THE IMPACT OF IMPERFECT FEEDBACK ON THE CAPACITY OF WIRELESS NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alireza Vahid

January 2015

THE IMPACT OF IMPERFECT FEEDBACK ON THE CAPACITY OF WIRELESS NETWORKS

Alireza Vahid, Ph.D.
Cornell University 2015

The defining metric for any wireless communication network is the maximum reliable data rate, also known as capacity. Before any data bit can be communicated over a wireless channel, information about the network state such as connectivity, channel statistics, or channel gains, is required. Receiver nodes feed such information back to other wireless nodes using feedback channels and feedback mechanisms. Considering the role of feedback makes the characterization of the network capacity a daunting theoretical task. As a result, researchers have overwhelmingly simplified the feedback channels to come up with tractable models. Specifically, it has commonly been assumed that feedback channel has infinite capacity and has no delay. While these assumptions could be justified for small, static, or slow-fading networks, they are not viable in the context of large-scale, mobile, or fast-fading wireless networks. In fact, feedback channel is low-rate, unreliable, scarce, and is subject to delay. The recent dramatic increase in wireless data traffic, caused by the success of online media streaming and the proliferation of smart phones and tablets, obliges researchers to understand the capacity of large-scale mobile networks. Thus, given the limited, scarce nature of feedback channels, future progress in wireless data communications crucially depends on a deeper understanding of the impact of feedback on the capacity of wireless networks.

In this work, we aim to adjust the assumptions on feedback channels in

a way to open doors to better understanding of real world, large-scale wireless networks. In particular, wireless networks are considered with rate-limited feedback links, with outdated channel state information, and with local views of the network state.

BIOGRAPHICAL SKETCH

Alireza Vahid received the B.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2009, and the M.Sc. degree in Electrical and Computer Engineering from Cornell University, Ithaca, NY, in 2012. As of September 2014, he is a postdoctoral scholar at Information Initiative at Duke University (*i*iD), Durham, NC. His research interests include information theory and wireless communications, statistics and machine learning.

He has received the Director's Ph.D. Teaching Assistant Award in 2010 from the school of electrical and computer engineering, Cornell University, and Jacobs Scholar Fellowship in 2009. He has also received Qualcomm Innovation Fellowship in 2013 for his research on "Collaborative Interference Management." He was ranked **second** among more than 450000 participant in Iranian National Entrance Exam 2004.

To my beloved family, Mohammad Bagher, Simin, Sara, and Sina.

ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my advisor, Professor Salman Avestimehr for his unconditional support of my research. Not only did he teach me how to conduct research, he also helped me pick problems of significant importance. I have truly learned from his unique style of research and presentation. His dedication to his work and his students has always been a source of inspiration to me, and his immense enthusiasm even after so many achievements has always surprised me.

I would like to thank my PhD committee members, Professor Lang Tong and Professor Stephen Wicker for their valuable comments and their guidance. I have gained valuable insights from them. As a student, I have also learned significant amount of knowledge in the courses they presented.

I would also like to thank Dr. Mohammad Ali Maddah-Ali for sharing his knowledge and experience with me. I had an amazing experience working with him. It was through him that I learned the meaning of "thinking outside the box!" He truly does his best to provide his colleagues with the best support he can. He has exceptional knowledge of the field that has helped me a lot during the last two years of my PhD studies.

I would also like to express my gratitude to Professor Ashutosh Sabharwal for his thoughtful comments and discussions throughout the joint projects. He is a great mentor and I have gathered priceless training from his style of writing and his thought process.

I would like to thank Dr. Vaneet Aggarwal for his help and support during our collaboration. Not only is he a colleague to me, he is also a close friend. I am confident that our collaboration and friendship will go on.

Finally, I would like to extend my deepest gratitude, love and affection to my beloved family for always being there for, supporting me unconditionally, believing in me and wishing me the best. I owe every single one of my achievements to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Recent years have seen a dramatic increase in the wireless data traffic, caused by the success of online media streaming services and the proliferation of smart phones, tablets, and netbooks. Several new advances such as fourth-generation (4G), LTE, and fifth-generation (5G) digital wireless standards have facilitated the data traffic in wireless networks. However, given the scarcity of wireless spectrum, the task of supporting the ever-increasing demand has become quite challenging and demands new innovations. The only solution, seems to be to exploit a much denser spatial reuse of the spectrum, by considering new paradigms in wireless communications such as heterogeneous networks and device-to-device communications.

Heterogeneous networks are created via the deployment of several low-power nodes such as pico stations, femto stations and relays, besides the existing high-power base station infrastructure. In this organically-grown multi-flow multi-terminal setting, interference becomes the dominating bottleneck for network capacity improvement. Similarly, device-to-device communication, while offering improvements in network coverage and quality of service, comes at the price of higher interference generation. Thus, interference management is of significant importance to the performance of wireless networks. As a result, the key to increase the performance of any wireless network is to maximize the operational efficiency by optimally using available spectral using interference management techniques.

In order to perform any interference management technique, wireless nodes need to acquire some information about the network status such as network connectivity or topology, and the channel gains associated with wireless links. The task of providing wireless nodes with network information is quite challenging due to the dynamic state of the networks, *i.e.* node and environmental mobility. Mobility leads to significant changes in network topology, network connectivity, and per-link channel gains at different time-scales. The constant flux in the state and in the available resources has made the problem nearly intractable. Hence, the question to be answered is, how should nodes choose their transmission and reception parameters to maximize the global network objective of maximal network efficiency based on *imperfect* knowledge of the network dynamics?

Although distributed decision making is a well known area and has been studied by many researchers, modeling and understanding the fundamental capacity limits of wireless networks with imperfect state knowledge is an untouched area. While development of network protocols is many decades old, fundamental results to bound the performance of distributed decision making are basically non-existent and most of the works in network information theory have been done on the case of *fully* coordinated nodes. Therefore, our motivation is to develop information-theoretic foundations for distributed decisions in wireless networks which lead to the design of future wireless networks. There are many challenging, yet promising directions left open in front of us and we will try to develop some guiding principles, which will have a direct impact on the design of future wireless networks.

In order to understand the fundamental limits of communications when *global* network dynamics are not available, we consider four main directions:

(a) **Communication with Delayed Channel State Information:** In this case, we take into account the delay at which network dynamics are tracked at each node. This direction provides a suitable guideline for fast-fading wireless networks. We discover new coding opportunities in wireless networks and we develop a framework to describe how these opportunities should be combined in an optimal fashion. We also develop novel techniques to derive outer-bounds and to prove the optimality of our transmission strategies.

(b) **Capacity results for Gaussian Networks with Delayed Channel State Information:** In the first direction, while we study the delay in distributed setting, we limit ourselves to a simple on-off channel model. To show the practicality of our results, we consider a two-user multiple-input single-output broadcast channel with delayed channel state information, and we derive a constant-gap approximation of the capacity region for this problem.

(c) **Communication with Local Network Views:** In this direction, we focus on the role of local network knowledge on network capacity and the associated optimal transmission strategies. We propose a new tractable framework that enables us to understand how to manage interference and how to optimally flow information with only local network views. This direction provides a suitable guideline for large-scale wireless networks where learning the entire network knowledge creates significant overhead.

(d) **Communication with Rate-Limited Feedback Links:** In this direction, we focus on the feedback links as they play a central role in learning algorithms. In the previous directions, while we have considered several constraints, we always assumed feedback links have infinite capacity. However, as expected, such assumption does not hold in practice and careful consideration is needed when feedback links have *limited* capacity.

1.2 Prior Work

As emphasized previously, to perform any interference management technique, wireless nodes need to acquire some information about the network status such as network connectivity or topology, and the channel gains associated with wireless links. This information has to be obtained via feedback mechanisms and feedback channels. Therefore, there is no surprise to see a large body of work on the role of feedback in wireless networks. In fact, the history of studying the impact of feedback channel in communication systems traces back to Shannon and in [47], he showed that feedback does not increase the capacity of discrete memoryless point-to-point channels. Ever since Shannon's work, there have been extensive efforts to discover new techniques that exploit feedback channels in order to benefit wireless networks.

First efforts on understanding the impact of feedback channels were focused on the slow-fading channels in multi-user networks where feedback links where utilized to provide transmitters with knowledge of previously transmitted signals. In particular, it was shown that in multi-user networks, even in the most basic case of the two-user memoryless multiple-access channel [20, 39]. Hence, there has been a growing interest in developing feedback strategies and understanding the fundamental limits of communication over multi-user networks with feedback, in particular the two-user interference channel (IC) (see [23, 31–33, 45, 50, 53]).

However, in today's wireless networks, a more important objective in utilizing feedback channels is to provide the transmitters with the knowledge of the channel state information (CSI). In slow-fading networks, this task could have been carried on with negligible overhead. However, as wireless networks started growing in size, as mobility became an inseparable part of networks, and as fast-fading networks started playing a more important role, the availability of up-to-date channel state information at the transmitters (CSIT) has become a challenging task to accomplish. Specifically, in fast-fading scenarios, the coherence time of the channel is smaller than the delay of the feedback channel, and thus, providing the transmitters with up-to-date channel state information is practically infeasible.

As a result, there has been a recent growing interest in studying the effect of lack of up-to-date channel state information at the transmitters in wireless networks. In particular, in the context of multiple-input single-output (MISO) broadcast channels (BC), it was shown that even completely stale CSIT (also known as Delayed-CSIT) can still be very useful and can change the scale of the capacity, measured by the degrees of freedom (DoF) [35]. A key idea behind the scheme proposed in [35] is that instead of predicting future channel state information, transmitters should focus on the side-information provided in the past signaling stages via the feedback channel, and try to create signals that are of *common interest* of multiple receivers. Hence, we can increase spectral efficiency by retransmission of such signals of common interest.

Motivated by the results of [35], there have been several results on the impact of delayed CSIT in wireless networks. This includes studying the DoF region of multi-antenna two-user Gaussian IC and X channel [25, 68], k-user Gaussian IC and X channel [2, 37], and multi-antenna two-user Gaussian IC with Delayed-CSIT and Shannon feedback [52, 70]. In particular, the DoF region of multi-antenna two-user Gaussian IC has been characterized in [71], and it has been shown that the k-user Gaussian IC and X channels can still achieve more than one DoF with Delayed-CSIT [1, 37] (for $k > 2$).

1.3 Contributions

1.3.1 Communication with Delayed Channel State Information

In wireless networks, transmitters obtain the channel state information (CSI) through feedback channels. A fundamental consideration for feedback channels is that they are subject to delay. While for networks with slow mobility, gathering and utilizing channel state information within the coherence time of the network is plausible, it becomes daunting as the mobility and size of the network increase (*e.g.*, millisecond time-scale coherence time). Hence, in fast fading scenarios, transmitters can no longer rely on up-to-date CSI. We would like to study how delayed channel state information at the transmitters (CSIT) can be utilized for interference management purposes in wireless networks.

In prior work, Maddah-Ali and Tse [35] showed that in broadcast channels delayed channel state information is very useful and can change the scale of the capacity, measured by the degrees-of-freedom (DoF). We would like to understand the potential of delayed CSI, in the context of interference channels (distributed transmitters). Consider the network in Figure 1.1(a), in which two transmitters, Tx_1 and Tx_2, are interfering with each other. We focus on a specific configuration, named the two-user Binary Fading Interference Channel (BFIC) as depicted in Fig. 1.1(a), in which the channel gains at each time instant are either 0 or 1 according to some distribution. The input-output relation of this channel at time instant t is given by

$$Y_i[t] = G_{ii}[t]X_i[t] \oplus G_{\bar{i}i}[t]X_{\bar{i}}[t], \quad i = 1, 2, \tag{1.1}$$

where $\bar{i} = 3 - i$, $G_{ii}[t], G_{\bar{i}i}[t] \in \{0, 1\}$, and all algebraic operations are in $\mathbb{F}_2$.

We define the channel state information at time t to be

$$G[t] \triangleq (G_{11}[t], G_{12}[t], G_{21}[t], G_{22}[t]). \tag{1.2}$$

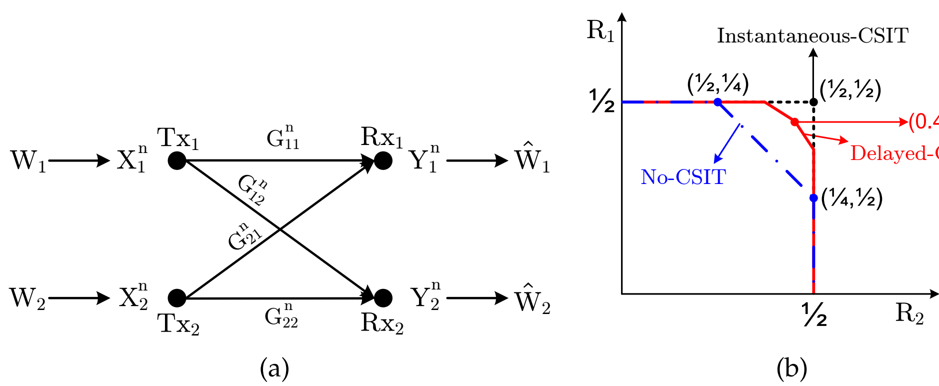


Figure 1.1: *(a) The two-user binary fading interference channel, and (b) the capacity region for three architectures.*

As the main motivation, we study the two-user BFIC as a stepping stone towards understanding the capacity of more complicated fading interference channels with Delayed-CSIT. Lately, the linear deterministic model introduced in [7], has been utilized to bridge the results from deterministic networks into Gaussian networks (*e.g.*, [7, 8, 11, 12, 46, 50, 59]). In the linear deterministic model, there is a non-negative integer representing the channel gain from a transmitter to a receiver. Hence, one can view the binary fading model as a fading interpretation of the linear deterministic model where the non-negative integer associated to each link is either 0 or 1. Furthermore, as demonstrated in [63], the binary fading model provides a simple, yet useful physical layer abstraction for wireless packet networks in which whenever a collision occurs, the receiver can store its received analog signal and utilize it for decoding the packets in future (for example, by successive interference cancellation techniques).

We consider three models for the available channel state information at the transmitters:

(a) Instantaneous-CSIT: In this model, the channel state information G^t is available at each transmitter at time instant t, $t = 1, 2, \ldots, n$;

(b) No-CSIT: In this model, transmitters only know the distribution from which the channel gains are drawn, but not the actual realizations of them;

(c) Delayed-CSIT: In this model, at time instant t, each transmitter has the knowledge of the channel state information up to the previous time instant (*i.e.* G^{t-1}) and the distribution from which the channel gains are drawn, $t = 1, 2, \ldots, n$.

We fully characterize the capacity region of the two-user BFIC under the three assumptions where

$$G_{ij}[t] \overset{d}{\sim} \mathcal{B}(p), \qquad i, j = 1, 2, \tag{1.3}$$

for $0 \leq p \leq 1$, and we define $q = 1 - p$.

The capacity region for three assumptions on the available channel state information at the transmitters is shown in Figure 1.1(b). The most challenging assumption is Delayed-CSIT where transmitters only become aware of CSI when it is already expired. In this scenario, we introduce and discuss several novel coding opportunities, created by outdated CSIT, which can enlarge the achievable rate region. In particular, we propose a new transmission strategy, which is carried on over several phases. Each channel realization creates multiple coding opportunities which can be exploited in the next phases, to improve the rate region. However, we observe that *merging* or *concatenating* some of the opportunities can offer even more gain. To achieve the capacity region, we find the most efficient arrangement of combination, concatenation, and merging of the opportunities, depending on the channel statistics. This can take up to five phases of communication for a two-user channel. For converse arguments, we start with a genie-aided interference channel and show that the problem can be reduced to some particular form of broadcast channels with Delayed-CSIT. We establish a new extremal inequality for the underlying BC, which leads to a tight outer-bound for the original interference channel. The established inequality provides an outer-bound on how much the transmitter in a BC can favor one receiver to the other using Delayed-CSIT (in terms of the entropy of the received signal at the two receivers).

The detailed discussions and proofs in this direction can be found in Chapter 2, and [62, 63, 65, 67].

1.3.2 Capacity results for Gaussian Networks with Delayed Channel State Information

In the first direction, while we study the delay in distributed setting, we limit ourselves to a simple on-off channel model. To get closer to real world settings, we consider the two-user multiple-input single-output (MISO) complex Gaussian broadcast channel (BC) as depicted in Fig. 1.2. At each receiver, the received signal can be expressed as follows.

$$y_1[t] = \mathbf{h}^\top[t]\mathbf{x}[t] + z_1[t],$$
$$y_2[t] = \mathbf{g}^\top[t]\mathbf{x}[t] + z_2[t], \tag{1.4}$$

where $\mathbf{x}[t] \in \mathbb{C}^{2\times 1}$ is the transmit signal subject to average power constraint P, *i.e.* $\mathbb{E}\left[\mathbf{x}^\dagger[t]\mathbf{x}[t]\right] \leq P$ for $P > 0$. The noise processes are independent from the transmit signal and are distributed i.i.d. as $z_k[t] \sim \mathcal{CN}(0, 1)$.

We focus on the impact of delayed CSIT at finite SNR regime, as opposed to prior works on the asymptotic DoF analysis. While there is strong body of work on the broadcast channels with perfect channel state information (see [40, 72, 73]), no capacity result has been reported for the delayed CSIT scenario.

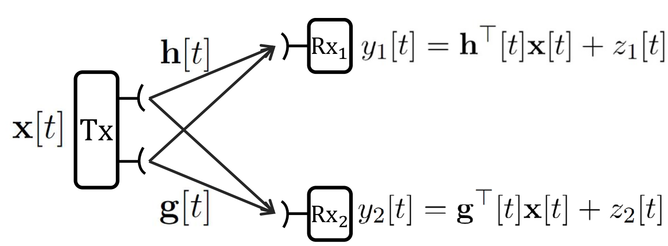


Figure 1.2: Two-user Multiple-Input Single-Output (MISO) Complex Gaussian Broadcast Channel.

We provide a constant-gap approximation of the capacity region of the two-user MISO BC with delayed CSIT. We obtain an achievable scheme and an outer-bound on the capacity region, and analytically show that they are within constant number of bits, for all values of transmit power. The detailed discussions and proofs in this direction can be found in Chapter 3, and [58, 66].

1.3.3 Communication with Local Network Views

In this case, we would like to understand the effect of local network knowledge on network capacity and the corresponding optimal transmission strategies. This local view can be different from node to node and can even be inaccurate at some nodes. Figure 1.3 depicts the issue of limited view in the networks. We propose a new tractable framework that enables us to understand how to manage interference and how to optimally flow information with only local network views.

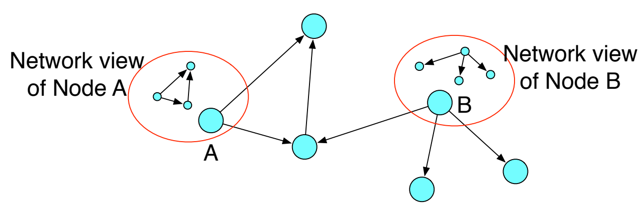


Figure 1.3: *Depiction of a typical scenario, where Nodes A and B have to rely on different partial views about the network to make their transmission decisions.*

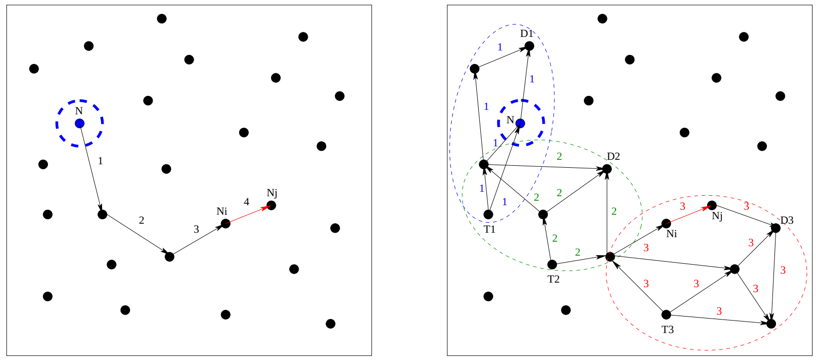
There is a general consensus that our current understanding of networks is so limited that attacking the problem directly is nearly impossible. Hence to make progress, we will focus on one of the dominant issues, *i.e.* the role of information about current state of the network on network capacity and the associated optimal strategies. To capture the dominant effects of the problem, we propose to model a node's view of the network with a class of distance-based accuracy models using network coordination protocols as a source of network state estimations.

Denote the nodes in the network as N_i. We say that a node pair (N_i, N_j), is connected if the channel state distribution is such that, the link between the two nodes is not always in a deep fade. The network topology T, is the graph where two nodes are connected when the channel between them satisfies the above condition. We will consider two different models for the local view at node N.

(a) *Hop-Based Model:* In this setting, we use the hop-distance between node N and link (N_i, N_j), in order to figure out whether the node is aware of the channel gain or not. This model is well suited for single-layer networks [4]. See figure 1.4(a) for a depiction.

(b) *Route-Based Model:* We will map each link (N_i, N_j), to those source-destination pairs that have (N_i, N_j) in one the routes between them. We further define the source-destination adjacency graph, P, which consists of n nodes corresponding to the n source-destination pairs in the network. Two source-destination pairs, (T_i, D_i) and (T_j, D_j), are called adjacent (i.e., there is an edge between them in P) if there is either a route from T_i to D_j or a route from T_j to D_i. We can now use the hop-distance between the source-destination pair (in P) corresponding to the node N and link (N_i, N_j). This model is inspired by learning algorithms that provide us with knowledge about the routes in the network and it is suitable for multi-layer networks [60]. See Figure 1.4(b) for a depiction.

One of our main objectives, is to study the loss in the network capacity due to



(a) Hop-based model for local view.

(b) Route-based model for local view.

Figure 1.4: *Illustration of two different model for network state at node N.*

incomplete knowledge at the nodes. In a recent work on single layer networks with partial information [4], a novel approach has been adopted by characterizing normalized sum-capacity in these networks, which is defined as the best guaranteed ratio of the sum-rate with local view to the sum-capacity with full global view at each node. We borrow this useful definition and we focus on multi-layer networks, where source-destination pairs are multi hops away. This significantly adds up to the complexity of the problem and therefore, the challenges we face. Our goal is to characterize normalized sum-capacity as a function of partial information available at different nodes, based on route-based model. See Figure 1.5 for a depiction.

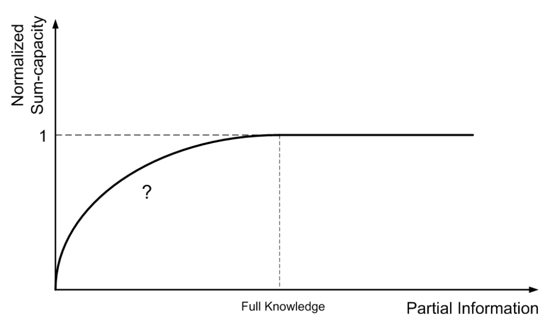
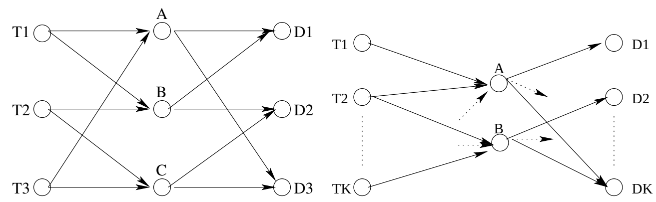


Figure 1.5: Normalized sum-capacity as a function of partial information available.



