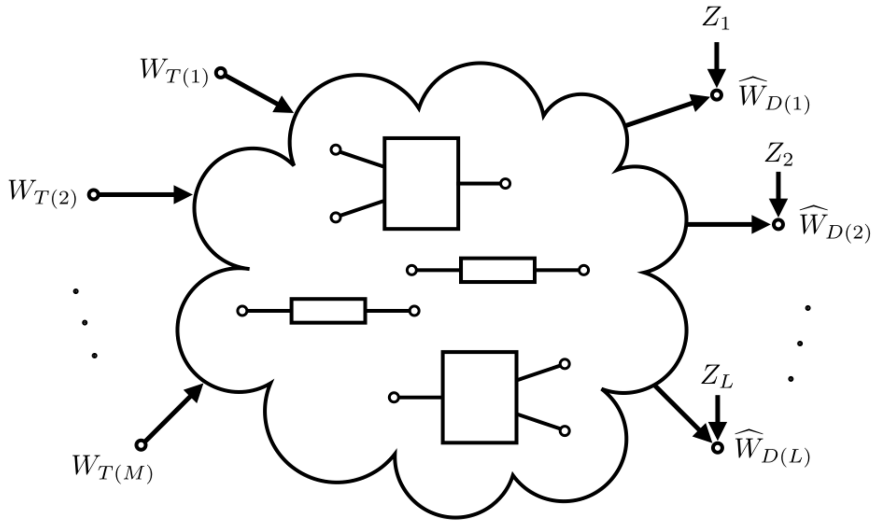

Figure D.1: The network $\mathcal{N}(\boldsymbol{\delta})$.

Proof. Suppose the rate vector $(R_k^*)_{k\in[K]}$ is achievable over $\mathcal{N}(\boldsymbol{\delta})$. Then for every $\epsilon > 0$, there exists a

$$\Big((2^{NR_1^*}, \ldots, 2^{NR_K^*}), N, \epsilon\Big)\text{-code}$$

for $\mathcal{N}(\boldsymbol{\delta})$. Fix $\epsilon > 0$ and a corresponding code C. For this code, let X_k^N denote the output of transmitter k and Y_ℓ^N denote the input to receiver ℓ. Furthermore, for $k \in [K]$, let W_k denote source message k. Note that by assumption, W_k is uniformly distributed on its source alphabet $[2^{NR_k^*}]$, and for all $\ell \in [L]$, $Z_\ell \in [2^{N\delta_\ell}]$. For $\ell \in [L]$ satisfying $k \in D(\ell)$, we have

$$\begin{aligned} R_k^* &= \frac{1}{N} H(W_k) \\ &= \frac{1}{N} I(W_k; Y_\ell^N, Z_\ell) + \epsilon_N(k,\ell) \\ &\leq \frac{1}{N} I(W_k; Y_\ell^N) + \delta_\ell + \epsilon_N(k,\ell) \\ &\leq \frac{1}{N} I(U_k; Y_\ell^N) + \bar{\delta}_k + \epsilon_N(k,\ell), \end{aligned}$$

where $U_k := W_k$, and

$$\epsilon_N(k,\ell) := \frac{1}{N} H(W_k | Y_\ell^N, Z_\ell).$$

By Fano's inequality [40, p. 38],

$$\epsilon_N(k,\ell) \leq \frac{1}{N}\big(h(\epsilon) + NR_k^*\epsilon\big) = o_\epsilon(1),$$

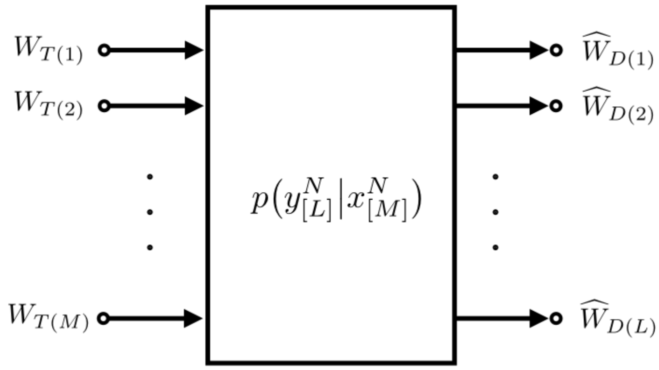

Figure D.2: The network $\mathcal{N}_{\mathrm{C}}(\mathbf{0})$.

where $h(\epsilon)$ is the binary entropy function. Thus for every $k \in [K]$,

$$R_k^* \leq \frac{1}{N} \min_{\ell : k \in D(\ell)} I(U_k; Y_\ell^N) + \bar{\delta}_k + o_\epsilon(1). \tag{D.1}$$

Note that the code C defines a conditional distribution

$$p\big(y_{[L]}^N \big| x_{[M]}^N\big) := p\big(y_1^N, \dots, y_L^N \big| x_1^N, \dots, x_M^N\big). \tag{D.2}$$