(a) (b)

Figure 1.6: *(a) An example of a two-layer network in which coding is required , and (b) the two-layer $K \times 2 \times K$ network.*

It is a well established fact that network coding is an effective technique to increase the spectral efficiency of the networks. However, a major problem is how to effectively exploit network coding with partial network-state information. The problem becomes really challenging when we are dealing with multi-commodity flows, such as the one depicted in figure 1.6(a). In our recent works [60] and [61], we have proved that by utilizing repetition coding at the transmitters and linear network coding at the relays, it is possible to achieve normalized sum-rate of $\frac{1}{2}$, which can be shown to be the maximum achievable normalized sum-rate with the assumption of 0-hop route-based partial information, *i.e.* each transmitter is only aware of network topology and all the channel gains of the links that are on a route to its destination. We reveal a deep connection between network geometry and optimal schemes with partial knowledge.

The detailed discussions and proofs in this direction can be found in Chapter 4, and [60, 61].

1.3.4 Communication with Rate-Limited Feedback Links

Finally, there is one subtle aspect of feedback channels that seems to be neglected by most researchers. In prior works, whether the feedback channel is used to provide transmitters with channel state information or previously transmitted signals, it has mostly been assumed that the feedback links have infinite capacity. A more realistic feedback model is one where feedback links are *rate-limited*. In this direction, we study the impact of the rate-limited feedback in the context of the two-user IC. We consider a two-user interference channel (IC) where a noiseless rate-limited feedback link is available from each receiver to its corresponding transmitter. See Figure 5.1. The feedback link from receiver k to transmitter k is assumed to have a capacity of C_{FBk}, $k = 1, 2$.

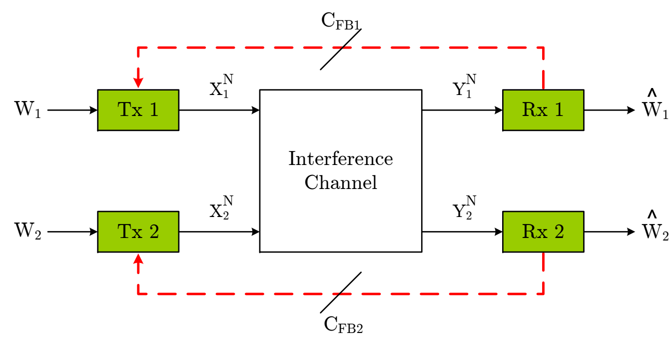


Figure 1.7: Two-user interference channel with rate-limited feedback.

We focus on two fundamental questions: (1) what is the maximum capacity gain that can be obtained with access to feedback links at a specific rate of C_{FB}? (2) what are the transmission strategies that exploit the available feedback links efficiently? Specifically, we address these questions under three channel models: the El Gamal-Costa deterministic model [16], the linear deterministic model of [7], and the Gaussian model.

Under the El Gamal-Costa deterministic model, we derive inner-bounds and outer-bounds on the capacity region. As a result, we show that the capacity region can be enlarged using feedback by at most the amount of available feedback, *i.e.*, "one bit of feedback is at most worth one bit of capacity". We show that our inner-bounds and outer-bounds match under the linear deterministic model, thus establishing the capacity region. For the Gaussian model and symmetric channel gains, we show that the gap between the achievable sum-rate and the outer-bounds can be bounded by constant number of bits independent of the channel gains. The detailed discussions and proofs in this direction can be found in Chapter 5, and [57, 59].

CHAPTER 2

COMMUNICATION WITH DELAYED CHANNEL STATE INFORMATION

2.1 Introduction

In this chapter, our goal is to shed light on fundamental limits of communications with Delayed-CSIT in interference channels. We consider a two-user interference channel as illustrated in Figure 2.1. In this network, the channel gains at each time instant are either 0 or 1 according to some Bernoulli distribution, and are independent from each other and *over time*. The input and output signals are also in the binary field and if two signals arrive simultaneously at a receiver, then the receiver obtains the XOR of them. We shall refer to this network as the two-user Binary Fading Interference Channel (BFIC).

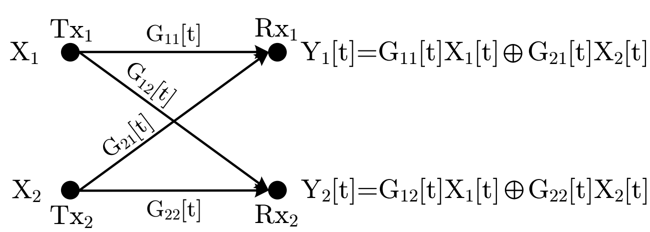


Figure 2.1: Binary fading channel model for a two-user interference channel. The channel gains, the transmitted signals and the received signals are in the binary field. The channel gains are distributed as i.i.d. Bernoulli random variables. The channel gains are independent across time so that the transmitters cannot predict future based on the past channel state information.

We fully characterize the capacity region of the two-user BFIC with Delayed-CSIT. We introduce and discuss several novel coding opportunities, created by outdated CSIT, which can enlarge the achievable rate region. In particular, we propose a new transmission strategy, which is carried on over several phases.

Each channel realization creates multiple coding opportunities which can be exploited in the next phases, to improve the rate region. However, we observe that *merging* or *concatenating* some of the opportunities can offer even more gain. To achieve the capacity region, we find the most efficient arrangement of combination, concatenation, and merging of the opportunities, depending on the channel statistics. This can take up to five phases of communication for a two-user channel. For converse arguments, we start with a genie-aided interference channel and show that the problem can be reduced to some particular form of broadcast channels with Delayed-CSIT. We establish a new extremal inequality for the underlying BC, which leads to a tight outer-bound for the original interference channel. The established inequality provides an outer-bound on how much the transmitter in a BC can favor one receiver to the other using Delayed-CSIT (in terms of the entropy of the received signal at the two receivers).

2.2 Problem Setting

We consider the two-user Binary Fading Interference Channel (BFIC) as illustrated in Figure 2.2 and defined below.

Definition 2.1 *The two-user Binary Fading Interference Channel includes two transmitter-receiver pairs in which the channel gain from transmitter* Tx_i *to receiver* Rx_j *at time instant* t *is denoted by* $G_{ij}[t]$, $i, j \in \{1,2\}$. *The channel gains are either* 0 *or* 1 (i.e. $G_{ij}[t] \in \{0,1\}$)*, and they are distributed as independent Bernoulli random variables (independent from each other and* over time*). We consider the homogeneous setting where*

$$G_{ij}[t] \overset{d}{\sim} \mathcal{B}(p), \qquad i, j = 1, 2, \tag{2.1}$$

for $0 \leq p \leq 1$, and we define $q = 1 - p$.

At each time instant t, the transmit signal at Tx_i is denoted by $X_i[t] \in \{0, 1\}$, and the received signal at Rx_i is given by

$$Y_i[t] = G_{ii}[t]X_i[t] \oplus G_{\bar{i}i}[t]X_{\bar{i}}[t], \quad i = 1, 2, \tag{2.2}$$

where the summation is in $\mathbb{F}_2$.

Definition 2.2 *We define the channel state information (CSI) at time instant t to be the quadruple*

$$G[t] \stackrel{\triangle}{=} (G_{11}[t], G_{12}[t], G_{21}[t], G_{22}[t]). \tag{2.3}$$

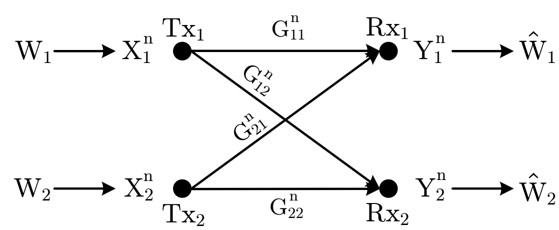


Figure 2.2: Two-user Binary Fading Interference Channel (BFIC). The channel gains, the transmit and the received signals are in the binary field. The channel gains are distributed as i.i.d. Bernoulli random variables.

We use the following notations in this chapter. We use capital letters to denote random variables (RVs), *e.g.*, $G_{ij}[t]$ is a random variable at time instant t. Furthermore for a natural number k, we set

$$G^k = [G[1], G[2], \ldots, G[k]]^\top. \tag{2.4}$$

Finally, we set

$$G_{ii}^t X_i^t \oplus G_{\bar{i}i}^t X_{\bar{i}}^t = [G_{ii}[1]X_i[1] \oplus G_{\bar{i}i}[1]X_{\bar{i}}[1], \ldots, G_{ii}[t]X_i[t] \oplus G_{\bar{i}i}[t]X_{\bar{i}}[t]]^\top. \tag{2.5}$$

In this chapter, we consider three models for the available channel state information at the transmitters:

(a) Instantaneous-CSIT: In this model, the channel state information G^t is available at each transmitter at time instant t, $t = 1, 2, \ldots, n$;

(b) No-CSIT: In this model, transmitters only know the distribution from which the channel gains are drawn, but not the actual realizations of them;

(c) Delayed-CSIT: In this model, at time instant t, each transmitter has the knowledge of the channel state information up to the previous time instant (*i.e.* G^{t-1}) and the distribution from which the channel gains are drawn, $t = 1, 2, \ldots, n$.

We assume that the receivers have instantaneous knowledge of the CSI. Consider the scenario in which Tx_i wishes to reliably communicate message $\mathrm{W}_i \in \{1, 2, \ldots, 2^{nR_i}\}$ to Rx_i during n uses of the channel, $i = 1, 2$. We assume that the messages and the channel gains are *mutually* independent and the messages are chosen uniformly. For each transmitter Tx_i, let message W_i be encoded as X_i^n using the encoding function $f_i(.)$, which depends on the available CSI at Tx_i. Receiver Rx_i is only interested in decoding W_i, and it will decode the message using the decoding function $\widehat{\mathrm{W}}_i = g_i(Y_i^n, G^n)$. An error occurs when $\widehat{\mathrm{W}}_i \neq \mathrm{W}_i$. The average probability of decoding error is given by

$$\lambda_{i,n} = \mathbb{E}[P[\widehat{\mathrm{W}}_i \neq \mathrm{W}_i]], \qquad i = 1, 2, \tag{2.6}$$

and the expectation is taken with respect to the random choice of the transmitted messages W_1 and W_2. A rate tuple (R_1, R_2) is said to be achievable, if there exists encoding and decoding functions at the transmitters and the receivers respectively, such that the decoding error probabilities $\lambda_{1,n}, \lambda_{2,n}$ go to zero as n goes to infinity. The capacity region is the closure of all achievable rate tuples.

In addition to the setting described above, we consider a separate scenario in which an output feedback (OFB) link is available from each receiver to its corresponding transmitter[1]. More precisely, we consider a noiseless feedback link of infinite capacity from each receiver to its corresponding transmitter.

Due to the presence of output feedback links, the encoded signal $X_i[t]$ of transmitter Tx_i at time t, would be a function of its own message, previous output sequence at its receiver, and the available CSIT. For instance, with Delayed-CSIT and OFB, we have

$$X_i[t] = f_i[t](\mathrm{W}_i, Y_i^{t-1}, G^{t-1}), \quad i = 1, 2. \tag{2.7}$$

As stated in the introduction, our goal is to understand the impact of the channel state information and the output feedback, on the capacity region of the two-user Binary Fading Interference Channel. Towards that goal, we consider several scenarios about the availability of the CSIT and the OFB. For all scenarios, we provide exact characterization of the capacity region. In the next section, we present the main results of the chapter.

2.3 Statement of Main Results

We basically focus on the following scenarios about the availability of the CSI and the OFB: (1) Delayed-CSIT and no OFB; (2) Delayed-CSIT and OFB; and (3)

[1] As we will see later, our result holds for the case in which output feedback links are available from each receiver to *both* transmitters.

Instantaneous-CSIT and OFB. In order to illustrate the results, we first establish the capacity region of the two-user BFIC with No-CSIT and Instantaneous-CSIT as our benchmarks.

2.3.1 Benchmarks

Our base line is the scenario in which there is no output feedback link from the receivers to the transmitters, and we assume the No-CSIT model. In other words, the only available knowledge at the transmitters is the distribution from which the channel gains are drawn. In this case, it is easy to see that for any input distribution, the two received signals are *statistically* the same, hence

$$\begin{aligned} I\left(X_1^n; Y_1^n | G^n\right) &= I\left(X_1^n; Y_2^n | G^n\right), \\ I\left(X_2^n; Y_1^n | G^n\right) &= I\left(X_2^n; Y_2^n | G^n\right). \end{aligned} \tag{2.8}$$

Therefore, the capacity region in this case, $C^{\text{No-CSIT}}$, is the same as the intersection of the capacity region of the multiple-access channels (MAC) formed at either of the receivers:

$$C^{\text{No-CSIT}} = \begin{cases} 0 \leq R_i \leq p, & i = 1, 2, \\ R_1 + R_2 \leq 1 - q^2. & \end{cases} \tag{2.9}$$

The other extreme point on the available CSIT is the Instantaneous-CSIT model. The capacity region in this case is given in the following theorem which is proved in Appendices A.1 and A.2.

Theorem 2.1 [Capacity Region with Instantaneous-CSIT] *The capacity region of the two-user Binary Fading IC with Instantaneous-CSIT (and no output feedback),*

C^{ICSIT}, is the set of all rate tuples (R_1, R_2) satisfying

$$C^{\text{ICSIT}} = \begin{cases} 0 \leq R_i \leq p, & i = 1, 2, \\ R_1 + R_2 \leq 1 - q^2 + pq. \end{cases} \tag{2.10}$$

Remark 2.1 *Comparing the capacity region of the two-user BFIC with No-CSIT (2.9) and Instantaneous-CSIT (2.10), we observe that the bounds on individual rates remain unchanged while the sum-rate outer-bound is increased by pq. This increase can be intuitively explained as follows. The outer-bound of $1 - q^2$ corresponds to the fraction of time in which at least one of the links to each receiver is equal to 1. Therefore, this outer-bound corresponds to the fraction of time that each receiver gets "useful" signal. This is tight with No-CSIT since each receiver should be able to decode both messages. However, once we move to Instantaneous-CSIT, we can send a private message to one of the receivers by using those time instants in which the link from the corresponding transmitter to that receiver is equal to 1, but that transmitter is not interfering with the other receiver. This corresponds to pq fraction of the time.*

Now that we have covered the benchmarks, we are ready to present our main results.

2.3.2 Main Results

As first step, we consider the Delayed-CSIT model. In this case, the following theorem establishes our result.

Theorem 2.2 [Capacity Region with Delayed-CSIT] *The capacity region of the two-user Binary Fading IC with Delayed-CSIT (and no output feedback), C^{DCSIT}, is the* *set of all rate tuples* (R_1, R_2) *satisfying*

$$C^{\mathrm{DCSIT}} = \begin{cases} 0 \leq R_i \leq p, & i = 1, 2, \\ R_i + (1+q) R_{\bar{i}} \leq p\,(1+q)^2\,, & i = 1, 2. \end{cases} \tag{2.11}$$

Remark 2.2 *Comparing the capacity region of the two-user BFIC with Delayed-CSIT (2.11) and Instantaneous-CSIT (2.10), we can show that for* $0 \leq p \leq \left(3 - \sqrt{5}\right)/2$*, the two regions are equal. However, for* $\left(3 - \sqrt{5}\right)/2 < p < 1$*, the capacity region of the two-user BFIC with Delayed-CSIT is strictly smaller than that of Instantaneous-CSIT. Moreover, we can show that the capacity region of the two-user BFIC with Delayed-CSIT is strictly larger than that of No-CSIT (except for* $p = 0$ *or* 1*).*

Furthermore, since the channel state information is acquired through the feedback channel, it is also important to understand the impact of output feedback on the capacity region of the two-user BFIC with Delayed-CSIT. In the study of feedback in wireless networks, one other direction is to consider the transmitter cooperation created through the output feedback links. In this context, it is well-known that feedback does not increase the capacity of discrete memoryless point-to-point channels [47]. However, feedback can enlarge the capacity region of multi-user networks, even in the most basic case of the two-user memoryless multiple-access channel [20, 39]. In [50, 59], the feedback capacity of the two-user Gaussian IC has been characterized to within a constant number of bits. One consequence of these results is that output feedback can provide an unbounded capacity increase. This is in contrast to point-to-point and multiple-access channels where feedback provides no gain and bounded gain respectively. In this work, we consider the scenario in which an output feedback link is available from each receiver to its corresponding transmitter on top of the delayed knowledge of the channel state information as depicted in Figure 2.3(a).

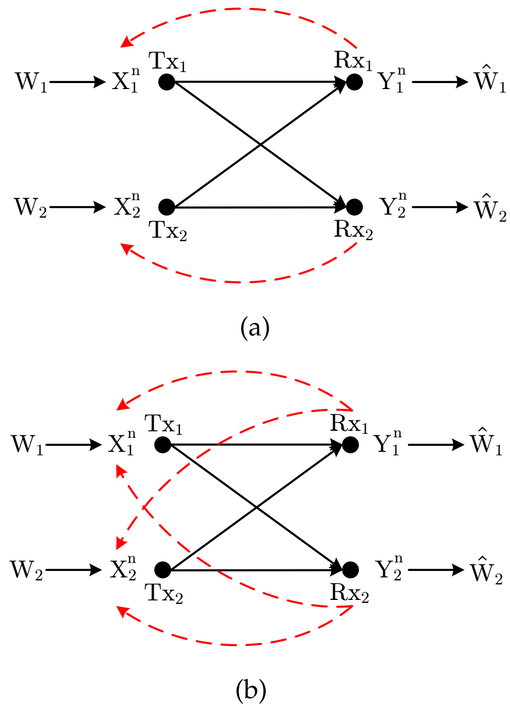


Figure 2.3: *Two-user Binary Fading Interference Channel: (a) with output feedback links from each receiver to its corresponding transmitter. In this setting, the transmit signal of* Tx_i *at time instant* t*, would be a function of the message* W_i*, the available CSIT, and the output sequences* Y_i^{t-1}*,* $i = 1, 2$*; and (b) with output feedback links from each receiver to both transmitters. In this setting, the transmit signal of* Tx_i *at time instant* t*, would be a function of the message* W_i*, the available CSIT, and the output sequences* Y_1^{t-1}, Y_2^{t-1}*,* $i = 1, 2$*.*

In the presence of output feedback and Delayed-CSIT, we have the following result.

Theorem 2.3 [Capacity Region with Delayed-CSIT and OFB] *For the two-user binary IC with Delayed-CSIT and OFB, the capacity region* $\mathcal{C}^{\mathrm{DCSIT,OFB}}$*, is given by*

$$\mathcal{C}^{\mathrm{DCSIT,OFB}} = \left\{ R_1, R_2 \in \mathbb{R}^+ \ s.t. \ R_i + (1+q)R_{\bar{i}} \leq p(1+q)^2, \ i = 1, 2 \right\}. \quad (2.12)$$

Remark 2.3 *The outer-bound on the capacity region with only Delayed-CSIT (2.11) is in fact the intersection of the outer-bounds on the individual rates* (i.e. $R_i \leq p$, $i = 1, 2$) *and the capacity region with Delayed-CSIT and OFB (2.12). Therefore, the impact of OFB is to remove the constraints on individual rates. This can be intuitively explained by noting that OFB creates new path to flow information from each transmitter to its corresponding receiver* (e.g., $\mathsf{Tx}_1 \rightarrow \mathsf{Rx}_2 \rightarrow \mathsf{Tx}_2 \rightarrow \mathsf{Rx}_1$)*. This opportunity results in elimination of the individual rate constraints in this case.*

Remark 2.4 *As we will see in Section 2.8, same outer-bounds hold in the presence of* global output feedback *where output feedback links are available from each receiver to both transmitters, see Figure 2.3(b). Therefore, the capacity region of two user binary IC with Delayed-CSIT and global output feedback is the same as the capacity region described in (2.12). This implies that in this case, global output feedback does not provide new coding opportunities, nor does it enhance the existing ones. Similar observation has been made in the context of MIMO Interference Channels [52, 70], even though the coding opportunities in Binary IC and MIMO IC are not the same.*

Finally, we present our result for the case of Instantaneous-CSIT and output feedback. Note that in this scenario, although transmitters have instantaneous knowledge of the channel state information, the output signals are available at the transmitters with unit delay. This scenario corresponds to a slow-fading channel where output feedback links are available from the receivers to the transmitters.

Theorem 2.4 [Capacity Region with Instantaneous-CSIT and OFB] *For the two-user binary IC with Instantaneous-CSIT and OFB, the capacity region* $\mathcal{C}^{\text{ICSIT,OFB}}$*,* *is the set of all rate tuples* (R_1, R_2) *satisfying*

$$C^{\text{ICSIT,OFB}} = \begin{cases} 0 \leq R_i \leq 1 - q^2, & i = 1, 2, \\ R_1 + R_2 \leq 1 - q^2 + pq. \end{cases} \tag{2.13}$$

Remark 2.5 *Comparing the capacity region of the two-user BFIC with Instantaneous-CSIT, with OFB (2.13) or without OFB (2.10), we observe that the outer-bound on the sum-rate remains unchanged. However, the bounds on individual rates are further increased to* $1 - q^2$*. Similar to the previous remark, this is again due to the additional communication path provided by OFB from each transmitter to its intended receiver. However, since the outer-bound on sum-rate with Instantaneous-CSIT and OFB (2.13) is higher than that of Delayed-CSIT and OFB (2.12), the bounds on individual rates cannot be eliminated.*

The proof of the results is organized as follows. The proof of Theorem 2.2 is presented in Sections 2.5 and 2.6. The proof of Theorem 2.3 is presented in Sections 2.7 and 2.8, and finally, the proof of Theorem 2.4 is presented in Sections 2.9 and 2.10. We end this section by illustrating our main results via an example in which $p = 0.5$.

2.3.3 Illustration of Main Results for $p = 0.5$

For this particular value of the channel parameter, the capacity region with Delayed-CSIT and Instantaneous-CSIT with or without output feedback is given in Table 2.1, and Figure 2.4 illustrates the results presented in this table. We notice the following remarks.

Table 2.1: Illustration of our main results through an example in which $p = 0.5$.

	Capacity Region with Delayed-CSIT	Capacity Region with Instantaneous-CSIT
No-OFB	$\begin{cases} R_i \leq \frac{1}{2} \\ R_i + \frac{3}{2}R_{\bar{i}} \leq \frac{9}{8} \end{cases}$	$\begin{cases} R_1 \leq \frac{1}{2} \\ R_2 \leq \frac{1}{2} \end{cases}$
OFB	$\begin{cases} R_1 + \frac{3}{2}R_2 \leq \frac{9}{8} \\ \frac{3}{2}R_1 + R_2 \leq \frac{9}{8} \end{cases}$	$\begin{cases} R_i \leq \frac{3}{4} \\ R_1 + R_2 \leq 1 \end{cases}$

Remark 2.6 *Note that for $p = 0.5$, we have*

$$\mathcal{C}^{\mathrm{No-CSIT}} \subset \mathcal{C}^{\mathrm{DCSIT}} \subset \mathcal{C}^{\mathrm{ICSIT}}.$$

In other words, the capacity region with Instantaneous-CSIT is strictly larger than that of Delayed-CSIT, which is in turn strictly larger than the capacity region with No-CSIT. Moreover, we have

$$\mathcal{C}^{\mathrm{DCSIT,OFB}} \subset \mathcal{C}^{\mathrm{ICSIT,OFB}},$$

meaning that the instantaneous knowledge of the CSIT enlarges the capacity region of the two-user BFIC with OFB compared to the case of Delayed-CSIT.

Remark 2.7 *In Figure 2.4(c), we have illustrated the capacity region with Delayed-CSIT, with and without output feedback. First, we observe that OFB enlarges the capacity region. Second, we observe that the optimal sum-rate point is the same for $p = 0.5$. However, this is not always the case. In fact, for some values of p, output feedback can even increase the optimal sum-rate. Using the results of Theorem 2.2 and Theorem 2.3,*

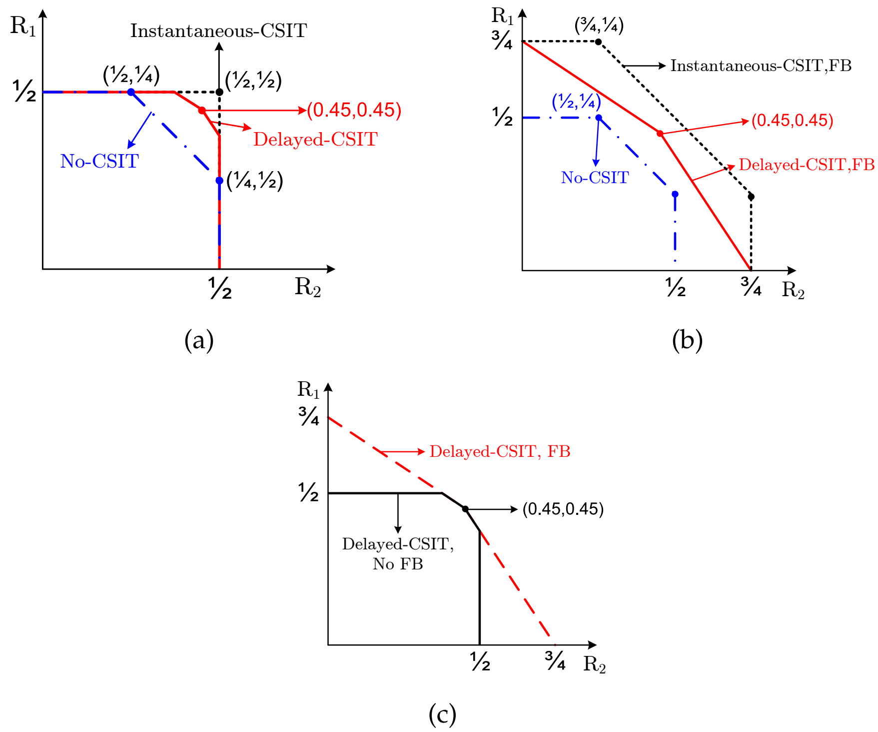


Figure 2.4: *Two-user Binary Fading IC: (a) the capacity region with No-CSIT, Delayed-CSIT, and Instantaneous-CSIT, without OFB; (b) the capacity region with No-CSIT, Delayed-CSIT, and Instantaneous-CSIT, with OFB; and (c) the capacity region with Delayed-CSIT, with and without output feedback.*

we have plotted the sum-rate capacity of the two-user Binary Fading IC with and without OFB for the Delayed-CSIT model in Figure 2.5. Note that for $0 < p < \left(3 - \sqrt{5}\right)/2$, *the sum-rate capacity with OFB is strictly larger than the no OFB scenario.*

Remark 2.8 *Comparing the capacity region of the two-user BFIC with Instantaneous-CSIT, with OFB (2.13) or without OFB (2.10), we observe that OFB enlarges the capacity region. Moreover, similar to the Delayed-CSIT scenario, the optimal sum-rate point is the same for* $p = 0.5$. *Again, this is not always the case. In fact, for* $0 < p < 0.5$, *output feedback can even increase the optimal sum-rate. Using the results of Theorem 2.1 and Theorem 2.4, we have plotted the sum-rate capacity of the two-user Binary Fading*

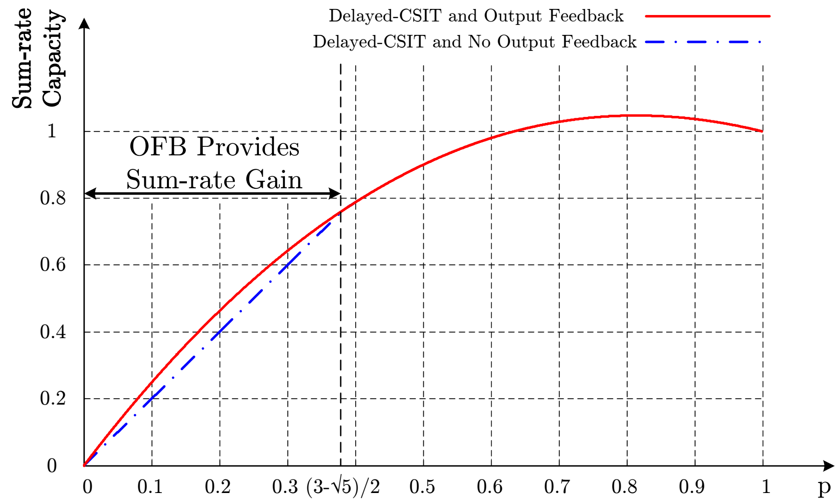


Figure 2.5: The sum-rate capacity of the two-user BFIC with Delayed-CSIT, with and without output feedback. For $0 < p < \left(3 - \sqrt{5}\right)/2$, the sum-rate capacity with OFB is strictly larger than the scenario where no OFB is available.

IC with and without OFB in Figure 2.6.

Remark 2.9 *In Figures 2.5 and 2.6, we have identified the range of p for which output feedback provides sum-rate gain. Basically, when the sum-rate capacity without OFB is dominated by the capacity of the direct links* (i.e. $2p$)*, and the additional communication paths created by the means of output feedback links help increase the optimal sum-rate.*

Remark 2.10 *While our capacity results in Theorems 2.2 and 2.3 are for binary fading interference channels, in [63], we have shown how they can also be utilized to obtain capacity results for a class of wireless packet networks.*

In the following section, we present the main ideas that we incorporate in this work.

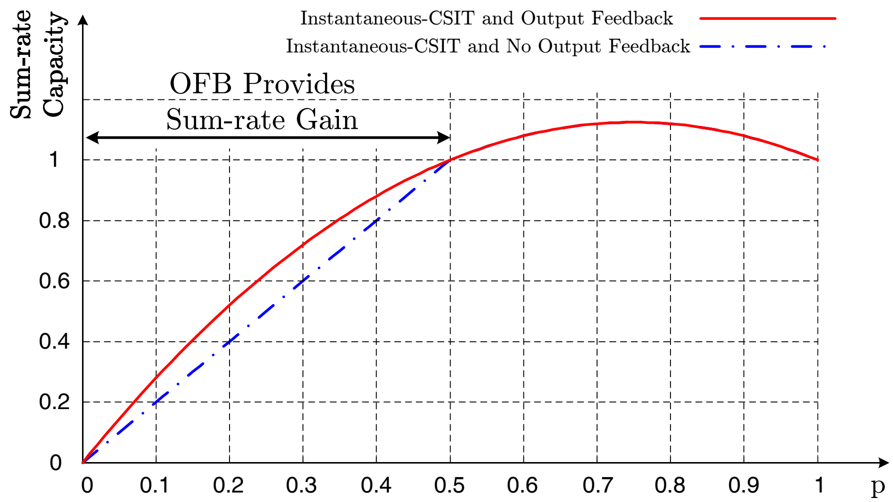


Figure 2.6: The sum-rate capacity of the two-user BFIC with Instantaneous-CSIT, with and without output feedback. For $0 < p < 0.5$, the sum-rate capacity with OFB is strictly larger than the scenario where no OFB is available.

2.4 Overview of Key Ideas

Our goal in this section is to present the key techniques we use in this work both for achievability and converse purposes. Although we will provide detailed explanation of the achievability strategy and converse proofs for all different scenarios, we found it instructive to elaborate the main ideas through several clarifying examples. Furthermore, the coding opportunities introduced in this section can be applicable to DoF analysis of wireless networks with linear schemes (Section III.A of [30]) or interference management in packet collision networks (Section IV of [63]).

2.4.1 Achievability Ideas with Delayed-CSIT

As we have described in Section 2.2, the channel gains are independent from each other and over time. This way, transmitters cannot use the delayed knowledge of the channel state information to predict future. However, this information can still be very useful. In particular, Delayed-CSIT allows us to evaluate the contributions of the desired signal and the interference at each receiver in the past signaling stages and exploit it as available side information for future communication.

Interference-free Bits

Using Delayed-CSIT transmitters can identify previously transmitted bits such that if retransmitted, they do not create any further interference. The following examples clarify this idea.

Example 1 [Creating interference channels with side information]: Suppose at a time instant, each one of the transmitters simultaneously sends one data bit. The bits of Tx_1 and Tx_2 are denoted by a_1 and b_1 respectively. Later, using Delayed-CSIT, transmitters figure out that only the cross links were equal to 1 at this time instant as shown in Figure 2.7(a). This means that in future, transmission of these bits will no longer create interference at the unintended receivers.

Example 2 [Creating interference channels with swapped receivers and side information]: Assume that at a time instant, transmitters one and two simultaneously send data bits, say a_2 and b_2 respectively. Again through Delayed-CSIT, transmitters realize that all links except the link between Tx_1 and Rx_2 were equal

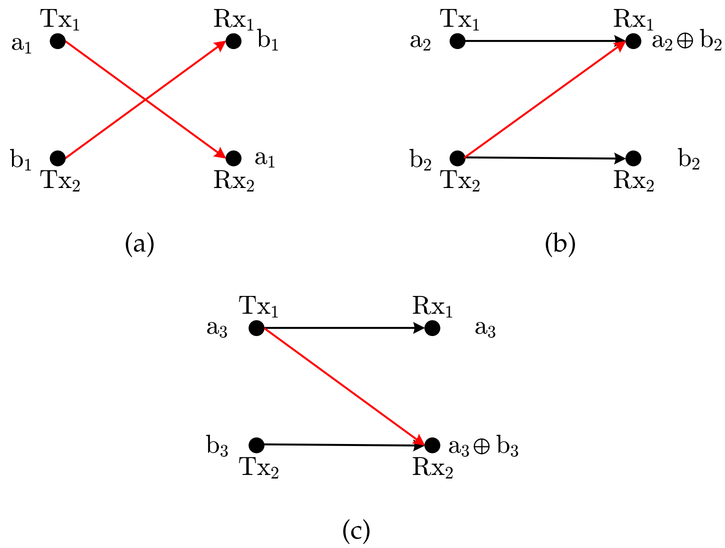


Figure 2.7: *Achievability ideas with Delayed-CSIT: (a) via Delayed-CSIT transmitters can figure out which bits are already known at the unintended receivers, transmission of these bits will no longer create interference at the unintended receivers; and (b) to decode the bits, it is sufficient that* Tx_2 *provides* Rx_1 *with* b_2 *while this bit is available at* Rx_2*; (c) similar to (b). Note that in (b) and (c) the intended receivers are swapped.*

to 1, see Figure 2.7(b). In a similar case, assume that at another time instant, transmitters one and two send data bits a_3 and b_3 at the same time. Through Delayed-CSIT, transmitters realize that all links except the link between Tx_2 and Rx_1 were connected, see Figure 2.7(c). Then it is easy to see that to successfully finish delivering these bits, it is enough that Tx_1 sends a_3 to Rx_2, while this bit is already available at Rx_1; and Tx_2 sends b_2 to Rx_1, while it is already available at Rx_2. Note that here the intended receivers are *swapped*.

Remark 2.11 *As described in Examples 1 and 2, an interference free bit can be retransmitted without worrying about creating interference at the unintended receiver. These bits can be transferred to a sub-problem, where in a two-user interference channel,* Rx_i *has apriori access to* $W_{\bar{i}}$ *as depicted in Figure 2.8,* $i = 1, 2$. *Since there will be no* *interference in this sub-problem, such bits can be communicated at higher rates.*

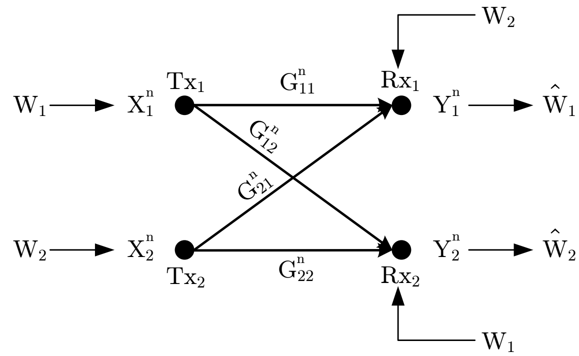


Figure 2.8: Interference channel with side information: the capacity region with no, delayed, or instantaneous CSIT is the same.

Bits of Common Interest

Transmitters can use the delayed knowledge of the channel state information to identify bits that are of interest of both receivers. Below, we clarify this idea through several examples.

Example 3 [Opportunistic creation of bits of common interest]: Suppose at a time instant, each one of the transmitters sends one data bit, say a_4 and b_4 respectively. Later, using Delayed-CSIT, transmitters figure out that all links were equal to 1. In this case, both receivers have an equation of the transmitted bits, see Figure 2.9(a). Now, we notice that it is sufficient to provide either of the transmitted bits, a_4 or b_4, to both receivers rather than retransmitting both bits. We refer to such bits as bits of common interest. Since such bits are useful for both receivers, they can be transmitted more efficiently.

Remark 2.12 (Pairing bits of common interest) *We note that in Example 3, one of the transmitters takes the responsibility of delivering one bit of common interest to both*

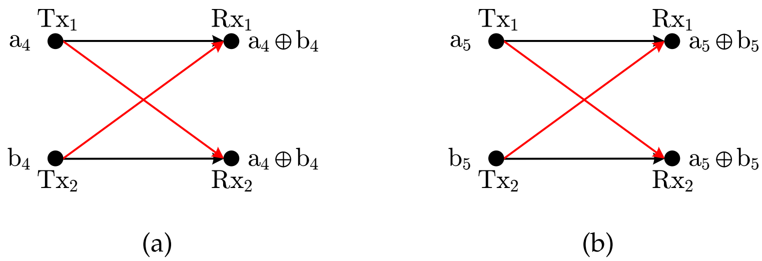


Figure 2.9: *In each case, it is sufficient to provide only one of the transmitted bits to both receivers. We refer to such bits as bits of common interest.*

receivers. To improve the performance, we can pair this problem with another similar problem as follows. Assume that in another time instant, each one of the transmitters sends one data bit, say a_5 *and* b_5 *respectively. Later, transmitters figure out that all links were equal to* 1*, see Figure 2.9(b). In this case, similar to Example 3, one of the bits* a_5 *and* b_5*, say* b_5*, can be chosen as the bit of common interest. Now we can pair cases depicted in Figures 2.9(a) and 2.9(b). Then transmitters can simultaneously send bits* a_4 *and* b_5*, to both receivers. With this pairing, we take advantage of all four links to transmit information.*

Remark 2.13 (Pairing ICs with side information (pairing Type-I)) *The advantage of interference channels with side information, explained in Examples 1 and 2, is that due to the side information, there is no interference involved in the problem. The downside is that half of the links in the channel become irrelevant and unexploited. More precisely, the cross links in Example 1 and the direct links in Example 2 are not utilized to increase the rate. Here we show that these two problems can be paired together to form an efficient two-multicast problem through creating bits of common interest. Referring to Figure 2.7, one can easily verify that it is enough to deliver* $a_1 \oplus a_3$ *and* $b_1 \oplus b_2$ *to both receivers. For instance, if* $a_1 \oplus a_3$ *and* $b_1 \oplus b_2$ *are available at* Rx_1*, it can remove* b_1 *from* $b_1 \oplus b_2$ *to decode* b_2*, then using* b_2 *and* $a_2 \oplus b_2$ *it can decode* a_2*; finally, using* a_3 *and* $a_1 \oplus a_3$ *it can decode* a_1*. Indeed, bit* $a_1 \oplus a_3$ *available at* Tx_1*, and bit* $b_1 \oplus b_2$ *available at*

Tx_2, are bits of common interest and can be transmitted to both receivers simultaneously in the efficient two-multicast problem as depicted in Figure 2.10. We note that for the two-multicast problem, the capacity region with no, delayed, or instantaneous CSIT is the same. We shall refer to this pairing as pairing **Type-I** *throughout the paper.*

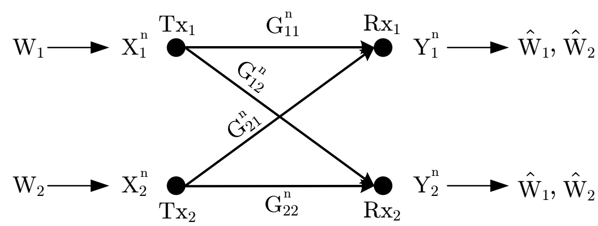
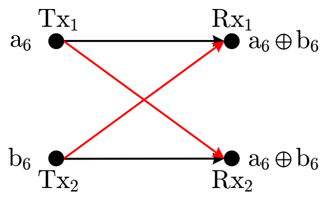


Figure 2.10: Two-multicast network. Transmitter Tx_i wishes to reliably communicate message W_i to both receivers, $i = 1, 2$. The capacity region with no, delayed, or instantaneous CSIT is the same.

Example 4 [Pairing interference-free bits with bits of common interest to create a two-multicast problem (pairing Type-II)]: Suppose at a time instant, each one of the transmitters sends one data bit, say a_6 and b_6 respectively. Later, using Delayed-CSIT, transmitters figure out that all links were equal to 1, see Figure 2.11(a). In another time instant, each one of the transmitters sends one data bit, say a_7 and b_7 respectively. Later, transmitters figure out that only the cross links were equal to 1, see Figure 2.11(b). Now, we observe that providing $a_6 \oplus a_7$ and $b_6 \oplus b_7$ to both receivers is sufficient to decode the bits. For instance if Rx_1 is provided with $a_6 \oplus a_7$ and $b_6 \oplus b_7$, then it will use b_7 to decode b_6, from which it can obtain a_6, and finally using a_6 and $a_6 \oplus a_7$, it can decode a_7. Thus, bit $a_6 \oplus a_7$ available at Tx_1, and bit $b_6 \oplus b_7$ available at Tx_2, are bits of common interest and can be transmitted to both receivers simultaneously in the efficient two-multicast problem. We shall refer to this pairing as pairing **Type-II** throughout the paper.



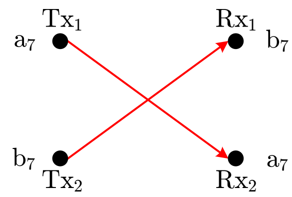


Figure 2.11: *Pairing Type-II: providing $a_6 \oplus a_7$ and $b_6 \oplus b_7$ to both receivers is sufficient to decode the bits. In other words, bit $a_6 \oplus a_7$ available at* Tx_1*, and bit $b_6 \oplus b_7$ available at* Tx_2*, are bits of common interest and can be transmitted to both receivers simultaneously in the efficient two-multicast problem. Note that in* (*b*)*, the cross links would have been irrelevant for future communications, however, using this pairing, we exploit all links.*

Example 5 [Pairing bits of common interest with interference-free bits with swapped receivers to create a two-multicast problem (pairing Type-III)]: Suppose at a time instant, each one of the transmitters sends one data bit, say a_8 and b_8 respectively. Later, using Delayed-CSIT, transmitters figure out that all links were equal to 1 as in Figure 2.12(a). In another time instant, each one of transmitters sends one data bit, say a_9 and b_9 respectively. Later, transmitters figure out that all links were equal to 1 except the link from Tx_2 to Rx_1, see Figure 2.12(b). In a similar case, assume that at another time instant, transmitters one and two send data bits a_{10} and b_{10} at the same time. Through Delayed-CSIT, transmitters realize that all links except the link between Tx_1 and Rx_2 were connected, see Figure 2.12(c). We observe that providing $a_8 \oplus a_9$ and $b_8 \oplus b_{10}$ to both receivers is sufficient to decode the bits. For instance, if Rx_1 is provided with $a_8 \oplus a_9$ and $b_8 \oplus b_{10}$, then it will use a_9 to decode a_8, from which it can obtain b_8, then using b_8 and $b_8 \oplus b_{10}$, it gains access to b_{10}, finally using b_{10}, it can decode a_{10} from $a_{10} \oplus b_{10}$. Thus, bit $a_8 \oplus a_9$ available at Tx_1, and bit $b_8 \oplus b_{10}$ available at Tx_2, are bits of common interest and can be transmitted to both receivers simultaneously in the efficient two-multicast problem. We shall refer to this pairing as pairing **Type-III** throughout the paper.

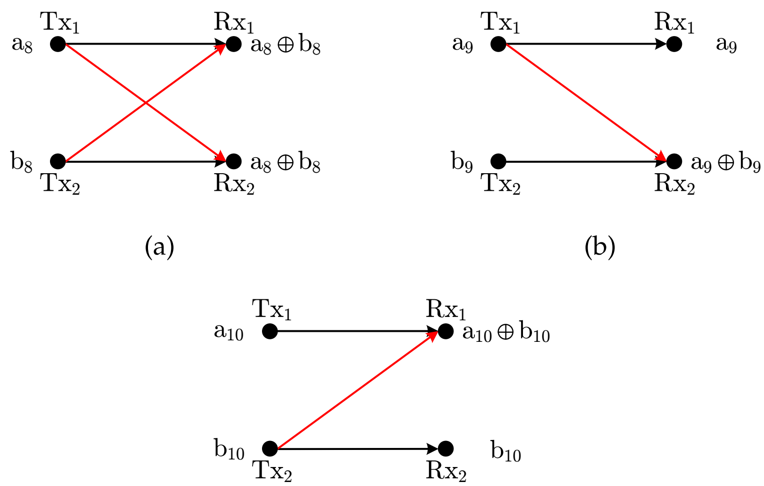


Figure 2.12: *Pairing Type-III: providing $a_8 \oplus a_9$ and $b_8 \oplus b_{10}$ to both receivers is sufficient to decode the bits. In other words, bit $a_8 \oplus a_9$ available at* Tx_1*, and bit $b_8 \oplus b_{10}$ available at* Tx_2*, are bits of common interest and can be transmitted to both receivers simultaneously in the efficient two-multicast problem.*

As explained in the above examples, there are several ways to exploit the available side information at each transmitter. To achieve the capacity region, the first challenge is to evaluate various options and choose the most efficient one. The second challenge is that different opportunities may occur with different probabilities. This makes the process of matching, combining, and upgrading the status of the bits very difficult. Unfortunately, there is no simple guideline to decide when to search for the most efficient combination of the opportunities and when to hold on to other schemes. It is also important to note that most of the opportunities we observe here, do not appear in achieving the DoF of the Gaussian multi-antenna interference channels (see *e.g.*, [25, 68]).