Now consider the network $\mathcal{N}_{\mathrm{C}}(\mathbf{0})$, depicted in Figure D.2, which consists of M transmitters, L receivers, and a single multi-terminal component defined by (D.2). Somekh-Baruch and Verdú [72] show that the set of all rate vectors $(R_k)_{k \in [K]}$ that for every $k \in [K]$, obey

$$R_k \leq \min_{\ell : k \in D(\ell)} I(U_k; Y_\ell^N)$$

for some distribution

$$p(u_1) \dots p(u_K) \prod_{m \in [M]} p\big(x_m^N | u_{T(m)}\big),$$

is a subset of the capacity region of $\mathcal{N}_{\mathrm{C}}(\mathbf{0})$. Now if a rate vector $(R_k)_{k \in [K]}$ is achievable over $\mathcal{N}_{\mathrm{C}}(\mathbf{0})$, then $(R_k/N)_{k \in [K]}$ is achievable over $\mathcal{N}(\mathbf{0})$. Thus by (D.1),

$$\Big(\big(R_k^* - \bar{\delta}_k\big)^+\Big)_{k \in [K]}$$

is in the capacity region of $\mathcal{N}(\mathbf{0})$. $\square$

BIBLIOGRAPHY

[1] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications", *Foundations and Trends in Networking*, vol. 1, no. 3-4, pp. 271–425, 2006.

[2] R. Ahlswede, "Multi-way communication channels", in *Proc. Int. Symp. Information Theory*, 1971.

[3] ——, "The capacity region of a channel with two senders and two receivers", *Ann. Prob.*, vol. 2, pp. 805–814, 1974.

[4] H. Liao, "Multiple access channels", PhD thesis, Dept. Elect. Eng., Univ. Hawaii, 1972.

[5] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders", *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, 1983.

[6] A. D. Wyner, "Recent results in the Shannon theory", *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 2–10, 1974.

[7] M. L. Ulrey, "The capacity region of a channel with s senders and r receivers", *Inform. Contr.*, vol. 29, pp. 185–203, 3 1975.

[8] K. Marton, "A coding theorem for the discrete memoryless broadcast channel", *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, 1979.

[9] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information", *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[10] S. Jafar, "Capacity with causal and noncausal side information: A unified view", *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, 2006.

[11] H. H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels", *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, 2011.

[12] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity", in *Information Theory and Applications Workshop*, 2011.

[13] R. G. Gallager, *Information theory and reliable communication*, 1st ed. Wiley, 1968.

[14] C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[15] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels", *Probl. Contr. Inform. Theory*, vol. 7, no. 1, pp. 11–19, 1978.

[16] F. M. J. Willems, "The maximal-error and average-error capacity region of the broadcast channel are identical: A direct proof", *Probl. Contr. Inform. Theory*, vol. 19, no. 4, pp. 339–347, 1990.

[17] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow", *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[18] W. Gu, M. Effros, and M. Bakshi, "A continuity theory for lossless source coding over networks", in *Proc. Allerton Conf. Communication, Control, and Computing*, 2008.

[19] T. Chan and A. Grant, "On capacity regions of non-multicast networks", in *Proc. IEEE Int. Symp. Information Theory*, 2010.

[20] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies", in *Proc. Allerton Conf. Communication, Control, and Computing*, 2010.

[21] O. Kosut and J. Kliewer, "On the relationship between edge removal and strong converses", in *Proc. IEEE Int. Symp. Information Theory*, 2016.

[22] E. J. Lee, M. Langberg, and M. Effros, "Outer bounds and a functional study of the edge removal problem", in *Proc. IEEE Information Theory Workshop*, 2013.

[23] R. W. Yeung, "A framework for linear information inequalities", *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1924–1934, 1997.

[24] W. Gu, "On achievable rate regions for source coding over networks", PhD thesis, Caltech, 2009.

[25] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems", in *Proc. IEEE Int. Symp. Information Theory*, 2009.

[26] M. Langberg and M. Effros, "Network coding: Is zero error always possible?", in *Proc. Allerton Conf. Communication, Control, and Computing*, 2011.

[27] ——, "Source coding for dependent sources", in *Proc. IEEE Information Theory Workshop*, 2012.

[28] M. F. Wong, M. Langberg, and M. Effros, "On a capacity equivalence between network and index coding and the edge removal problem", in *Proc. IEEE Int. Symp. Information Theory*, 2013.

[29] M. F. Wong, M. Effros, and M. Langberg, "On an equivalence of the reduction of k-unicast to 2-unicast capacity and the edge removal property", in *Proc. IEEE Int. Symp. Information Theory*, 2015.

[30] ——, "On tightness of an entropic region outer bound for network coding and the edge removal property", in *Proc. IEEE Int. Symp. Information Theory*, 2016.