2.4.2 Achievability Ideas with Output Feedback

In this subsection, we focus on the impact of the output feedback in the presence of Delayed-CSIT. The first observation is that through output feedback, each transmitter can evaluate the interference of the other transmitter, and therefore has access to the previously transmitted signal of the other user. Thus, output feedback can create new path of communication between each transmitter and the corresponding receiver, *e.g.*,

$$\mathsf{Tx}_1 \rightarrow \mathsf{Rx}_2 \rightarrow \mathsf{Tx}_2 \rightarrow \mathsf{Rx}_1.$$

Although this additional path can improve the rate region, the advantage of output feedback is not limited to that. We explain the new opportunities through two examples

Example 6 [Creating two-multicast problem from ICs with side information]: In the previous subsection, we showed that interference-free transmissions can be upgraded to two-multicast problems through pairing. However, it is important to note that the different channel realizations used for pairing do not occur at the same probability. Therefore, it is not always possible to fully implement pairing in all cases. In particular, in some cases, some interference-free transmissions are left alone without possibility of pairing. In this example, we show that output feedback allows us to create bits of common interest out of these cases, which in turn allows us to create two-multicast problems. Referring to Figure 2.13, one can see that through the output feedback links, transmitters one and two can learn b_{11} and a_{11} respectively. Therefore, either of the transmitters is able to create $a_{11} \oplus b_{11}$. It is easy to see that $a_{11} \oplus b_{11}$ is of interest of both receivers. Indeed, feedback allows us to form a bit of common interest which can be delivered through the efficient two-multicast problem.

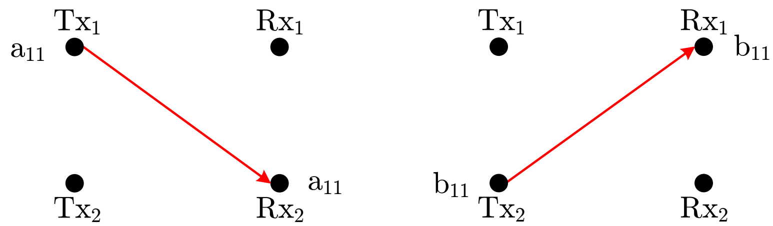


Figure 2.13: *Opportunistic creation of bits of common interest using output feedback: bit b_{11} is available at Tx_1 via the feedback link from Rx_1; it is sufficient that Tx_1 provides $a_{11} \oplus b_{11}$ to both receivers.*

Example 7 [Creating two-multicast problem from ICs with swapped receivers and side information]: As another example, consider the two channel gain realizations depicted in Figure 2.14. In these cases, using output feedback Tx_1 can learn the transmitted bit of Tx_2 (*i.e.* b_{12}), and then form $a_{13} \oplus b_{12}$. It is easy to see that $a_{13} \oplus b_{12}$ is useful for both receivers and thus is a bit of common interest. Similar argument is valid for the second receiver. This means that output feedback allows us to upgrade interference-free transmissions with swapped receivers to bits of common interest that can be used to form efficient two-multicast problems.

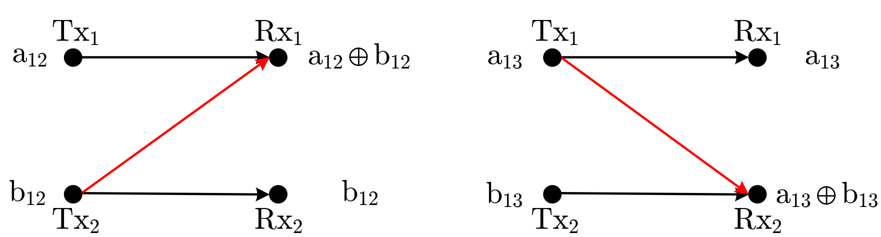


Figure 2.14: *Opportunistic creation of bits of common interest using output feedback: using output feedback Tx_1 can learn the transmitted bit of Tx_2* (i.e. *b_{12}); now, we observe that providing $a_{13} \oplus b_{12}$ to both receivers is sufficient to decode the intended bits.*

2.4.3 Key Idea for Converse Proofs with Delayed-CSIT

While we provide detailed proofs in Sections 2.6 and 2.8, we try to describe the main challenge in deriving the outer-bounds in this subsection. Consider the Delayed-CSIT scenario and suppose rate tuple (R_1, R_2) is achievable. Then for $\beta > 0$, we have

$$
\begin{aligned}
n\left(R_1 + \beta R_2\right) &= H(W_1|W_2, G^n) + \beta H(W_2|G^n) \\
&\overset{\text{(Fano)}}{\leq} I(W_1; Y_1^n|W_2, G^n) + \beta I(W_2; Y_2^n|G^n) + n\epsilon_n \\
&= \beta H(Y_2^n|G^n) + \underbrace{H(G_{11}^n X_1^n|G^n) - \beta H(G_{12}^n X_1^n|G^n)} + n\epsilon_n.
\end{aligned} \tag{2.14}
$$

We refer the reader to Section 2.6 for the detailed derivation of each step. Here, we would like to find a value of β such that

$$
H(G_{11}^n X_1^n|G^n) - \beta H(G_{12}^n X_1^n|G^n) \leq 0, \tag{2.15}
$$

for *any* input distribution. Note that since the terms involved are only a function of X_1^n and the channel gains, this term resembles a broadcast channel formed by Tx_1 and the two receivers. Therefore, the main challenge boils down to understanding the ratio of the entropies of the received signals in a broadcast channel, and this would be the main focus of this subsection.

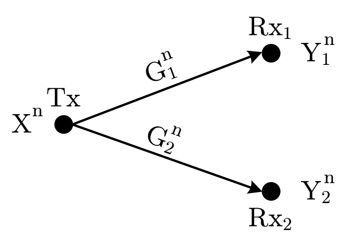


Figure 2.15: A transmitter connected to two receivers through binary fading channels.

Consider a transmitter that is connected to two receivers through binary fading channels as depicted in Figure 2.15. We would like to understand how much this transmitter can privilege receiver one to receiver two, given outdated knowledge of the channel state information. Our metric would be the ratio of the entropies of the received signals[2]. In other words, we would like to understand what is the lower-bound on the ratio of the entropy of the received signal at Rx_2 to that of Rx_1. We first point out the result for the No-CSIT and Instantaneous-CSIT cases. With No-CSIT, from transmitter's point of view the two receivers are identical and it cannot favor one over the other and as a result, the two entropies would be equal. However with Instantaneous-CSIT, transmitter can choose to transmit at time t only if $G_1[t] = 1$ and $G_2[t] = 0$. Thus, with Instantaneous-CSIT the ratio of interest could be as low as 0. For the Delayed-CSIT case, we have the following lemma which we will formally prove in Section 2.6. Here, we try to provide some intuition about the problem by describing an input distribution that utilizes delayed knowledge of the channel state information in order to favor receiver one. It is important to keep in mind that this should not be considered as a proof but rather just a helpful intuition. Also, we point out that for the two-user BFIC with Delayed-CSIT and OFB, we will derive a variation of this lemma in Section 2.8.

Lemma 2.1 [Entropy Leakage] *For the channel described above with Delayed-CSIT, and for* any *input distribution, we have*

$$H\left(Y_2^n|G^n\right) \geq \frac{p}{1-q^2} H\left(Y_1^n|G^n\right). \tag{2.16}$$

As mentioned before, we do not intend to prove this lemma here. We only provide an input distribution for which this lower-bound is tight. Consider m

[2]We point out that if $H(G_{11}^n X_1^n|G^n) = 0$, then ratio is not defined. But we keep in mind that what we really care about is (2.15).

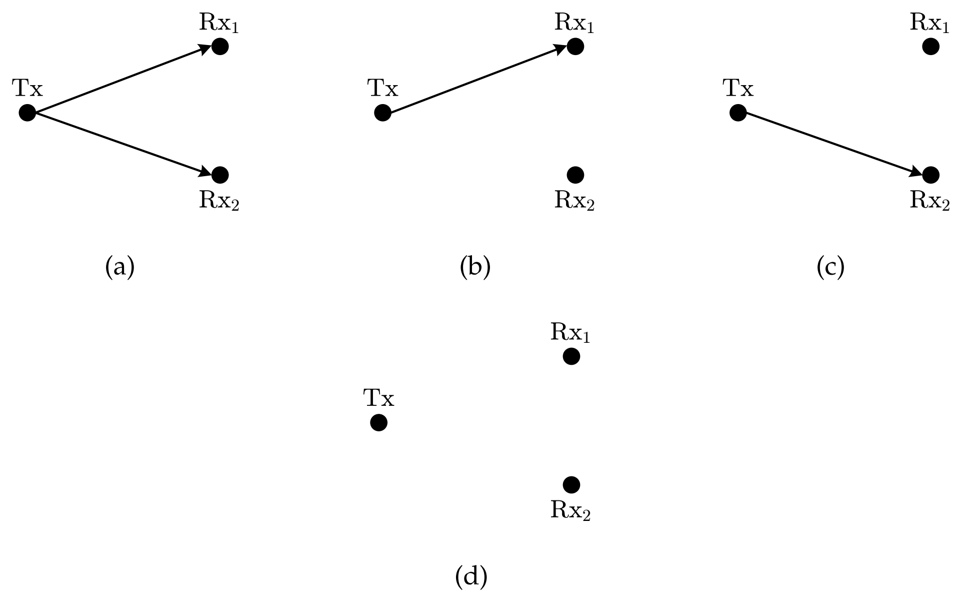


Figure 2.16: *Four possible channel realizations for the network in Figure 2.15. The transmitter sends out a data bit at time instant t, and at the next time instant, using Delayed-CSIT, he knows which channel realization has occurred. If either of the realizations (a) or (b) occurred at time t, then we remove the transmitted bit from the initial queue. However, if either of the realizations (c) or (d) occurred at time t, we leave this bit in the initial queue. This way the transmitter favors receiver one over receiver two.*

bits drawn from i.i.d. Bernoulli 0.5 random variables and assume these bits are in some initial queue. At any time instant t, the transmitter sends one of the bits in this initial queue (if the queue is empty, then the scheme is terminated). At time instant $t + 1$, using Delayed-CSIT, the transmitter knows which one of the four possible channel realizations depicted in Figure 2.16 has occurred at time t. If either of the realizations (a) or (b) occurred at time t, then we remove the transmitted bit from the initial queue. However, if either of the realizations (c) or (d) occurred at time t, we leave this bit in the initial queue (*i.e.* among the bits that can be transmitted at any future time instant). Note that this way, any bit that is available at Rx_2 would be available at Rx_1, however, there will be bits that are only available at Rx_1. Hence, the transmitter has favored receiver one over receiver two. If we analyze this scheme, we get

$$H\left(Y_2^n|G^n\right) = \frac{p}{1-q^2} H\left(Y_1^n|G^n\right), \tag{2.17}$$

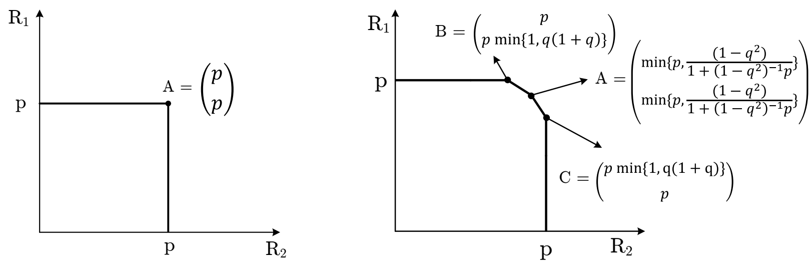
meaning that the bound given in (2.16) is achievable and thus, it is tight.

Now that we have described the key ideas we incorporate in this paper, starting next section, we provide the proof of our main results.

2.5 Achievability Proof of Theorem 2.2 [Delayed-CSIT]

For $0 \leq p \leq \left(3 - \sqrt{5}\right)/2$, the capacity of the two-user BFIC with Delayed-CSIT is depicted in Figure 2.17(a) and as a result, it is sufficient to describe the achievability for point $A = (p, p)$. However, for $\left(3 - \sqrt{5}\right)/2 < p \leq 1$, all bounds are active and the region, as depicted in Figure 2.17(b), is the convex hull of points A, B, and C. By symmetry, it is sufficient to describe the achievability for points A and C in this regime.

We first provide the achievability proof of point A for $0.5 \leq p \leq 1$ in this section. Then, we provide an overview of the achievability proof of corner point C and we postpone the detailed proof to Appendix A.4. Finally in Appendix A.3, we present the achievability proof of point A for $0 \leq p < 0.5$.



(a) (b)

Figure 2.17: *Capacity Region of the two-user Binary Fading IC with Delayed-CSIT for:* (*a*) $0 \leq p \leq \left(3 - \sqrt{5}\right)/2$; *and* (*b*) $\left(3 - \sqrt{5}\right)/2 < p \leq 1$.

2.5.1 Achievability Strategy for Corner Point A

In this subsection, we describe a transmission strategy that achieves a rate tuple arbitrary close to corner point A for $0.5 \leq p \leq 1$ as depicted in Figure 2.17(b), *i.e.*

$$R_1 = R_2 = \frac{(1-q^2)}{1+(1-q^2)^{-1}p}. \tag{2.18}$$

Let the messages of transmitters one and two be denoted by $\mathrm{W}_1 = a_1, a_2, \ldots, a_m$, and $\mathrm{W}_2 = b_1, b_2, \ldots, b_m$, respectively, where data bits a_i's and b_i's are picked uniformly and independently from $\{0,1\}$, $i = 1, \ldots, m$. We show that it is possible to communicate these bits in

$$n = \left(1-q^2\right)^{-1} m + \left(1-q^2\right)^{-2} pm + O\left(m^{2/3}\right) \tag{2.19}$$

time instants[3] with vanishing error probability (as $m \to \infty$). Therefore achieving the rates given in (2.18) as $m \to \infty$. Our transmission strategy consists of two phases as described below.

Phase 1 [uncategorized transmission]: At the beginning of the communication

[3]Throughout the paper whenever we state the number of bits or time instants, say n, if the expression for a given value of p is not an integer, then we use the ceiling of that number $\lceil n \rceil$, where $\lceil . \rceil$ is the smallest integer greater than or equal to n. Note that since we will take the limit as $m \to \infty$, this does not change the end results.

block, we assume that the bits at Tx_i are in queue $Q_{i\rightarrow i}$ (the initial state of the bits), $i = 1, 2$. At each time instant t, Tx_i sends out a bit from $Q_{i\rightarrow i}$, and this bit will either stay in the initial queue or transition to one of the following possible queues will take place according to the description in Table 2.2. If at time instant t, $Q_{i\rightarrow i}$ is empty, then Tx_i, $i = 1, 2$, remains silent until the end of Phase 1.

(A) Q_{i,C_1}: The bits that at the time of communication, all channel gains were equal to 1.

(B) $Q_{i\rightarrow\{1,2\}}$: The bits that are of common interest of both receivers and do not fall in category (A).

(C) $Q_{i\rightarrow i|\bar{i}}$: The bits that are required by Rx_i but are available at the unintended receiver $\mathsf{Rx}_{\bar{i}}$. A bit is in $Q_{i\rightarrow i|\bar{i}}$ if $\mathsf{Rx}_{\bar{i}}$ gets it without interference and Rx_i does not get it with or without interference.

(D) $Q_{i\rightarrow\bar{i}|i}$: The bits that are required by $\mathsf{Rx}_{\bar{i}}$ but are available at the intended receiver Rx_i. More precisely, a bit is in $Q_{i\rightarrow\bar{i}|i}$ if Rx_i gets the bit without interference and $\mathsf{Rx}_{\bar{i}}$ gets it with interference.

(E) $Q_{i\rightarrow F}$: The bits that we consider delivered and no retransmission is required.

More precisely, based on the channel realizations, a total of 16 possible configurations may occur at any time instant as summarized in Table 2.2. The transition for each one of the channel realizations is as follows.

- Case 1 $\left(\boxtimes\right)$: If at time instant t, Case 1 occurs, then each receiver gets a linear combination of the bits that were transmitted. Then as illustrated in Figure 2.18, if either of such bits is provided to both receivers then the receivers can decode both bits. The transmitted bit of Tx_i leaves $Q_{i\rightarrow i}$ and joins Q_{i,C_1}[4], $i = 1, 2$. Although we can consider such bits as bits of common interest, we keep them in an intermediate queue for now and as we describe later, we combine them with other bits to create bits of common interest.

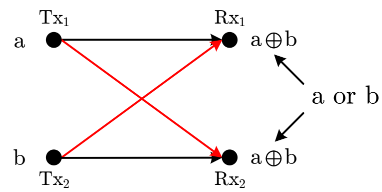


Figure 2.18: Suppose transmitters one and two send out data bits a and b respectively and Case 1 occurs. Now, if either of the transmitted bits is provided to both receivers, then each receiver can decode its corresponding bit.

- Case 2 $\left(\substack{\searrow\\\rightarrow}\right)$: In this case, Rx_1 has already received its corresponding bit while Rx_2 has a linear combination of the transmitted bits, see Table 2.2. As a result, if the transmitted bit of Tx_1 is provided to Rx_2, it will be able to decode both bits. In other words, the transmitted bit from Tx_1 is available at Rx_1 and is required by Rx_2. Therefore, transmitted bit of Tx_1 leaves $Q_{1\rightarrow 1}$ and joins $Q_{1\rightarrow 2|1}$. Note that the bit of Tx_2 will not be retransmitted since upon delivery of the bit of Tx_1, Rx_2 can decode its corresponding bit. Since no retransmission is required, the bit of Tx_2 leaves $Q_{2\rightarrow 2}$ and joins $Q_{2,F}$ (the final state of the bits).

- Case 3 $\left(\substack{\nearrow\\\rightarrow}\right)$: This is similar to Case 2 with swapping user IDs.

[4]In this paper, we assume that the queues are ordered. Meaning that the first bit that joins the queue is placed at the head of the queue and any new bit occupies the next empty position. For instance, suppose there are ℓ bits in Q_{1,C_1} and ℓ bits in Q_{2,C_1}, then the next time Case 1 occurs, the transmitted bit of Tx_i is placed at position $\ell + 1$ in Q_{i,C_1}, $i = 1, 2$.

- Case 4 $\left(\substack{\rightarrow \\ \rightarrow}\right)$: In this case, each receiver gets its corresponding bit without any interference. We consider such bits to be delivered and no retransmission is required. Therefore, the transmitted bit of Tx_i leaves $Q_{i\rightarrow i}$ and joins $Q_{i,F}$, $i = 1, 2$.

- Case 5 $\left(\substack{\rightarrow \\ }\right)$ and Case 6 $\left(\substack{\rightarrow \\ \searrow}\right)$: In these cases, Rx_1 gets its corresponding bit interference free. We consider this bit to be delivered and no retransmission is required. Therefore, the transmitted bit of Tx_1 leaves $Q_{1\rightarrow 1}$ and joins $Q_{1,F}$, while the transmitted bit of Tx_2 remains in $Q_{2\rightarrow 2}$.

- Case 7 $\left(\substack{\rightarrow \\ \nearrow}\right)$: In this case, Rx_1 has a linear combination of the transmitted bits, while Rx_2 has not received anything, see Table 2.2. It is sufficient to provide the transmitted bit of Tx_2 to both receivers. Therefore, the transmitted bit of Tx_2 leaves $Q_{2\rightarrow 2}$ and joins $Q_{2\rightarrow\{1,2\}}$. Note that the bit of Tx_1 will not be retransmitted since upon delivery of the bit of Tx_2, Rx_1 can decode its corresponding bit. This bit leaves $Q_{1\rightarrow 1}$ and joins $Q_{1,F}$. Similar argument holds for Case 8 $\left(\substack{\rightarrow \\ \nearrow\!\!\!\searrow}\right)$.

- Cases 9,10,11, and 12: Similar to Cases 5,6,7, and 8 with swapping user IDs respectively.

- Case 13 $(\nearrow)$: In this case, Rx_1 has received the transmitted bit of Tx_2 while Rx_2 has not received anything, see Table 2.2. Therefore, the transmitted bit of Tx_1 remains in $Q_{1\rightarrow 1}$, while the transmitted bit of Tx_2 is required by Rx_2 and it is available at Rx_1. Hence, the transmitted bit of Tx_2 leaves $Q_{2\rightarrow 2}$ and joins $Q_{2\rightarrow 2|1}$. Queue $Q_{2\rightarrow 2|1}$ represents the bits at Tx_2 that are available at Rx_1, but Rx_2 needs them.

- Case 14 $(\searrow)$: This is similar to Case 13 with swapping user IDs.

- Case 15 ($\times$): In this case, Rx_1 has received the transmitted bit of Tx_2 while Rx_2 has received the transmitted bit of Tx_1, see Table 2.2. In other words, the transmitted bit of Tx_2 is available at Rx_1 and is required by Rx_2; while the transmitted bit of Tx_1 is available at Rx_2 and is required by Rx_1. Therefore, we have transition from $Q_{i\rightarrow i}$ to $Q_{i\rightarrow i|\bar{i}}$, $i = 1, 2$.
- Case 16: The transmitted bit of Tx_i remains in $Q_{i\rightarrow i}$, $i = 1, 2$.

Phase 1 goes on for

$$\left(1 - q^2\right)^{-1} m + m^{\frac{2}{3}} \tag{2.20}$$

time instants, and if at the end of this phase, either of the queues $Q_{i\rightarrow i}$ is not empty, we declare error type-(i) and halt the transmission (we assume m is chosen such that $m^{\frac{2}{3}} \in \mathbb{Z}$).