[31] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part I: Point-to-point channels", *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, 2011.

[32] P. Noorzad, M. Effros, M. Langberg, and T. Ho, "On the power of cooperation: Can a little help a lot?", in *Proc. IEEE Int. Symp. Information Theory*, 2014.

[33] J. Spencer, *Ten lectures on the probabilistic method*. SIAM, 1994.

[34] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[35] P. Noorzad, M. Effros, and M. Langberg, "The unbounded benefit of encoder cooperation for the k-user MAC", in *Proc. IEEE Int. Symp. Information Theory*, 2016.

[36] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources", *Inform. Contr.*, vol. 40, pp. 37–60, 1 1979.

[37] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel", *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, 1981.

[38] A. El Gamal and Y.-H. Kim, *Network information theory*, 2nd ed. Cambridge University Press, 2012.

[39] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part II: Multiterminal channels", *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3709–3732, 2014.

[40] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Wiley, 2006.

[41] O. Simeone, O. Somekh, G. Kramer, H. V. Poor, and S. Shamai, "Three-user Gaussian multiple access channel with partially cooperating encoders", in *Asilomar Conf. on Signals, Systems, and Computers*, 2008.

[42] J. D. Hunter, "Matplotlib: A 2D graphics environment", *IEEE Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[43] P. Noorzad, M. Effros, and M. Langberg, "On the cost and benefit of cooperation", in *Proc. IEEE Int. Symp. Information Theory*, 2015.

[44] S. Watanabe, “Information theoretical analysis of multivariate correlation”, *IBM J. Res. Dev.*, vol. 4, pp. 66–82, 1 1960.

[45] A. Schrijver, *Combinatorial optimization: Polyhedra and efficiency*. Springer-Verlag, 2003, vol. B.

[46] P. Noorzad, M. Effros, and M. Langberg, “The benefit of encoder cooperation in the presence of state information”, in *Proc. IEEE Int. Symp. Information Theory*, 2017.

[47] D. N. C. Tse and S. V. Hanly, “Multiaccess fading channels–Part I: Polymatroid structure, optimal resource allocation and throughput capacities”, *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.

[48] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, “Cooperative multiple-access encoding with states available at one transmitter”, *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, 2008.

[49] C. Heegard and A. A. El Gamal, “On the capacity of computer memory with defects”, *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, 1983.

[50] M. H. M. Costa, “Writing on dirty paper”, *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, 1983.

[51] Y. Cemal and Y. Steinberg, “The multiple-access channel with partial state information at the encoders”, *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3992–4003, 2005.

[52] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Common state”, *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, 2013.

[53] ——, “The multiple-access channel with causal side information: Double state”, *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1379–1393, 2013.

[54] J. E. Angus, “The probability integral transform and related results”, *SIAM Review*, vol. 36, no. 4, pp. 652–654, 1994.

[55] P. Noorzad, M. Effros, and M. Langberg, “Can negligible cooperation increase network reliability?”, in *Proc. IEEE Int. Symp. Information Theory*, 2016.

[56] A. D. Sarwate and M. Gastpar, “Some observations on limited feedback for multiaccess channels”, in *Proc. IEEE Int. Symp. Information Theory*, 2009.

[57] M. Langberg and M. Effros, “On the capacity advantage of a single bit”, in *IEEE Globecom Workshops*, 2016.

[58] N. Cai, “The maximum error probability criterion, random encoder, and feedback, in multiple input channels”, *Entropy*, vol. 16, no. 3, pp. 1211–1242, 2014.

[59] R. Schneider, *Convex bodies: The Brunn-Minkowski theory*. Cambridge University Press, 1993.

[60] R. Lucchetti, *Convexity and well-posed problems*, 1st ed. Springer, 2006.

[61] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, 2nd ed. MIT Press, 2001.

[62] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed. Cambridge University Press, 2004.

[63] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels", *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[64] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels", *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 758–782, 2015.

[65] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving Marton's region for broadcast channels using polar codes", *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 783–800, 2015.

[66] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel", *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, 2012.

[67] S. Verdú, "Non-asymptotic covering lemmas", in *IEEE Information Theory Workshop (ITW)*, 2015.

[68] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and Marton's inner bound with a common message", in *Proc. IEEE Int. Symp. Information Theory*, 2015.

[69] P. Billingsley, *Probability and measure*, 3rd ed. SIAM, 1995.

[70] G. Dueck, "The strong converse of the coding theorem for the multiple-access channel", *Journal of Combinatorics, Information, and System Sciences*, vol. 6, pp. 187–196, 3 1981.

[71] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrary correlated sources", *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, 1980.

[72] A. Somekh-Baruch and S. Verdú, "General relayless networks: Representation of the capacity region", in *Proc. IEEE Int. Symp. Information Theory*, 2006.