Assuming that the transmission is not halted, let N_{i,C_1}, $N_{i\rightarrow j|\bar{j}}$, and $N_{i\rightarrow\{1,2\}}$ denote the number of bits in queues Q_{i,C_1}, $Q_{i\rightarrow j|\bar{j}}$, and $Q_{i\rightarrow\{1,2\}}$ respectively at the end of the transitions, $i = 1, 2$, and $j = i, \bar{i}$. The transmission strategy will be halted and an error type-(ii) will occur, if any of the following events happens.

$$\begin{aligned}
&N_{i,\mathsf{C}_1} > \mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}} \triangleq n_{i,\mathsf{C}_1}, \quad i = 1, 2; \\
&N_{i\rightarrow j|\bar{j}} > \mathbb{E}[N_{i\rightarrow j|\bar{j}}] + m^{\frac{2}{3}} \triangleq n_{i\rightarrow j|\bar{j}}, \quad i = 1, 2, \text{ and } j = i, \bar{i}; \\
&N_{i\rightarrow\{1,2\}} > \mathbb{E}[N_{i\rightarrow\{1,2\}}] + m^{\frac{2}{3}} \triangleq n_{i\rightarrow\{1,2\}}, \quad i = 1, 2.
\end{aligned} \tag{2.21}$$

From basic probability, we know that

$$\begin{aligned}
\mathbb{E}[N_{i,\mathsf{C}_1}] &= \frac{\Pr\left(\text{Case 1}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p^4 m, \\
\mathbb{E}[N_{i\rightarrow i|\bar{i}}] &= \frac{\sum_{j=14,15} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p q^2 m,
\end{aligned}$$

$$\mathbb{E}[N_{i\to\bar{i}|i}] = \frac{\Pr(\text{Case 2})}{1-\sum_{i=9,10,13,16}\Pr(\text{Case i})}m = (1-q^2)^{-1}p^3qm,$$
$$\mathbb{E}[N_{i\to\{1,2\}}] = \frac{\sum_{j=11,12}\Pr(\text{Case j})}{1-\sum_{i=9,10,13,16}\Pr(\text{Case i})}m = (1-q^2)^{-1}p^2qm. \tag{2.22}$$

Furthermore, we can show that the probability of errors of types I and II decreases exponentially with m. More precisely, we use Chernoff-Hoeffding bound[5], to bound the error probabilities of types I and II. For instance, to bound the probability of error type-I, we have

$$\begin{aligned}\Pr[\text{error type}-\text{I}] &\leq \sum_{i=1}^{2}\Pr[Q_{i\to i}\text{ is not empty}]\\ &\leq 4\exp\left(\frac{-m^{4/3}}{4n(1-q^2)q^2}\right)\\ &= 4\exp\left(\frac{-m^{4/3}}{4(1-q^2)q^2\left[(1-q^2)^{-1}m+m^{\frac{2}{3}}\right]}\right),\end{aligned} \tag{2.23}$$

which decreases exponentially to zero as $m\to\infty$.

At the end of Phase 1, we add 0's (if necessary) in order to make queues Q_{i,C_1}, $Q_{i\to j|\bar{j}}$, and $Q_{i\to\{1,2\}}$ of size equal to n_{i,C_1}, $n_{i\to j|\bar{j}}$, and $n_{i\to\{1,2\}}$ respectively as defined in (2.21), $i=1,2$, and $j=i,\bar{i}$. For the rest of this subsection, we assume that Phase 1 is completed and no error has occurred.

We now use the ideas described in Section 2.4.1, to further create bits of common interest. Depending on the value of p, we use different ideas. We break the rest of this subsection into two parts: (1) $0.5\leq p\leq\left(\sqrt{5}-1\right)/2$; and (2) $\left(\sqrt{5}-1\right)/2<p\leq 1$. In what follows, we first describe the rest of the achievability strategy for $0.5\leq p\leq\left(\sqrt{5}-1\right)/2$. In particular, we demonstrate how to incorporate the ideas of Section 2.4.1 to create bits of common interest in an

[5]We consider a specific form of the Chernoff-Hoeffding bound [29] described in [41], which is simpler to use and is as follows. If $X_1,\ldots,X_r$ are r independent random variables, and $M=\sum_{i=1}^{r}X_i$, then $Pr[|M-\mathbb{E}[M]|>\alpha]\leq 2\exp\left(\frac{-\alpha^2}{4\sum_{i=1}^{r}\mathrm{Var}(X_i)}\right)$.

optimal way.

- **Type I** Combining bits in $Q_{i \to \bar{i}|i}$ and $Q_{i \to i|\bar{i}}$: Consider the bits that were transmitted in Cases 2 and 14, see Figure 2.19. Observe that if we provide $a_1 \oplus a_2$ to *both* receivers then Rx_1 can decode bits a_1 and a_2, whereas Rx_2 can decode bit b_1. Therefore, $a_1 \oplus a_2$ is a bit of common interest and can join $Q_{1 \to \{1,2\}}$. Hence, as illustrated in Figure 2.20, we can remove two bits in $Q_{1 \to 2|1}$ and $Q_{1 \to 1|2}$, by inserting their XOR in $Q_{1 \to \{1,2\}}$, and we deliver this bit of common interest to both receivers during the second phase. Note that due to the symmetry of the channel, similar argument holds for $Q_{2 \to 1|2}$ and $Q_{2 \to 2|1}$.

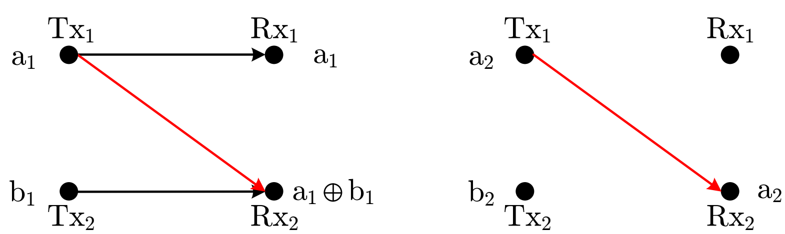


Figure 2.19: Suppose at a time instant, transmitters one and two send out data bits a_1 and b_1 respectively, and later using Delayed-CSIT, transmitters figure out Case 2 occurred at that time. At another time instant, suppose transmitters one and two send out data bits a_2 and b_2 respectively, and later using Delayed-CSIT, transmitters figure out Case 14 occurred at that time. Now, bit $a_1 \oplus a_2$ available at Tx_1 is useful for both receivers and it is a bit of common interest. Hence, $a_1 \oplus a_2$ can join $Q_{1 \to \{1,2\}}$.

For $0.5 \leq p \leq \left(\sqrt{5}-1\right)/2$, we have $\mathbb{E}[N_{i \to \bar{i}|i}] \leq \mathbb{E}[N_{i \to i|\bar{i}}]$. Therefore, after this combination, queue $Q_{i \to \bar{i}|i}$ becomes empty and we have

$$\mathbb{E}[N_{i \to i|\bar{i}}] - \mathbb{E}[N_{i \to \bar{i}|i}] = \left(1-q^2\right)^{-1} pq\left(q-p^2\right)m \tag{2.24}$$

bits left in $Q_{i \to i|\bar{i}}$.

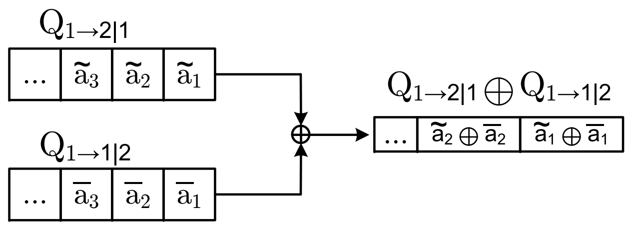


Figure 2.20: Creating XOR of the bits in two different queues. We pick one bit from each queue and create the XOR of the two bits.

- **Type II** Combining the bits in Q_{i,C_1} and $Q_{i\rightarrow i|\bar{i}}$: Consider the bits that were transmitted in Cases 1 and 15, see Figure 2.21. It is easy to see that providing $a_1 \oplus a_2$ and $b_1 \oplus b_2$ to both receivers is sufficient to decode their corresponding bits. For instance, Rx_1 removes b_2 from $b_1 \oplus b_2$ to decode b_1, then uses b_1 to decode a_1 from $a_1 \oplus b_1$. Therefore, $a_1 \oplus a_2$ and $b_1 \oplus b_2$ are bits of common interest and can join $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ respectively. Hence, we can remove two bits in Q_{i,C_1} and $Q_{i\rightarrow i|\bar{i}}$, by inserting their XOR in $Q_{i\rightarrow\{1,2\}}$, $i = 1, 2$, and then deliver this bit of common interest to both receivers during the second phase.

  For $0.5 \leq p \leq \left(\sqrt{5}-1\right)/2$, we have $\left(1-q^2\right)^{-1} pq\left(q-p^2\right)m \leq \mathbb{E}[N_{i,\mathsf{C}_1}]$. Therefore after combining the bits, queue $Q_{i\rightarrow i|\bar{i}}$ becomes empty and we have

$$\mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}} - \left(\mathbb{E}[N_{i\rightarrow i|\bar{i}}] - \mathbb{E}[N_{i\rightarrow \bar{i}|i}]\right) = \left(1-q^2\right)^{-1} p\left(p-q\right)m + m^{\frac{2}{3}} \tag{2.25}$$

  bits left in Q_{i,C_1}, $i = 1, 2$.

Finally, we need to describe what happens to the remaining

$$\left(1-q^2\right)^{-1} p\left(p-q\right)m + m^{\frac{2}{3}}$$

bits in Q_{i,C_1}. As mentioned before, a bit in Q_{i,C_1} can be viewed as a bit of common interest by itself. For the remaining bits in Q_{1,C_1}, we put the first half in $Q_{1\rightarrow\{1,2\}}$

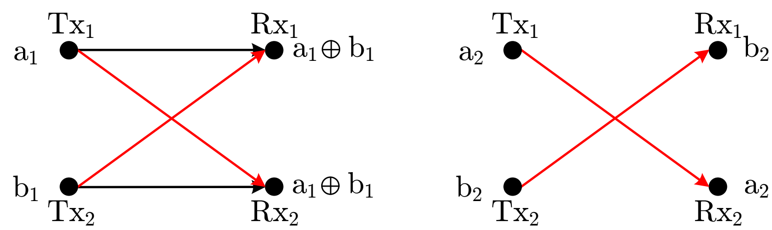


Figure 2.21: Suppose at a time instant, transmitters one and two send out data bits a_1 and b_1 respectively, and later using Delayed-CSIT, transmitters figure out Case 1 occurred at that time. At another time instant, suppose transmitters one and two send out data bits a_2 and b_2 respectively, and later using Delayed-CSIT, transmitters figure out Case 15 occurred at that time. Now, bit $a_1 \oplus a_2$ available at Tx_1 and bit $b_1 \oplus b_2$ available at Tx_2 are useful for both receivers and they are bits of common interest. Therefore, bits $a_1 \oplus a_2$ and $b_1 \oplus b_2$ can join $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ respectively.

(suppose m is picked such that the remaining number of bits is even). Note that if these bits are delivered to Rx_2, then Rx_2 can decode the first half of the remaining bits in Q_{2,C_1} as well. Therefore, the first half of the bits in Q_{2,C_1} can join $Q_{2,F}$.

Then, we put the second half of the remaining bits in Q_{2,C_1} in $Q_{2\rightarrow\{1,2\}}$. Similar to the argument presented above, the second half of the bits in Q_{1,C_1} join $Q_{1,F}$.

Hence at the end of Phase 1, if the transmission is not halted, we have a total of

$$
\begin{aligned}
&(1-q^2)^{-1}\left[\underbrace{p^2q}_{\text{Cases 11 and 12}} + \underbrace{pq^2}_{\text{XOR opportunities}} + 0.5\underbrace{(p^4 - pq^2 + p^3q)}_{\text{remaining Case 1}}\right] m + 2.5m^{2/3} \\
&= \left(1-q^2\right)^{-1} 0.5pm + 2.5m^{2/3} \qquad (2.26)
\end{aligned}
$$

number of bits in $Q_{1\rightarrow\{1,2\}}$. Same result holds for $Q_{2\rightarrow\{1,2\}}$.

This completes the description of Phase 1 for $0.5 \leq p \leq \left(\sqrt{5}-1\right)/2$. For

$\left(\sqrt{5}-1\right)/2 < p \leq 1$, we combine the bits as follows. Fot this range of p, after Phase 1, the number of bits in each queue is such that the mergings described above are not optimal and we have to rearrange them as decsribed below.

- **Type I** Combining $Q_{i\rightarrow\bar{i}|i}$ and $Q_{i\rightarrow i|\bar{i}}$: We have already described this opportunity for $0.5 \leq p \leq \left(\sqrt{5}-1\right)/2$. We create the XOR of the bits in $Q_{1\rightarrow 2|1}$ and $Q_{1\rightarrow 1|2}$ and put the XOR of them in $Q_{1\rightarrow\{1,2\}}$. Note that due to the symmetry of the channel, similar argument holds for $Q_{2\rightarrow 1|2}$ and $Q_{2\rightarrow 2|1}$.

  For $\left(\sqrt{5}-1\right)/2 < p \leq 1$, $\mathbb{E}[N_{i\rightarrow i|\bar{i}}] \leq \mathbb{E}[N_{i\rightarrow\bar{i}|i}]$, $i = 1,2$. Therefore after combining the bits, queue $Q_{i\rightarrow i|\bar{i}}$ becomes empty, and we have

  $$\mathbb{E}[N_{i\rightarrow\bar{i}|i}] - \mathbb{E}[N_{i\rightarrow i|\bar{i}}] = \left(1-q^2\right)^{-1} pq\left(p^2-q\right)m \tag{2.27}$$

  bits left in $Q_{i\rightarrow\bar{i}|i}$, $i = 1,2$.

- **Type III** Combining the bits in Q_{i,C_1} and $Q_{i\rightarrow\bar{i}|i}$: Consider the bits that were transmitted in Cases 1, 2, and 3, see Figure 2.22. Now, we observe that providing $a \oplus c$ and $b \oplus f$ to both receivers is sufficient to decode their corresponding bits. For instance, Rx_1 will have $a \oplus b$, c, $e \oplus f$, $a \oplus c$, and $b \oplus f$, from which it can recover a, c, and e. Similar argument holds for Rx_2. Therefore, $a \oplus c$ and $b \oplus f$ are bits of common interest and can join $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ respectively. Hence, we can remove two bits in Q_{i,C_1} and $Q_{i\rightarrow\bar{i}|i}$, by inserting their XORs in $Q_{i\rightarrow\{1,2\}}$, $i = 1,2$, and then deliver this bit of common interest to both receivers during the second phase.

  For $\left(\sqrt{5}-1\right)/2 < p \leq 1$, we have $\left(1-q^2\right)^{-1} pq\left(p^2-q\right)m \leq \mathbb{E}[N_{i,\mathsf{C}_1}]$. Therefore after combining the bits, queue $Q_{i\rightarrow\bar{i}|i}$ becomes empty and we have

  $$\mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}} - \left(\mathbb{E}[N_{i\rightarrow\bar{i}|i}] - \mathbb{E}[N_{i\rightarrow i|\bar{i}}]\right) = \left(1-q^2\right)^{-1}\left(p^4 - p^3q + pq^2\right)m + m^{\frac{2}{3}} \tag{2.28}$$

  bits left in Q_{i,C_1}.

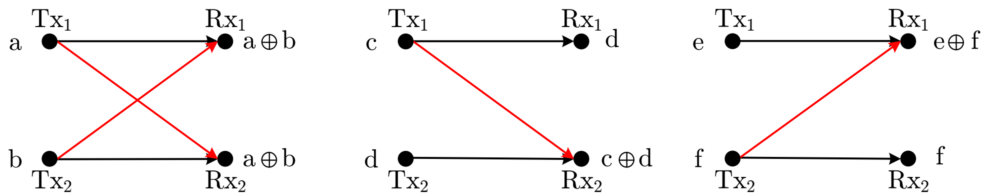


Figure 2.22: Consider the bits transmitted in Cases 1,2, and 3. Now, bit $a \oplus c$ available at Tx_1 and bit $b \oplus f$ available at Tx_2 are useful for both receivers and they are bits of common interest. Therefore, bits $a \oplus c$ and $b \oplus f$ can join $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ respectively.

We treat the remaining bits in Q_{i,C_1} as described before. Hence at the end of Phase 1, if the transmission is not halted, we have a total of

$$\begin{aligned}
&(1-q^2)^{-1}\left[\underbrace{p^2 q}_{\text{Cases 11 and 12}} + \underbrace{p^3 q}_{\text{XOR opportunities}} + 0.5\underbrace{(p^4 - p^3 q + pq^2)}_{\text{remaining Case 1}}\right] m + 2.5 m^{2/3} \\
&= \left(1-q^2\right)^{-1} 0.5pm + 2.5m^{2/3} \qquad (2.29)
\end{aligned}$$

number of bits in $Q_{1\rightarrow\{1,2\}}$. Same result holds for $Q_{2\rightarrow\{1,2\}}$.

To summarize, at the end of Phase 1 assuming that the transmission is not halted, by using coding opportunities of types I, II, and III, we are only left with $\left(1-q^2\right)^{-1} 0.5pm + 2.5m^{2/3}$ bits in queue $Q_{i\rightarrow\{1,2\}}$, $i = 1, 2$.

We now describe how to deliver the bits of common interest in Phase 2 of the transmission strategy. The problem resembles a network with two transmitters and two receivers where each transmitter Tx_i wishes to communicate an independent message W_i to *both* receivers, $i = 1, 2$. The channel gain model is the same as described in Section 2.2. We refer to this network as the two-multicast network as depicted in Figure 2.23. We have the following result for this network.

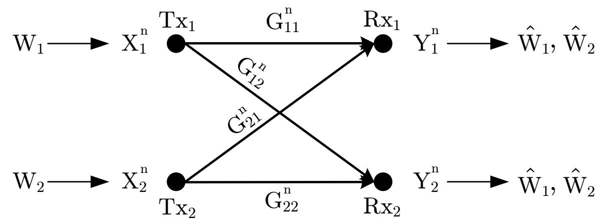


Figure 2.23: Two-multicast network. Transmitter Tx_i wishes to reliably communicate message W_i to both receivers, $i = 1, 2$. The capacity region with no, delayed, or instantaneous CSIT is the same.

Lemma 2.2 *For the two-multicast network as described above, we have*

$$C^{\text{No-CSIT}}_{\text{multicast}} = C^{\text{DCSIT}}_{\text{multicast}} = C^{\text{ICSIT}}_{\text{multicast}}, \tag{2.30}$$

and, we have

$$C^{\text{ICSIT}}_{\text{multicast}} = \begin{cases} R_i \leq p, & i = 1, 2, \\ R_1 + R_2 \leq 1 - q^2. \end{cases} \tag{2.31}$$

This result basically shows that the capacity region of the two-multicast network described above is equal to the capacity region of the multiple-access channel formed at either of the receivers. The proof of Lemma 2.2 is presented in Appendix A.5.

Phase 2 [transmitting bits of common interest]: In this phase, we deliver the bits in $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ using the transmission strategy for the two-multicast problem. More precisely, the bits in $Q_{i\rightarrow\{1,2\}}$ will be considered as the message of Tx_i and they will be encoded as in the achievability scheme of Lemma 2.2, $i = 1, 2$. Fix $\epsilon, \delta > 0$, from Lemma 2.2, we know that rate tuple

$$(R_1, R_2) = \frac{1}{2}\left((1 - q^2) - \delta, (1 - q^2) - \delta\right)$$

is achievable with decoding error probability less than or equal to ϵ. Therefore, transmission of the bits in $Q_{1\to\{1,2\}}$ and $Q_{2\to\{1,2\}}$, will take

$$t_{\text{total}} = \frac{\left(1-q^2\right)^{-1} pm + 5m^{2/3}}{(1-q^2)-\delta}. \tag{2.32}$$

Therefore, the total transmssion time of our two-phase achievability strategy is equal to

$$(1-q^2)^{-1}m + m^{\frac{2}{3}} + \frac{\left(1-q^2\right)^{-1} pm + 5m^{2/3}}{(1-q^2)-\delta}, \tag{2.33}$$

hence, if we let $\epsilon, \delta \to 0$ and $m \to \infty$, the decoding error probability of delivering bits of common interest goes to zero, and we achieve a symmetric sum-rate of

$$R_1 = R_2 = \lim_{\substack{\epsilon,\delta\to 0\\ m\to\infty}} \frac{m}{t_{\text{total}}} = \frac{(1-q^2)}{1+(1-q^2)^{-1}p}. \tag{2.34}$$

This completes the achievability proof of point A for $0.5 \leq p \leq 1$.

2.5.2 Overview of the Achievability Strategy for Corner Point C

We now provide an overview of the achievability strategy for corner point C depicted in Figure 2.17 for $\left(3-\sqrt{5}\right)/2 < p \leq 1$, *i.e.*

$$(R_1, R_2) = (pq(1+q), p), \tag{2.35}$$

and we postpone the detailed proof to Appendix A.4.

Compared to the achievability strategy of the sum-rate point, the challenges in achieving the other corner points arise from the asymmetricity of the rates.

At this corner point, while Tx_2 (the primary user) communicates at full rate of p, Tx_1 (the secondary user) communicates at a lower rate and tries to coexist with the primary user. The achievability strategy is based on the following two principles.

(a) If the secondary user creates interference at the primary receiver, it is the secondary user's responsibility to resolve this interference;

(b) For the achievability of the optimal sum-rate point A (see Figure 2.17(b)), the bits of common interest were transmitted such that both receivers could decode them. However, for corner point C, when the primary receiver obtains a bit of common interest, we revise the coding scheme in a way that favors the primary receiver.

Our transmission strategy consists of five phases as summarized below.

- Phase 1 [uncategorized transmission]: This phase is similar to Phase 1 of the achievability of the optimal sum-rate point A. The main difference is due to the fact that the transmitters have unequal number of bits at the beginning. In Phase 1, Tx_1 (the secondary user) transmits all its initial bits while Tx_2 (the primary user) only transmits part of its initial bits. Transmitter two postpones the transmission of its remaining bits to Phase 3.
- Phase 2 [updating status of the bits transmitted when either of the Cases 7 or 8 (11 or 12) occurred]: For the achievability of optimal sum-rate point A, we transferred the transmitted bits of Tx_2 (Tx_1) to the two-multicast sub-problem by viewing them as bits of common interest. However, this scheme turns out to be suboptimal for corner point C. In this case, we retransmit these bits during Phase 2 and update their status based on the

channel realization at the time of transmission. Phase 2 provides coding opportunities that we exploit in Phases 4 and 5.

- Phase 3 [uncategorized transmission vs interference management]: In this phase, the primary user transmits the remaining initial bits while the secondary user tries to resolve as much interference as it can at the primary receiver. To do so, the secondary user sends the bits that caused interference at the primary receiver during Phase 1, at a rate low enough such that both receivers can decode and remove them regardless of what the primary transmitter does. Note that pq of the time, each receiver gets interference-free signal from the secondary transmitter, hence, the secondary transmitter can take advantage of these time instants to deliver its bits during Phase 3.

- Phases 4 and 5 [delivering interference-free bits and interference management]: In the final phases, each transmitter has two main objectives: (1) communicating the bits required by its own receiver but available at the unintended receiver; and (2) mitigating interference at the unintended receiver. This task can be accomplished by creating the XOR of the bits similar to coding type-I described in Section 2.4 with a modification: we first encode these bits and then create the XOR of the encoded bits. Moreover, the balance of the two objectives is different between the primary user and the secondary user.

As mentioned before, the detailed proof of the achievability for corner point C is provided in Appendix A.4. In the following section, we describe the converse proof for the two-user BFIC with Delayed-CSIT.

2.6 Converse Proof of Theorem 2.2 [Delayed-CSIT]

In this section, we provide the converse proof for Theorem 2.2. As mentioned in Remark 2.3, The outer-bound on the capacity region with only Delayed-CSIT (2.11) is in fact the intersection of the outer-bounds on the individual rates (*i.e.* $R_i \leq p$, $i = 1, 2$) and the capacity region with Delayed-CSIT and OFB (2.12). Therefore, the converse proof that we will later present in Section 2.8 for the case of Delayed-CSIT and OFB suffices. However, specific challenges arise when OFB is present and careful considerations must be taken into account. Here, we independently present the converse proof of Theorem 2.2 to highlight the key techniques without worrying about the details needed for the case of OFB.

We first present the Entropy Leakage Lemma that plays a key role in deriving the converse. Consider the scenario where a transmitter is connected to two receivers through binary fading channels as in Figure 2.24. Suppose $G_1[t]$ and $G_2[t]$ are distributed as i.i.d. Bernoulli RVs (*i.e.* $G_i[t] \overset{d}{\sim} \mathcal{B}(p)$), $i = 1, 2$. In this channel the received signals are given as

$$Y_i[t] = G_i[t]X[t], \qquad i = 1, 2, \tag{2.36}$$

where $X[t]$ is the transmit signal at time t. We have the following lemma.

Lemma 2.3 [Entropy Leakage] *For the channel described above with Delayed-CSIT and for* any *input distribution, we have*

$$H\left(Y_2^n|G^n\right) \geq \frac{1}{2-p} H\left(Y_1^n|G^n\right). \tag{2.37}$$

Remark 2.14 *Note that with No-CSIT, from the transmitter's point of view, the two receivers are identical and it cannot favor one over the other and as a result, we have*

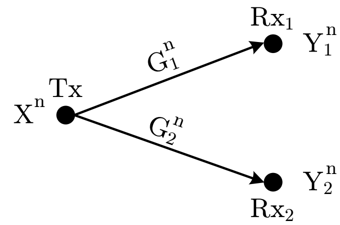


Figure 2.24: A transmitter connected to two receivers through binary fading channels. Using Delayed-CSIT, the transmitter can privilege reveiver one to receiver two. Lemma 2.3 formalizes this privilege.

$H\left(Y_2^n|G^n\right) = H\left(Y_1^n|G^n\right)$. *With Instantaneous-CSIT this ratio can become zero*[6]. *Therefore, this lemma captures the impact of Delayed-CSIT on the entropy of the received signals at the two receivers.*

Proof: For time instant t where $1 \leq t \leq n$, we have

$$\begin{aligned}
&H\left(Y_2[t]|Y_2^{t-1}, G^t\right) \\
&\quad = pH\left(X[t]|Y_2^{t-1}, G_2[t] = 1, G^{t-1}\right) \\
&\quad \overset{(a)}{=} pH\left(X[t]|Y_2^{t-1}, G^t\right) \\
&\quad \overset{(b)}{\geq} pH\left(X[t]|Y_1^{t-1}, Y_2^{t-1}, G^t\right) \\
&\quad \overset{(c)}{=} \frac{p}{1-q^2} H\left(Y_1[t], Y_2[t]|Y_1^{t-1}, Y_2^{t-1}, G^t\right),
\end{aligned} \tag{2.38}$$

where (a) holds since $X[t]$ is independent of the channel realization at time instant t; (b) follows from the fact that conditioning reduces entropy; and (c) follows from the fact that $\Pr\left[G_1[t] = G_2[t] = 0\right] = q^2$. Therefore, we have

$$\sum_{t=1}^{n} H\left(Y_2[t]|Y_2^{t-1}, G^t\right) \geq \frac{1}{2-p} \sum_{t=1}^{n} H\left(Y_1[t], Y_2[t]|Y_1^{t-1}, Y_2^{t-1}, G^t\right), \tag{2.39}$$

[6]This can be done by simply remaining silent whenever $G_2[t] = 1$.

and since the transmit signals at time instant t are independent from the channel realizations in future time instants, we have

$$\sum_{t=1}^{n} H\left(Y_2[t]|Y_2^{t-1}, G^n\right) \geq \frac{1}{2-p} \sum_{t=1}^{n} H\left(Y_1[t], Y_2[t]|Y_1^{t-1}, Y_2^{t-1}, G^n\right), \tag{2.40}$$

hence, we get

$$H\left(Y_2^n|G^n\right) \geq \frac{1}{2-p} H\left(Y_1^n, Y_2^n|G^n\right) \geq \frac{1}{2-p} H\left(Y_1^n|G^n\right). \tag{2.41}$$

This completes the proof of the lemma. ■

We now derive the converse for Theorem 2.2. The outer-bound on R_i is the same under no, delayed, and instantaneous CSIT, and we present it in Appendix A.2. In this section, we provide the proof of

$$R_i + (1+q)R_{\bar{i}} \leq p(1+q)^2, \qquad i = 1, 2. \tag{2.42}$$

By symmetry, it is sufficient to prove it for $i = 1$. Let $\beta = (1+q)$, and suppose rate tuple (R_1, R_2) is achievable. Then we have

$$\begin{aligned}
& n\left(R_1 + \beta R_2\right) = H(W_1) + \beta H(W_2) \\
& \overset{(a)}{=} H(W_1|W_2, G^n) + \beta H(W_2|G^n) \\
& \overset{\text{(Fano)}}{\leq} I(W_1; Y_1^n|W_2, G^n) + \beta I(W_2; Y_2^n|G^n) + n\epsilon_n \\
& = H(Y_1^n|W_2, G^n) - \underbrace{H(Y_1^n|W_1, W_2, G^n)}_{=\,0} + \beta H(Y_2^n|G^n) - \beta H(Y_2^n|W_2, G^n) + n\epsilon_n \\
& \overset{(b)}{=} \beta H(Y_2^n|G^n) + H(Y_1^n|W_2, X_2^n, G^n) - \beta H(Y_2^n|W_2, X_2^n, G^n) + n\epsilon_n \\
& = \beta H(Y_2^n|G^n) + H(G_{11}^n X_1^n|W_2, X_2^n, G^n) - \beta H(G_{12}^n X_1^n|W_2, X_2^n, G^n) + n\epsilon_n \\
& \overset{(c)}{=} \beta H(Y_2^n|G^n) + H(G_{11}^n X_1^n|W_2, G^n) - \beta H(G_{12}^n X_1^n|W_2, G^n) + n\epsilon_n \\
& \overset{(d)}{=} \beta H(Y_2^n|G^n) + H(G_{11}^n X_1^n|G^n) - \beta H(G_{12}^n X_1^n|G^n) + n\epsilon_n \\
& \overset{\text{Lemma 2.3}}{\leq} \beta H(Y_2^n|G^n) + n\epsilon_n
\end{aligned}$$

$$= \beta \sum_{t=1}^{n} H(Y_2[t]|Y_2^{t-1}, G^n) + n\epsilon_n$$

$$\overset{(e)}{\leq} \beta \sum_{t=1}^{n} H(Y_2[t]|G^n) + n\epsilon_n$$

$$\overset{(f)}{\leq} n\beta(1-q^2) + \epsilon_n = np(1+q)^2 + n\epsilon_n. \tag{2.43}$$

where (a) holds since $\mathrm{W}_1, \mathrm{W}_2$ and G^n are mutually independent; (b) and (c) hold since X_2^n is a deterministic function of W_2 and G^n; (d) follows from

$$\begin{aligned} 0 &\leq H(G_{11}^n X_1^n|G^n) - H(G_{11}^n X_1^n|W_2, G^n) \\ &= I\left(G_{11}^n X_1^n; W_2|G^n\right) \leq I\left(W_1, G_{11}^n X_1^n; W_2|G^n\right) \\ &= \underbrace{I\left(W_1; W_2|G^n\right)}_{=\ 0\ \text{since}\ \mathrm{W}_1 \perp \mathrm{W}_2 \perp \mathrm{G}^n} + \underbrace{I\left(G_{11}^n X_1^n; W_2|W_1, G^n\right)}_{=\ 0\ \text{since}\ \mathrm{X}_1^n = \mathrm{f}_1(\mathrm{W}_1,\ \mathrm{G}^n)} = 0, \end{aligned} \tag{2.44}$$

which implies $H(G_{11}^n X_1^n|G^n) = H(G_{11}^n X_1^n|W_2, G^n)$, and similarly $H(G_{12}^n X_1^n|G^n) = H(G_{12}^n X_1^n|W_2, G^n)$; (e) is true since conditioning reduces entropy; and (f) holds since the probability that at least one of the links connected to Rx_2 is equal to 1 at each time instant is $(1-q^2)$. Dividing both sides by n and let $n \to \infty$, we get

$$R_1 + (1+q)R_2 \leq p(1+q)^2. \tag{2.45}$$

This completes the converse proof for Theorem 2.2.

2.7 Achievability Proof of Theorem 2.3 [Delayed-CSIT and OFB]

We now focus on the impact of the output feedback in the presence of Delayed-CSIT. In particular, we demonstrate how output feedback can be utilized to further improve the achievable rates. The capacity region of the two-user BFIC with Delayed-CSIT and OFB is given by

$$\mathcal{C}^{\text{DCSIT,OFB}} = \left\{R_1, R_2 \in \mathbb{R}^{+} \ s.t. \ R_i + (1+q)R_{\bar{i}} \leq p(1+q)^2, \ i = 1,2\right\}, \tag{2.46}$$

and is depicted in Figure 2.25.

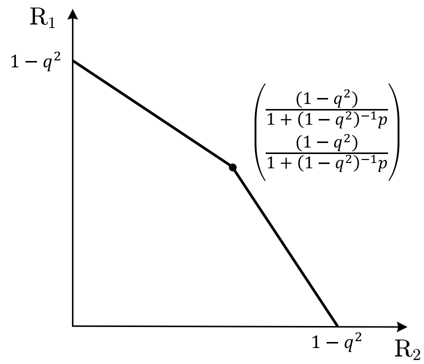


Figure 2.25: Capacity region of the two-user BFIC with Delayed-CSIT and output feedback.

The achievability strategy of the corner points $\left(1-q^2, 0\right)$ and $\left(0, 1-q^2\right)$, is based on utilizing the additional communication paths created by the means of the output feedback links, *e.g.*,

$$\mathsf{Tx}_1 \rightarrow \mathsf{Rx}_2 \rightarrow \mathsf{Tx}_2 \rightarrow \mathsf{Rx}_1,$$

and is presented in Appendix A.6. Here, we only describe the transmission strategy for the sum-rate point, *i.e.*

$$R_1 = R_2 = \frac{(1-q^2)}{1+(1-q^2)^{-1}p}. \tag{2.47}$$

Let the messages of transmitters one and two be denoted by $\mathrm{W}_1 = a_1, a_2, \ldots, a_m$, and $\mathrm{W}_2 = b_1, b_2, \ldots, b_m$, respectively, where data bits a_i's and b_i's are picked uniformly and independently from $\{0,1\}$, $i = 1, \ldots, m$. We show that it is possible to communicate these bits in

$$n = \left(1-q^2\right)^{-1} m + \left(1-q^2\right)^{-2} pm + O\left(m^{2/3}\right) \tag{2.48}$$

time instants with vanishing error probability (as $m \to \infty$). Therefore achieving the rates given in (2.47) as $m \to \infty$. Our transmission strategy consists of two phases as described below.

Phase 1 [uncategorized transmission]: This phase is identical to Phase 1 of Section 2.5. At the beginning of the communication block, we assume that the bits at Tx_i are in queue $Q_{i \to i}$, $i = 1, 2$. At each time instant t, Tx_i sends out a bit from $Q_{i \to i}$, and this bit will either stay in the initial queue or transition to a new queue will take place. The transitions are identical to what we have already described in Table 2.2, therefore, we are not going to repeat them here. Phase 1 goes on for

$$\left(1 - q^2\right)^{-1} m + m^{\frac{2}{3}} \tag{2.49}$$

time instants and if at the end of this phase, either of the queues $Q_{1 \to 1}$ or $Q_{2 \to 2}$ is not empty, we declare error type-I and halt the transmission.

The transmission strategy will be halted and an error type-II will occur, if any of the following events happens.

$$\begin{aligned}
&N_{i,\mathsf{C}_1} > \mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}} \triangleq n_{i,\mathsf{C}_1}, \quad i = 1, 2; \\
&N_{i \to j|\bar{j}} > \mathbb{E}[N_{i \to j|\bar{j}}] + m^{\frac{2}{3}} \triangleq n_{i \to j|\bar{j}}, \quad i = 1, 2, \text{ and } j = i, \bar{i}; \\
&N_{i \to \{1,2\}} > \mathbb{E}[N_{i \to \{1,2\}}] + m^{\frac{2}{3}} \triangleq n_{i \to \{1,2\}}, \quad i = 1, 2.
\end{aligned} \tag{2.50}$$

From basic probability, we know that

$$\begin{aligned}
\mathbb{E}[N_{i,\mathsf{C}_1}] &= \frac{\Pr\left(\text{Case 1}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1 - q^2)^{-1} p^4 m, \\
\mathbb{E}[N_{i \to i|\bar{i}}] &= \frac{\sum_{j=14,15} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1 - q^2)^{-1} p q^2 m, \\
\mathbb{E}[N_{i \to \bar{i}|i}] &= \frac{\Pr\left(\text{Case 2}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1 - q^2)^{-1} p^3 q m, \\
\mathbb{E}[N_{i \to \{1,2\}}] &= \frac{\sum_{j=11,12} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1 - q^2)^{-1} p^2 q m.
\end{aligned} \tag{2.51}$$

Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

At the end of Phase 1, we add 0's (if necessary) in order to make queues Q_{i,C_1}, $Q_{i\to j|\bar{j}}$, and $Q_{i\to\{1,2\}}$ of size equal to n_{i,C_1}, $n_{i\to j|\bar{j}}$, and $n_{i\to\{1,2\}}$ respectively as defined in (2.50), $i = 1, 2$, and $j = i, \bar{i}$. For the rest of this subsection, we assume that Phase 1 is completed and no error has occurred. We now use the ideas described in Section 2.4 for output feedback, to further create bits of common interest.

• Updating the status of bits in Q_{i,C_1} to bits of common interest: A bit in Q_{i,C_1} can be considered as a bit of common interest. Also note that it is sufficient to deliver only one of the two bits transmitted simultaneously during Case 1. Therefore, Tx_1 updates the status of the first half of the bits in Q_{1,C_1} to $Q_{1\to\{1,2\}}$, whereas Tx_2 updates the status of the second half of the bits in Q_{2,C_1} to $Q_{2\to\{1,2\}}$. Hence, after updating the status of bits in Q_{i,C_1}, we have

$$(1-q^2)^{-1}\left[p^2q+\frac{1}{2}p^4\right]m+\frac{3}{2}m^{\frac{2}{3}} \tag{2.52}$$

bits in $Q_{i\to\{1,2\}}$, $i = 1, 2$.

• Upgrading ICs with side information to a two-multicast problem using OFB: Note that through the output feedback links, each transmitter has access to the transmitted bits of the other user during Phase 1. As described above, there are $\mathbb{E}[N_{i\to i|\bar{i}}] + m^{\frac{2}{3}}$ bits in $Q_{i\to i|\bar{i}}$ at the end of Phase 1. Now, Tx_1 creates the XOR of the first half of the bits in $Q_{1\to 1|2}$ and $Q_{2\to 2|1}$ and updates the status of the resulting bits to $Q_{1\to\{1,2\}}$. Note that as described in Example 6 of Section 2.4, the XOR of these bits is a bit of common interest. On the other hand, Tx_2 creates the XOR of the second half of the bits in $Q_{1\to 1|2}$ and $Q_{2\to 2|1}$ and updates the status of the resulting bits to $Q_{2\to\{1,2\}}$. Thus, we have

$$(1-q^2)^{-1}\left[p^2q+\frac{1}{2}p^4+\frac{1}{2}pq^2\right]m+2m^{\frac{2}{3}} \tag{2.53}$$

bits in $Q_{i\to\{1,2\}}$, $i=1,2$.

• Upgrading ICs with side information and swapped receivers to a two-multicast problem using OFB: As described above, there are $\mathbb{E}[N_{i\to\bar{i}|i}]+m^{\frac{2}{3}}$ bits in $Q_{i\to\bar{i}|i}$, Tx_1 creates the XOR of the first half of the bits in $Q_{1\to2|1}$ and $Q_{2\to1|2}$ and updates the status of the resulting bits to $Q_{1\to\{1,2\}}$. Note that as described in Example 7 of Section 2.4, the XOR of these bits is a bit of common interest. On the other hand, Tx_2 creates the XOR of the second half of the bits in $Q_{1\to2|1}$ and $Q_{2\to1|2}$ and updates the status of the resulting bits to $Q_{2\to\{1,2\}}$. Hence, we have

$$(1-q^2)^{-1}\left[p^2q+\frac{1}{2}p^4+\frac{1}{2}pq^2+\frac{1}{2}p^3q\right]m+\frac{5}{2}m^{\frac{2}{3}}=(1-q^2)^{-1}\frac{p}{2}m+\frac{5}{2}m^{\frac{2}{3}} \tag{2.54}$$

bits in $Q_{i\to\{1,2\}}$, $i=1,2$. This completes the description of Phase 1.

Phase 2 [transmitting bits of common interest]: In this phase, we deliver the bits in $Q_{1\to\{1,2\}}$ and $Q_{2\to\{1,2\}}$ using the transmission strategy for the two-multicast problem. More precisely, the bits in $Q_{i\to\{1,2\}}$ will be considered as the message of Tx_i and they will be encoded as in the achievability scheme of Lemma 2.2, $i=1,2$. Fix $\epsilon,\delta>0$, from Lemma 2.2 we know that the rate tuple

$$(R_1,R_2)=\frac{1}{2}\Big((1-q^2)-\delta/2,(1-q^2)-\delta/2\Big)$$

is achievable with decoding error probability less than or equal to ϵ. Therefore, transmission of the bits in $Q_{1\to\{1,2\}}$ and $Q_{2\to\{1,2\}}$ requires

$$t_{\text{total}}=\frac{\left(1-q^2\right)^{-1}pm+5m^{2/3}}{(1-q^2)-\delta} \tag{2.55}$$

time instants. Therefore, the total transmission time of our two-phase achievability strategy is equal to

$$(1-q^2)^{-1}m + m^{\frac{2}{3}} + \frac{\left(1-q^2\right)^{-1} pm + 5m^{2/3}}{(1-q^2)-\delta}. \tag{2.56}$$

The probability that the transmission strategy halts at any point can be bounded by the summation of error probabilities of types I and II, and the probability that an error occurs in decoding the encoded bits. This probability approaches zero for $\epsilon, \delta \to 0$ and $m \to \infty$.

Hence, if we let $\epsilon, \delta \to 0$ and $m \to \infty$, the decoding error probability goes to zero, and we achieve a symmetric sum-rate of

$$R_1 = R_2 = \lim_{\substack{\epsilon,\delta\to 0\\ m\to\infty}} \frac{m}{t_{\text{total}}} = \frac{(1-q^2)}{1+(1-q^2)^{-1}p}. \tag{2.57}$$

2.8 Converse Proof of Theorem 2.3 [Delayed-CSIT and OFB]

In this section, we prove the converse for Theorem 2.3. Suppose rate tuple (R_1, R_2) is achievable, then by letting $\beta = 1+q$, we have

$$\begin{aligned}
n\left(R_1+\beta R_2\right) &= H(W_1)+\beta H(W_2)\\
&\overset{(a)}{=} H(W_1|W_2,G^n)+\beta H(W_2|G^n)\\
&\overset{\text{Fano}}{\leq} I(W_1;Y_1^n|W_2,G^n)+\beta I(W_2;Y_2^n|G^n)+n\epsilon_n\\
&\leq I(W_1;Y_1^n,Y_2^n|W_2,G^n)+\beta I(W_2;Y_2^n|G^n)+n\epsilon_n\\
&= H(Y_1^n,Y_2^n|W_2,G^n)-\underbrace{H(Y_1^n,Y_2^n|W_1,W_2,G^n)}_{=0}+\beta H(Y_2^n|G^n)-\beta H(Y_2^n|W_2,G^n)+n\epsilon_n\\
&= \beta H(Y_2^n|G^n)+H(Y_1^n,Y_2^n|W_2,G^n)-\beta H(Y_2^n|W_2,G^n)+n\epsilon_n\\
&= \beta H(Y_2^n|G^n)+\sum_{t=1}^{n} H(Y_1[t],Y_2[t]|W_2,Y_1^{t-1},Y_2^{t-1},G^n)-\beta\sum_{t=1}^{n} H(Y_2[t]|W_2,Y_2^{t-1},G^n)+n\epsilon_n
\end{aligned}$$

$$\overset{(b)}{\leq} \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(Y_1[t], Y_2[t]|W_2, Y_1^{t-1}, Y_2^{t-1}, X_2^t, G^n)$$

$$- \beta \sum_{t=1}^{n} H(Y_2[t]|W_2, Y_2^{t-1}, X_2^t, G^n) + n\epsilon_n$$

$$= \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t]|W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^n)$$

$$- \beta \sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^n) + n\epsilon_n$$

$$\overset{(c)}{=} \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t]|W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t)$$

$$- \beta \sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) + n\epsilon_n$$

$$\overset{(d)}{\leq} \beta H(Y_2^n|G^n) + n\epsilon_n$$

$$\leq p(1+q)^2 n + n\epsilon_n, \tag{2.58}$$

where (a) holds since the channel gains and the messages are mutually independent; (b) follows from the fact that X_2^t is a deterministic function of $\left(W_2, Y_2^{t-1}\right)$[7] and the fact that conditioning reduces entropy; (c) follows from the fact that condition on W_2, X_1^{t-1}, X_2^t, $X_1[t]$ is independent of the channel realization at future time instants, hence, we can replace G^n by G^t; and (d) follows from Lemma 2.4 below. Dividing both sides by n and let $n \to \infty$, we get

$$R_1 + (1+q)R_2 \leq p(1+q)^2. \tag{2.59}$$

Similarly, we can get $(1+q)R_1 + R_2 \leq p(1+q)^2$.

Lemma 2.4

$$\sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t)$$

[7] We have also added Y_1^{t-1} in the condition for the scenario in which output feedback links are available from each receiver to both transmitters.

$$\geq \frac{1}{2-p} \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t] | W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t). \tag{2.60}$$

Remark 2.15 *Lemma 2.4 is the counterpart of Lemma 2.3 when Output Feedback is available. Note that in the condition we have X_2^t, and due to the presence of output feedback the proof is different than that of Lemma 2.3.*

Proof: We have

$$\begin{aligned}
&H(G_{12}[t]X_1[t] | W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\
&\quad = pH(X_1[t] | G_{12}[t] = 1, W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^{t-1}) \\
&\quad \overset{(a)}{=} pH(X_1[t] | W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^{t-1}) \\
&\quad \overset{(b)}{=} pH(X_1[t] | W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\
&\quad = \frac{p}{1-q^2} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t] | W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\
&\quad \overset{(c)}{\geq} \frac{1}{2-p} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t] | W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t),
\end{aligned} \tag{2.61}$$

where (a) and (b) follow from the fact that condition on W_2, $G_{12}^{t-1}X_1^{t-1}$, X_2^t and G^n, $X_1[t]$ is independent of the channel realization at time t; and (c) holds since conditioning reduces entropy.

Therefore, we have

$$\begin{aligned}
&\sum_{t=1}^{n} H(G_{12}[t]X_1[t] | W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\
&\quad \geq \frac{1}{2-p} \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t] | W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t).
\end{aligned}$$

■

In the following two sections, we consider the last scenario we are interested in, *i.e.* Instantaneous-CSIT and OFB, and we provide the proof of Theorem 2.4. First, we present the achievability strategy, and we demonstrate how OFB can enhance our achievable rate region. We then present the converse proof.

2.9 Achievability Proof of Theorem 2.4 [Instantaneous-CSIT and OFB]

In this section, we describe our achievability strategy for the case of Instantaneous-CSIT and output feedback. Note that in this scenario, although transmitters have instantaneous knowledge of the channel state information, the output signals are available at the transmitters with unit delay. We first provide a brief overview of our scheme.

2.9.1 Overview

By symmetry, it suffices to describe the achievability scheme for corner point

$$(R_1, R_2) = \left(1 - q^2, pq\right),$$

as depicted in Figure 2.26. Similarly, we can achieve corner point $(R_1, R_2) = \left(pq, 1 - q^2\right)$, and therefore by time sharing, we can achieve the region.

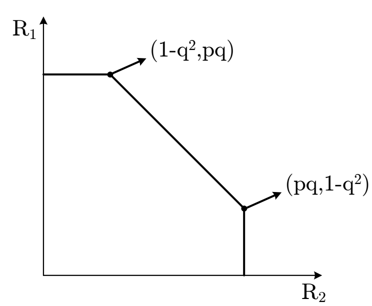


Figure 2.26: Two-user Binary Fading IC: capacity region with Instantaneous-CSIT and output feedback. By symmetry, it suffices to describe the achievability scheme for corner point $(R_1, R_2) = \left(1 - q^2, pq\right)$.

Our achievability strategy is carried on over $b+1$ communication blocks, each block with n time instants. Transmitters communicate fresh data bits in the first b blocks and the final block is to help the receivers decode their corresponding bits. At the end, using our scheme, we achieve a rate tuple arbitrary close to $\frac{b}{b+1}\left(1-q^2, pq\right)$ as $n \to \infty$. Finally letting $b \to \infty$, we achieve the desired rate tuple.

2.9.2 Achievability Strategy

Let W_i^j be the message of transmitter i in block j, $i = 1, 2$, and $j = 1, 2, \ldots, b$. Moreover, let $\mathrm{W}_1^j = a_1^j, a_2^j, \ldots, a_m^j$, and $\mathrm{W}_2^j = b_1^j, b_2^j, \ldots, b_{m_2}^j$, for $j = 1, 2, \ldots, b$, where data bits a_i^j's and b_i^j's are picked uniformly and independently from $\{0, 1\}$, $i = 1, 2, \ldots, m$, and

$$m_2 = \frac{q}{1+q} m. \tag{2.62}$$

We also set $n = m/(1-q^2) + m^{2/3}$, where n is the length of each block.

Achievability strategy for block 1: In the first communication block, at each time instant t, if at least one of the outgoing links from Tx_1 is on, then it sends one of its initial m bits that has not been transmitted before (note that this happens with probability $(1-q^2)$). On the other hand, Tx_2 communicates a new bit (a bit that has not been transmitted before) if the link to its receiver is on and it does not interfere with receiver one (*i.e.* $G_{22}[t] = 1$ and $G_{21}[t] = 0$). In other words, Tx_2 communicates a new bit if either one of Cases $2, 4, 9$, or 11 in Table 2.2 occurs (note that this happens with probability pq).

The first block goes on for n time instants. If at the end of the first block, there exists a bit at either of the transmitters that has not yet been transmitted, we consider it as error type-I and halt the transmission.

Assuming that the transmission in not halted, using output feedback links, transmitter two has access to the bits of transmitter one communicated in the first block. In particular, Tx_2 has access to the bits of Tx_1 transmitted in Cases $2, 11, 12, 14$, and 15 during block 1. Note that the bits communicated in Cases $11, 12, 14$, and 15 from Tx_1 have to be provided to Rx_1. However, the bits communicated in Case 2 from Tx_1 are already available at Rx_1 but needed at Rx_2, see Figure 2.27. Transmitter two will provide such bits to Rx_2 in the following communication block.

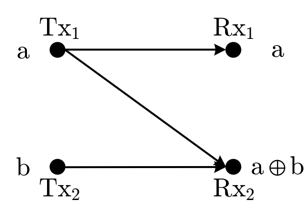


Figure 2.27: The bit communicated in Case 2 from Tx_1 is already available at Rx_1 but it is needed at Rx_2. Transmitter two learns this bit through the feedback channel and will provide it to Rx_2 in the following communication block.

Now, Tx_2 transfers the bits of Tx_1 communicated in Cases $2, 11, 12, 14$, and 15, during the first communication block to queues Q^1_{1,C_2}, $Q^1_{1,\mathsf{C}_{11}}$, $Q^1_{1,\mathsf{C}_{12}}$, $Q^1_{1,\mathsf{C}_{14}}$, and $Q^1_{1,\mathsf{C}_{15}}$ respectively.

Let random variable N^1_{1,C_ℓ} denote the number of bits in Q^1_{1,C_ℓ}, $\ell = 2, 11, 12, 14, 15$. Since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of block 1, we have

$$\mathbb{E}[N^1_{1,\mathsf{C}_\ell}] = \frac{\Pr(\text{Case } \ell)}{1 - \sum_{i=9,10,13,16} \Pr(\text{Case i})} m = (1-q^2)^{-1} \Pr(\text{Case } \ell)\, m, \tag{2.63}$$

for $\ell = 2, 11, 12, 14, 15$.

If the event $\left[N^1_{1,\mathsf{C}_\ell} \geq \mathbb{E}[N^1_{1,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to Q^1_{1,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^1_{1,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Furthermore, using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Achievability strategy for block j, $j = 2, 3, \ldots, b$: The transmission strategy for Tx_1 is the same as block 1 for the first b blocks (all but the last block). In other words, at time instant t, Tx_1 transmits one of its initial m bits (that has not been transmitted before) if at least one of its outgoing links is on. On the other hand, Tx_2 communicates W_2^2 using similar strategy as the first block, *i.e.* Tx_2 communicates a new bit if either one of Cases 2, 4, 9, or 11 occurs.

Transmitter two transfers the bits communicated in Cases 2, 11, 12, 14, and 15, during communication block j to queues Q^j_{1,C_2}, $Q^j_{1,\mathsf{C}_{11}}$, $Q^j_{1,\mathsf{C}_{12}}$, $Q^j_{1,\mathsf{C}_{14}}$, and $Q^j_{1,\mathsf{C}_{15}}$ respectively.

Moreover, at time instant t,

- if Case 3 occurs, Tx_2 sends one of the bits from Q^{j-1}_{1,C_2} and removes it from this queue since it has been delivered successfully to Rx_2, see Figure 2.28. If Case 3 occurs and Q^{j-1}_{1,C_2} is empty, Tx_2 remains silent;
- if Case 10 occurs, Tx_2 sends one of the bits from $Q^{j-1}_{1,\mathsf{C}_{11}}$ and removes it from this queue, see Figure 2.29. If Case 10 occurs and $Q^{j-1}_{1,\mathsf{C}_{11}}$ is empty, Tx_2 remains silent;
- if Case 12 occurs, it sends one of the bits from $Q^{j-1}_{1,\mathsf{C}_{12}}$ and removes it from

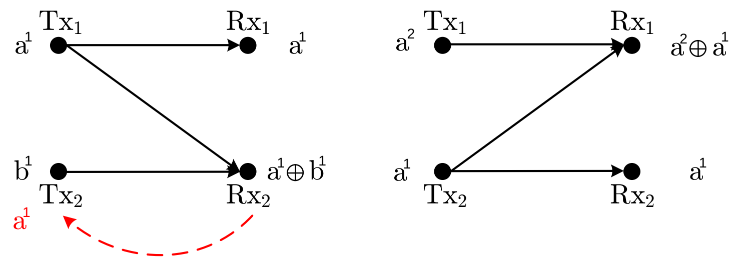


Figure 2.28: In block j when Case 3 occurs, Tx_2 retransmits the bit of Tx_1 communicated in Case 2 during block $j-1$. Note that this bit does not cause interference at Rx_1 and it is needed at Rx_2.

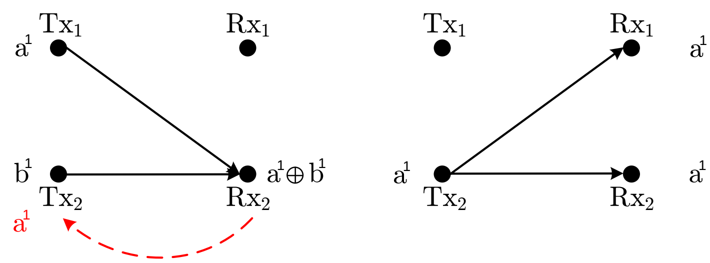


Figure 2.29: In block j when Case 10 occurs, Tx_2 retransmits the bit of Tx_1 communicated in Case 11 during block $j-1$. Note that this bit is needed at both receivers.

this queue. If Case 12 occurs and $Q^{j-1}_{1,\mathsf{C}_{12}}$ is empty, Tx_2 remains silent;

- if Case 13 occurs, it sends one of the bits from $Q^{j-1}_{1,\mathsf{C}_{14}}$ and removes it from this queue. If Case 13 occurs and $Q^{j-1}_{1,\mathsf{C}_{14}}$ is empty, Tx_2 remains silent;
- if Case 15 occurs, it sends one of the bits from $Q^{j-1}_{1,\mathsf{C}_{15}}$ and removes it from this queue. If Case 15 occurs and $Q^{j-1}_{1,\mathsf{C}_{15}}$ is empty, Tx_2 remains silent.

If at the end of block j, there exists a bit at either of the transmitters that has not yet been transmitted, or any of the queues Q^{j-1}_{1,C_2}, $Q^{j-1}_{1,\mathsf{C}_{11}}$, $Q^{j-1}_{1,\mathsf{C}_{12}}$, $Q^{j-1}_{1,\mathsf{C}_{14}}$, or $Q^{j-1}_{1,\mathsf{C}_{15}}$ is not empty, we consider this event as error type-I and halt the transmission.

Assuming that the transmission is not halted, let random variable N^j_{1,C_ℓ} denote the number of bits in Q^j_{1,C_ℓ}, $\ell = 2, 11, 12, 14, 15$. From basic probability, we have

$$\mathbb{E}[N^j_{1,\mathsf{C}_\ell}] = \frac{\Pr(\text{Case } \ell)}{1-\sum_{i=9,10,13,16}\Pr(\text{Case i})} m = (1-q^2)^{-1}\Pr(\text{Case } \ell)\, m, \tag{2.64}$$

for $\ell = 2, 11, 12, 14, 15$.

If the event $\left[N^j_{1,\mathsf{C}_\ell} \geq \mathbb{E}[N^j_{1,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to Q^j_{1,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^j_{1,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II and decreases exponentially with m.

Achievability strategy for block $b+1$: Finally in block $b+1$, no new data bit is transmitted (*i.e.* $\mathsf{W}_1^{b+1}, \mathsf{W}_2^{b+1} = 0$), and Tx_2 only communicates the bits of Tx_1 communicated in the previous block in Cases $2, 11, 12, 14$, and 15 as described above. If at the end of block $b+1$, any of the queues Q^b_{1,C_2}, $Q^b_{1,\mathsf{C}_{11}}$, $Q^b_{1,\mathsf{C}_{12}}$, $Q^b_{1,\mathsf{C}_{14}}$, or $Q^b_{1,\mathsf{C}_{15}}$ is not empty, we consider this event as error type-I and halt the transmission.

The probability that the transmission strategy halts at the end of each block can be bounded by the summation of error probabilities of types I and II. Using Chernoff-Hoeffding bound, we can show that the probability that the transmission strategy halts at any point approaches zero as $m \to \infty$.

2.9.3 Decoding

At the end of block $j+1$, Rx_1 has acces to W_1^j with no interference, $j = 1, 2, \ldots, b$. At the end of block $b+1$, Rx_2 uses the bits communicated in Cases 3 and 10 from Tx_2 to cancel out the interference it has received from Tx_1 during the previous block in Cases 2 and 11. Therefore, at the end of block $b+1$, Rx_1 has access to W_2^b with no interference. Then, Rx_2 follows the same strategy for blocks b and $b-1$. Therefore, using similar idea, Rx_2 uses backward decoding to cancel out interference in the previous blocks to decode all messages.

Now, since each block has $n = m/(1-q^2)+m^{2/3}$ time instants and the probability that the transmission strategy halts at any point approaches zero for $m \to \infty$, we achieve a rate tuple

$$\frac{b}{b+1}\left(1-q^2, pq\right), \tag{2.65}$$

as $m \to \infty$. Finally letting $b \to \infty$, we achieve the desired rate tuple.

2.10 Converse Proof of Theorem 2.4 [Instantaneous-CSIT and OFB]

To derive the outer-bound on individual rates, we have

$$\begin{aligned}
nR_1 &= H(W_1) \overset{(a)}{=} H(W_1|G^n) \\
&\overset{\text{(Fano)}}{\leq} I(W_1; Y_1^n|G^n) + n\epsilon_n \\
&= H(Y_1^n|G^n) - H(Y_1^n|W_1, G^n) + n\epsilon_n \\
&\leq H(Y_1^n|G^n) + n\epsilon_n \\
&\leq (1-q^2)n + n\epsilon_n,
\end{aligned} \tag{2.66}$$

where $\epsilon_n \to 0$ as $n \to \infty$; (a) holds since message W_1 is independent of G^n. Similarly, we have

$$nR_2 \leq (1-q^2)n + n\epsilon_n, \tag{2.67}$$

dividing both sides by n and let $n \to \infty$, we have

$$\begin{cases} R_1 \leq 1 - q^2, \\ R_2 \leq 1 - q^2. \end{cases} \tag{2.68}$$

The outer-bound on $R_1 + R_2$, *i.e.*

$$R_1 + R_2 \leq 1 - q^2 + pq, \tag{2.69}$$

can be obtained as follows.

$$\begin{aligned}
& n(R_1 + R_2 - 2\epsilon_n) \\
& \overset{(a)}{\leq} H(W_1|W_2, G^n) + H(W_2|G^n) \\
& \overset{\text{Fano}}{\leq} I(W_1; Y_1^n|W_2, G^n) + I(W_2; Y_2^n|G^n) \\
& = H(Y_1^n|W_2, G^n) - \underbrace{H(Y_1^n|W_1, W_2, G^n)}_{=\,0} + I(W_2; Y_2^n|G^n) \\
& = H(Y_1^n|W_2, G^n) + H(Y_2^n|G^n) - H(Y_2^n|W_2, G^n) \\
& = H(Y_1^n|W_2, G^n) + H(Y_2^n|G^n) - \left[H(Y_1^n, Y_2^n|W_2, G^n) - H(Y_1^n|Y_2^n, W_2, G^n)\right] \\
& = H(Y_1^n|Y_2^n, W_2, G^n) + H(Y_2^n|G^n) \\
& \overset{(b)}{=} H(Y_2^n|G^n) + \sum_{t=1}^{n} H(Y_1[t]|W_2, Y_2^n, Y_1^{t-1}, X_2^t, G_{12}^t X_1^t, G^n) \\
& \overset{(c)}{\leq} H(Y_2^n|G^n) + H(Y_1^n|G_{12}^n X_1^n, G_{21}^n X_2^n, G^n) \\
& \overset{(d)}{\leq} \sum_{t=1}^{n} H(Y_2[t]|G^n) + \sum_{t=1}^{n} H(Y_1[t]|G_{12}[t]X_1[t], G_{21}[t]X_2[t], G^n) \\
& \overset{(e)}{\leq} \left(1 - q^2\right) n + pqn,
\end{aligned} \tag{2.70}$$

where $\epsilon_n \to 0$ as $n \to \infty$; and (a) follows from the fact that the messages and G^n are mutually independent; (b) holds since X_2^t is a function of $W_2, Y_1^{t-1}, Y_2^{t-1}$, and G^t; (c) and (d) follow from the fact that conditioning reduces entropy; and (e) holds since

$$H(Y_2[t]|G^n) \leq 1 - q^2,$$

$$H(Y_1[t]|G_{12}[t]X_1[t], G_{21}[t]X_2[t], G^n) \leq pq. \tag{2.71}$$

Dividing both sides by n and let $n \to \infty$, we get

$$R_1 + R_2 \leq 1 - q^2 + pq. \tag{2.72}$$

We also note this outer-bound on $R_1 + R_2$ can be also applied to the case of Instantaneous-CSIT and no output feedback (*i.e.* Theorem 2.1).

2.11 Extension to the Non-Homogeneous Setting

In this section, we discuss the extension of our results to the non-homogeneous case. More precisely, we consider the two-user Binary Fading Interference Channel of Section 2.2 where

$$G_{ii}[t] \overset{d}{\sim} \mathcal{B}(p_d), \qquad G_{\bar{i}i}[t] \overset{d}{\sim} \mathcal{B}(p_c), \tag{2.73}$$

for $0 \leq p_d, p_c \leq 1$, $\bar{i} = 3 - i$, and $i = 1, 2$. We define $q_d = 1 - p_d$ and $q_c = 1 - p_c$. We study the non-homogeneous BFIC in two settings: (1) Delayed-CSIT and output feedback; and (2) Delayed-CSIT (and no output feedback). For the case of Delayed-CSIT and output feedback, we fully characterize the capacity region as follows.

Theorem 2.5 *The capacity region of the two-user Binary Fading IC with Delayed-CSIT and output feedback, $\mathcal{C}^{\text{DCSIT,OFB}}(p_d, p_c)$ is given by*

$$\mathcal{C}^{\text{DCSIT,OFB}}(p_d, p_c) = \left\{R_1, R_2 \in \mathbb{R}^+ \ s.t. \ p_c R_i + (1 - q_d q_c) R_{\bar{i}} \leq (1 - q_d q_c)^2, \ i = 1, 2\right\}. \tag{2.74}$$

Proof: We first prove the converse. The converse proof follows similar steps as the case of the homogeneous setting described in Scetion 2.8 for Theorem 2.3. Set[8]

$$\beta = \frac{(1 - q_d q_c)}{p_c}. \tag{2.75}$$

We have

$$\begin{aligned}
n\left(R_1 + \beta R_2\right) &= H(W_1) + \beta H(W_2) \\
&\stackrel{(a)}{=} H(W_1|W_2, G^n) + \beta H(W_2|G^n) \\
&\stackrel{\text{Fano}}{\leq} I(W_1; Y_1^n|W_2, G^n) + \beta I(W_2; Y_2^n|G^n) + n\epsilon_n \\
&\leq I(W_1; Y_1^n, Y_2^n|W_2, G^n) + \beta I(W_2; Y_2^n|G^n) + n\epsilon_n \\
&= H(Y_1^n, Y_2^n|W_2, G^n) - \underbrace{H(Y_1^n, Y_2^n|W_1, W_2, G^n)}_{=0} + \beta H(Y_2^n|G^n) - \beta H(Y_2^n|W_2, G^n) + n\epsilon_n \\
&= \beta H(Y_2^n|G^n) + H(Y_1^n, Y_2^n|W_2, G^n) - \beta H(Y_2^n|W_2, G^n) + n\epsilon_n \\
&= \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(Y_1[t], Y_2[t]|W_2, Y_1^{t-1}, Y_2^{t-1}, G^n) - \beta \sum_{t=1}^{n} H(Y_2[t]|W_2, Y_2^{t-1}, G^n) + n\epsilon_n \\
&\stackrel{(b)}{\leq} \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(Y_1[t], Y_2[t]|W_2, Y_1^{t-1}, Y_2^{t-1}, X_2^t, G^n) - \beta \sum_{t=1}^{n} H(Y_2[t]|W_2, Y_2^{t-1}, X_2^t, G^n) + n\epsilon_n \\
&= \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t]|W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^n) \\
&\quad - \beta \sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^n) + n\epsilon_n \\
&\stackrel{(c)}{=} \beta H(Y_2^n|G^n) + \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t]|W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\
&\quad - \beta \sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) + n\epsilon_n \\
&\stackrel{(d)}{\leq} \beta H(Y_2^n|G^n) + n\epsilon_n \\
&\leq \frac{(1 - q_d q_c)^2}{p_c} n + n\epsilon_n,
\end{aligned} \tag{2.76}$$

[8]For $p_c = 0$ the result is trivial, so we assume that β is well defined.

where (a) holds since the channel gains and the messages are mutually independent; (b) follows from the fact that X_2^t is a deterministic function of $\left(W_2, Y_2^{t-1}\right)$[9] and the fact that conditioning reduces entropy; (c) follows from the fact that condition on W_2, X_1^{t-1}, X_2^t, $X_1[t]$ is independent of the channel realization at future time instants, hence, we can replace G^n by G^t; and (d) follows from Lemma 2.5 below. Dividing both sides by n and let $n \rightarrow \infty$, we get

$$p_c R_1 + (1 - q_d q_c) R_2 \leq (1 - q_d q_c)^2, \tag{2.77}$$

and the derivation of the other bound would be similar.

Lemma 2.5 [Non-Homogeneous Entropy Leakage with Output Feedback] *For the broadcast channel described in Fig. 2.24 with parameters p_d and p_c, and with Delayed-CSIT and output feedback, for* any *input distribution, we have*

$$\begin{aligned}(1 - q_d q_c) &\sum_{t=1}^{n} H(G_{12}[t]X_1[t]|W_2, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t) \\ &\geq p_c \sum_{t=1}^{n} H(G_{11}[t]X_1[t], G_{12}[t]X_1[t]|W_2, G_{11}^{t-1}X_1^{t-1}, G_{12}^{t-1}X_1^{t-1}, X_2^t, G^t).\end{aligned} \tag{2.78}$$

The proof of Lemma 2.5 follows the same steps as the proof of Lemma 2.4. We note that this lemma, is the generalization of the Entropy Leakage Lemma to the case where $G_1[t] \overset{d}{\sim} \mathcal{B}(p_d)$ and $G_2[t] \overset{d}{\sim} \mathcal{B}(p_c)$ and where the output feedback is present.

We now describe the achievability proof. The achievability proof is also similar to that of the homogeneous setting described in Section 2.7. Hence, we provide an outline of the achievability strategy here. The achievability strategy

[9] We have also added Y_1^{t-1} in the condition for the scenario in which output feedback links are available from each receiver to both transmitters.

of corner points $(1 - q_d q_c, 0)$ and $(0, 1 - q_d q_c)$, is based on utilizing the additional communication paths created by the means of the output feedback links, *e.g.*,

$$\mathsf{Tx}_1 \rightarrow \mathsf{Rx}_2 \rightarrow \mathsf{Tx}_2 \rightarrow \mathsf{Rx}_1.$$

In the rest of the proof, we provide the outline for the achievability of corner point

$$R_1 = R_2 = \frac{(1 - q_d q_c)}{1 + (1 - q_d q_c)^{-1} p_c}. \tag{2.79}$$

The strategy is carried on over two phases similar to Phase 1 and Phase 2 of Section 2.7. We assume that at the beginning of the communication block, there are m bits in $Q_{i \rightarrow i}$, $i = 1, 2$. Phase 1 is the uncategorized transmission, and it goes on for

$$(1 - q_d q_c)^{-1} m + m^{\frac{2}{3}} \tag{2.80}$$

time instants and if at the end of this phase, either of the queues $Q_{1 \rightarrow 1}$ or $Q_{2 \rightarrow 2}$ is not empty, we declare an error and halt the transmission. Upon completion of Phase 1, using the ideas described in Section 2.4 for output feedback, we further create bits of common interest. More precisely, we use the following ideas: we update the status of bits in Q_{i,C_1} to bits of common interest; as described in Example 6 of Section 2.4, using output feedback, we combine bits in $Q_{1 \rightarrow 1|2}$ and $Q_{2 \rightarrow 2|1}$ to create bits of common interest; as described in Example 7 of Section 2.4, using output feedback, we combine bits in $Q_{1 \rightarrow 2|1}$ and $Q_{2 \rightarrow 1|2}$ to create bits of common interest.

In the second phase, we deliver the bits in $Q_{1 \rightarrow \{1,2\}}$ and $Q_{2 \rightarrow \{1,2\}}$ using the transmission strategy for the two-multicast problem. For

$$0 \leq p_c \leq \frac{p_d}{1 + p_d}, \tag{2.81}$$

the cross links become the bottleneck for the two-multicast network depicted in Figure 2.23, and as a result, using the two-multicast problem as discussed in Lemma 2.2 is sub-optimal. However, using the output feedback link, Tx_i learns the interfering bit of $\mathsf{Tx}_{\bar{i}}$, $i = 1, 2$; and considering this side information available at the transmitters, we can show that a sum-rate of $(1 - q_d q_c)$ for the two-multicast problem is in fact achievable. At the end of Phase 2, all bits are delivered. It takes

$$(1 - q_d q_c)^{-2} \, p_c m + O\left(m^{\frac{2}{3}}\right) \tag{2.82}$$

time instants to complete Phase 2. Therefore, we achieve a symmetric sum-rate of

$$R_1 = R_2 = \frac{(1 - q_d q_c)}{1 + (1 - q_d q_c)^{-1} p_c}, \tag{2.83}$$

as $m \to \infty$. ■

For the case of Delayed-CSIT (and no output feedback), we partially solve the problem as described below.

Theorem 2.6 [Capacity Region with Delayed-CSIT and Non-Homogeneous Channel Gains] *The capacity region of the two-user Binary Fading IC with Delayed-CSIT (and no output feedback),* $\mathcal{C}^{\text{DCSIT}}(p_d, p_c)$ *for*

$$\frac{p_d}{1 + p_d} \leq p_c \leq 1, \tag{2.84}$$

is the set of all rate tuples (R_1, R_2) *satisfying*

$$\mathcal{C}^{\text{DCSIT}}(p_d, p_c) = \begin{cases} 0 \leq R_i \leq p_d, & i = 1, 2, \\ p_c R_i + (1 - q_d q_c) R_{\bar{i}} \leq (1 - q_d q_c)^2, & i = 1, 2. \end{cases} \tag{2.85}$$

Proof: The proof of converse follows from the previous theorem since the outer-bound of Theorem 2.5 also serves as an outer-bound for Theorem 2.6. Here, we discuss the achievability strategy. The achievability proof is similar to that of the homogeneous setting as described in Section 2.5 for Theorem 2.2. The corner points are as follows.

$$
\begin{aligned}
&(R_1, R_2) = \left(\min\left\{p_d, \frac{(1-q_dq_c)}{1+(1-q_dq_c)^{-1}p_c}\right\}, \min\left\{p_d, \frac{(1-q_dq_c)}{1+(1-q_dq_c)^{-1}p_c}\right\}\right),\\
&(R_1, R_2) = \left(p_d, \min\left\{p_d, (1-q_dq_c)\, q_d\right\}\right),\\
&(R_1, R_2) = \left(\min\left\{p_d, (1-q_dq_c)\, q_d\right\}, p_d\right).
\end{aligned}
\tag{2.86}
$$

Here, we provide the outline for the achievability of the first corner point, *i.e.*

$$
R_1 = R_2 = \min\left\{p_d, \frac{(1-q_dq_c)}{1+(1-q_dq_c)^{-1}p_c}\right\}. \tag{2.87}
$$

We assume that at the beginning of the communication block, there are m bits in $Q_{i\to i}$, $i = 1, 2$. Phase 1 is the uncategorized transmission. Upon completion of Phase 1, using the ideas described in Section 2.4.1, we upgrade the status of the bits to bits of common interest. We use the following coding opportunities.

- **Type I** Combining bits in $Q_{i\to\bar{i}|i}$ and $Q_{i\to i|\bar{i}}$ to create bits of common interest, $i = 1, 2$;
- **Type II** Combining the bits in Q_{i,C_1} and $Q_{i\to i|\bar{i}}$ to create bits of common interest, $i = 1, 2$;
- **Type III** Combining the bits in Q_{i,C_1} and $Q_{i\to\bar{i}|i}$ to create bits of common interest, $i = 1, 2$.

Then, in the final phase, we deliver the bits in $Q_{1\to\{1,2\}}$ and $Q_{2\to\{1,2\}}$ using the transmission strategy for the two-multicast problem. ■

Remark 2.16 *We note that in Theorem 2.6, we partially characterized the capacity region. In fact, for*

$$0 \leq p_c \leq \frac{p_d}{1 + p_d}, \tag{2.88}$$

our achievability region does not match the outer-bounds. Closing the gap in this regime could be an interesting future direction. The reason one might think the achievability could be improved is that in this regime the cross links become the bottleneck for the two-multicast network depicted in Figure 2.23, and as a result, using the two-multicast problem might be sub-optimal. On the other hand, as we demonstrated in [64], even under No-CSIT assumption, this regime requires a different outer-bound compared to other regimes.

2.12 Conclusion and Future Directions

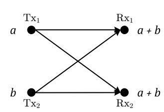
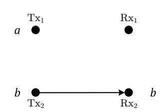
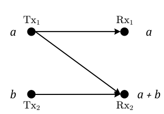
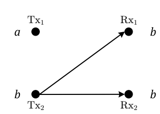
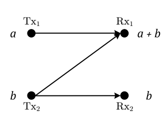
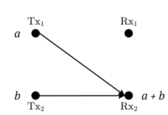
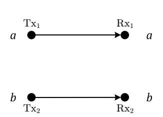
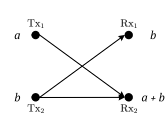
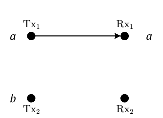
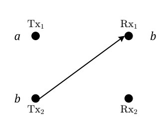
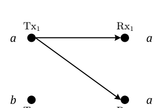
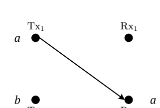
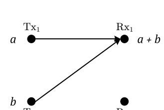
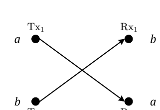
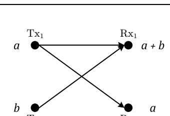
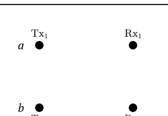
We studied the impact of delayed knowledge of the channel state information at the transmitters, on the capacity region of the two-user binary fading interference channels. We introduced various coding opportunities, created by Delayed-CSIT, and presented an achievability strategy that systematically exploits the coding opportunities. We derived an achievable rate region that matches the outer-bounds for this problem, hence, characterizing the capacity region. We have also derived the capacity region of this problem with Delayed-CSIT and output feedback.

A future direction would be to extend our results to the case of two-user Gaussian fading interference channel with Delayed-CSIT. As discussed in the introduction, one can view our binary fading model as a fading interpretation of the linear deterministic model where the non-negative integer associated to each link is at most 1. Therefore, one approach is to extend the current results to the case of fading linear deterministic interference channel and then, further extend that result to the case of Gaussian fading interference channel, in order to obtain approximate capacity characterization. This approach has been taken for the No-CSIT assumption in [55], in the context of fading broadcast channels, and in [78] in the context of one-sided fading interference channels. In fact, one can view our Binary Fading model as the model introduced in [55, 78] with only one layer.

Another future direction is to consider the k-user setting of the problem. In [38], authors have shown that for the k-user fading interference channel with instantaneous knowledge of the channel state information, sum degrees of freedom (DoF) of $k/2$ is achievable. However, in the absence of the CSIT, the achievable sum DoF collapses to 1. As a result a large degradation in network capacity, due to lack of the CSIT, is observed. It has been recently shown that, with Delayed-CSIT, it is possible to achieve more than one sum DoF [1, 37], however, the achievable sum DoFs are less than 1.5 for any number of users. This together with lack of nontrivial DoF upper bounds leaves the problem of sum DoF characterization of interference channels with Delayed-CSIT still open and challenging, to the extent that it is even unknown whether the sum DoF of such networks *scales* with the number of users or not. A promising direction may be to study this problem in the context of our simpler binary fading model, to understand whether the sum capacity of such network with Delayed-CSIT scales with the number of users or it will saturate.

Finally, motivated by recent results that demonstrate that, with Instantaneous-CSIT, multi-hopping can significantly increase the capacity of interference networks (*e.g.*, [26, 49] for two-unicast networks and [48] for multi-unicast networks), an interesting future direction would be explore the impact of Delayed-CSIT on the capacity of muti-hop binary interference networks.

Table 2.2: All possible channel realizations and transitions from the initial queue to other queues; solid arrow from tranmsitter Tx_i to receiver Rx_j indicates that $G_{ij}[t] = 1$, $i, j \in \{1, 2\}$, $t = 1, 2, \ldots, n$. Bit "a" represents a bit in $Q_{1\to1}$ while bit "b" represents a bit in $Q_{2\to2}$.

ID	ch. at time n	transition	ID	ch. at time n	transition
1		$\begin{cases} a \to Q_{1,\mathsf{C}_1} \\ b \to Q_{2,\mathsf{C}_1} \end{cases}$	9		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to F} \end{cases}$
2		$\begin{cases} a \to Q_{1\to2\|1} \\ b \to Q_{2\to F} \end{cases}$	10		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to F} \end{cases}$
3		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to1\|2} \end{cases}$	11		$\begin{cases} a \to Q_{1\to\{1,2\}} \\ b \to Q_{2\to F} \end{cases}$
4		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to F} \end{cases}$	12		$\begin{cases} a \to Q_{1\to\{1,2\}} \\ b \to Q_{2\to F} \end{cases}$
5		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to2} \end{cases}$	13		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to2\|1} \end{cases}$
6		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to2} \end{cases}$	14		$\begin{cases} a \to Q_{1\to1\|2} \\ b \to Q_{2\to2} \end{cases}$
7		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to\{1,2\}} \end{cases}$	15		$\begin{cases} a \to Q_{1\to1\|2} \\ b \to Q_{2\to2\|1} \end{cases}$
8		$\begin{cases} a \to Q_{1\to F} \\ b \to Q_{2\to\{1,2\}} \end{cases}$	16		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to2} \end{cases}$

CHAPTER 3

CAPACITY RESULTS FOR GAUSSIAN NETWORKS WITH DELAYED CHANNEL STATE INFORMATION

3.1 Introduction

In wireless networks, receivers estimate the channel state information (CSI) and pass this information to the transmitters through feedback mechanisms. The extent to which channel state information is available at the transmitters has a direct impact on the capacity of wireless networks and the optimal strategies. In fast-fading scenarios where the coherence time of the channel is smaller than the delay of the feedback channel, providing the transmitters with up-to-date CSI is practically infeasible. Consequently, we are left with no choice but to understand the behavior of wireless networks under such constraint.

Our objective is to understand the effect of lack of up-to-date CSI on the *capacity* region of wireless networks by considering a fundamental building block, namely the multiple-input single-output (MISO) broadcast channel (BC). In the context of MISO BC, it has been shown that even completely stale CSIT (also known as delayed CSIT) can still be very useful and can change the scale of the capacity, measured by the degrees of freedom (DoF) [36]. The degrees of freedom by definition provides a first order approximation of the capacity, thus it is mainly useful in understanding the behavior of the capacity in high power regimes. However, it is not a suitable measure for practical settings with finite signal-to-noise ratio (SNR).

We first consider the two-user MISO BC and we focus on the effect of delayed CSIT at finite SNR regimes as opposed to the asymptotic DoF analysis. While there is a strong body of work on broadcast channels with perfect channel state information (see [40, 72, 73]), no capacity result has been reported for the delayed CSIT scenario. There are some prior results in the literature (for example [75]) that have proposed and analyzed several achievability strategies at finite SNR regimes. Nonetheless, characterizing the capacity region of the two-user MISO BC with delayed CSIT has remained open.

In this paper, we provide the first constant-gap approximation of the capacity region of the two-user MISO BC with delayed CSIT. We obtain an achievable scheme and an outer-bound on the capacity region, and we analytically show that they are within 1.81 bits/sec/Hz per user, for all values of the transmit power. Our numerical analysis shows that the gap is in fact smaller and in the worst case, it is at most 1.1 bits/sec/Hz per user.

The proposed achievability scheme for the two-user MISO BC with delayed CSIT has three phases. In Phase 1 and Phase 2, transmitter respectively sends messages intended for receivers one and two. In each one of these phases, the unintended receiver overhears and saves some signal (interference) which is only useful for the other receiver. At this point, transmitter can evaluate the overheard signals using the delayed CSI. In the third phase, the transmitter will swap the overheard signals between the receivers by sending a signal of common interest. This signal is the quantized version of the summation of the previously transmitted signals. The swapping is performed by exploiting the overheard signals as available side-information at receivers side. The overall information that each receiver collects in the three phases is enough to decode the intended message. Although the above three phases follows the scheme of [36], as we will show, some important additional ingredients are needed to make an approximately optimal scheme.

To derive the outer-bound, we create a physically degraded BC by providing the received signal of one user one to the other user. Then, since we know that feedback does not enlarge the capacity region of a physically degraded BC [22], we ignore the delayed knowledge of the channel state information at the transmitter (*i.e.* no CSIT assumption). We then derive an outer-bound on the capacity region of this degraded broadcast channel which in turn serves as an outer-bound on the capacity region of the two-user MISO BC with delayed CSIT. We show that the achievable rate region and the outer-bound are within 1.81 bits/sec/Hz per user. Using numerical analysis, we can show that the gap is in fact smaller than 1.1 bits/sec/Hz per user.

We then consider the K-user MISO BC with delayed CSIT and we focus on the symmetric capacity. We show how to extend our ideas for the achievability and converse to this setting, and we show that for the symmetric capacity the gap between the achievable rate and the outer-bound is less than $2\log_2(K+2)$ bits/sec/Hz independent of the transmit power.

In the literature, there have been several results on the impact of delayed CSIT in wireless networks. However, these results are either focused on the DoF region (*e.g*, [13, 52, 68]) or capacity results for noiseless channels (*e.g.*, [63, 67]). The present work provides constant-gap approximation of the capacity region of the two-user Complex Gaussian MISO BC with delayed CSIT which is of great importance in practical wireless settings. Existing results on constant-gap approximation of the capacity region of wireless networks mainly consider the scenario in which transmitters have perfect instantaneous knowledge of the channel state information (*e.g*, [11, 50, 59]). Thus the current result opens the door for constant-gap approximation of the capacity region of wireless networks with delayed CSIT.

The rest of the paper is organized as follows. In Section 3.2, we formulate our problem. In Section 3.3, we present our main results. We describe our achievability strategy in Section 3.4. Section 3.5 is dedicated to deriving the outer-bound. In Section 3.6, we show that our inner-bound and outer-bound are within constant number of bits. We extend our results to the K-user MISO BC with delayed CSIT in Section 3.7. Finally, Section 3.8 concludes the paper.

3.2 Problem Setting

We start by considering the two-user multiple-input single-output (MISO) complex Gaussian broadcast channel (BC) with Rayleigh fading as depicted in Fig. 3.1. The channel gains from the transmitter to receivers one and two are denoted by $\mathbf{h}[t], \mathbf{g}[t] \in \mathbb{C}^{2\times 1}$, respectively, where the entries of $\mathbf{h}[t]$ and $\mathbf{g}[t]$ are distributed as i.i.d. $\mathcal{CN}(0,1)$ (independent across time, antenna, and users). At each receiver, the received signal can be expressed as follows.

$$y_1[t] = \mathbf{h}^\top[t]\mathbf{x}[t] + z_1[t], \qquad y_2[t] = \mathbf{g}^\top[t]\mathbf{x}[t] + z_2[t], \tag{3.1}$$

where $\mathbf{x}[t] \in \mathbb{C}^{2\times 1}$ is the transmit signal subject to average power constraint P, *i.e.* $\mathbb{E}\left[\mathbf{x}^\dagger[t]\mathbf{x}[t]\right] \leq P$ for $P > 0$. The noise processes are independent from the transmit signal and are distributed i.i.d. as $z_k[t] \sim \mathcal{CN}(0,1)$. Furthermore, we define

$$s_1[t] = \mathbf{h}^\top[t]\mathbf{x}[t], \qquad s_2[t] = \mathbf{g}^\top[t]\mathbf{x}[t], \tag{3.2}$$

to be the noiseless versions of the received signals.

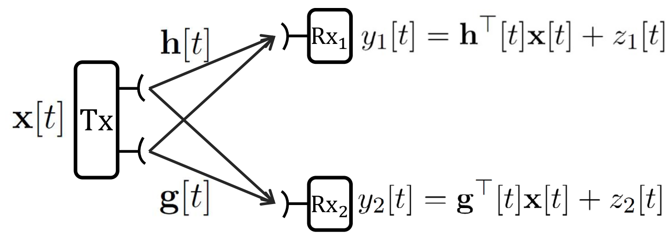


Figure 3.1: Two-user Multiple-Input Single-Output (MISO) Complex Gaussian Broadcast Channel.

Transmitter wishes to reliably communicate independent and uniformly distributed messages $w_1 \in \{1, 2, \ldots, 2^{nR_1}\}$ and $w_2 \in \{1, 2, \ldots, 2^{nR_2}\}$ to receivers 1 and 2, respectively, during n uses of the channel. We denote the channel state information at time t by

$$(\mathbf{h}[t], \mathbf{g}[t]), \qquad t = 1, 2, \ldots, n. \tag{3.3}$$

The transmitter has access to the delayed (outdated) channel state information, meaning that at time instant t, the transmitter has access to

$$(\mathbf{h}[\ell], \mathbf{g}[\ell])_{\ell=1}^{t-1}, \qquad t = 1, 2, \ldots, n. \tag{3.4}$$

Due to the delayed knowledge of the channel state information, the encoded signal $\mathbf{x}[t]$ is a function of both the messages and the previous channel realizations.

Each receiver k, $k = 1, 2$, uses a decoding function $\varphi_{k,n}$ to get the estimate $\hat{w}_k$ from the channel outputs $\{y_k[t] : t = 1, \ldots, n\}$. An error occurs whenever $\hat{w}_k \neq w_k$. The average probability of error is given by

$$\lambda_{k,n} = \mathbb{E}[P(\hat{w}_k \neq w_k)], \qquad k = 1, 2, \tag{3.5}$$

where the expectation is taken with respect to the random choice of the transmitted messages w_1 and w_2.

We say that a rate pair (R_1, R_2) is achievable, if there exists a block encoder at the transmitter, and a block decoder at each receiver, such that $\lambda_{k,n}$ goes to zero as the block length n goes to infinity, $k = 1, 2$. The capacity region C is the closure of the set of the achievable rate pairs.

3.3 Statement of Main Result

Our main contributions are: (1) characterization of the capacity region of the two-user MISO complex Gaussian BC with delayed CSIT to within 1.6 bits/sec/Hz; and (2) characterizing the symmetric capacity of the K-user MISO complex Gaussian BC to within $2\log_2(K+2)$ bits/sec/Hz.

For the two-user setting, the achievability scheme has three phases. In Phase 1 and Phase 2, the transmitter respectively sends messages intended for receiver one and receiver two. In each of these phases, the unintended receiver overhears and saves some signal (interference), which is only useful for the other receiver. Later, in the third phase, the transmitter evaluates what each receiver overheard about the other receiver's message using the delayed knowledge of the channel state information and provides these overheard signals efficiently to both receivers exploiting available side information at each receiver. Transmitter provides the overheard signals to the receivers by sending a signal of common interest. This way transmitter reduces the overall communication time and in turn, increases the achievable rate.

The outer-bound is derived based on creating a physically degraded broadcast channel where one receiver is enhanced by having two antennas. In this channel, feedback and in particular delayed knowledge of the channel state information, does not increase the capacity region. Thus, we can ignore the delayed knowledge of the channel state information and consider a degraded BC with no CSIT. This would provide us with the outer-bound.

We then show how to extend our arguments for achievability and converse to the K-user setting to derive approximate symmetric capacity under delayed CSIT assumption. Before stating our results, we need to define some notations.

Definition 3.1 *For a region $\mathcal{R} \subseteq \mathbb{R}^2$, we define*

$$\mathcal{R} \ominus (\tau, \tau) \stackrel{\triangle}{=} \{(R_1, R_2) \,|R_1, R_2 \geq 0, (R_1 + \tau, R_2 + \tau) \in \mathcal{R}\}. \tag{3.6}$$

Definition 3.2 *The ergodic capacity of a point-to-point complex Gaussian channel with k transmit antennas and ℓ receive antennas with no CSIT, transmit power P, and additive zero-mean Gaussian noise process of variance m_j at the receive antenna j, is denoted by*

$$C_{k\times\ell}\left(P; m_1, m_2, \ldots, m_\ell\right). \tag{3.7}$$

For simplicity of notations, we drop P, and whenever noise variances are all equal to 1, we do not mention them. Fig. 3.2 depicts a point-to-point complex Gaussian channel with k transmit antennas and k receive antennas.

Definition 3.3 *Rate region C_k, $k = 1, 2$, is defined as*

$$C_k = \{R_1, R_2 \geq 0 \,|R_k + 2R_{\bar{k}} \leq 2C_{2\times1}\} \quad k = 1, 2, \tag{3.8}$$

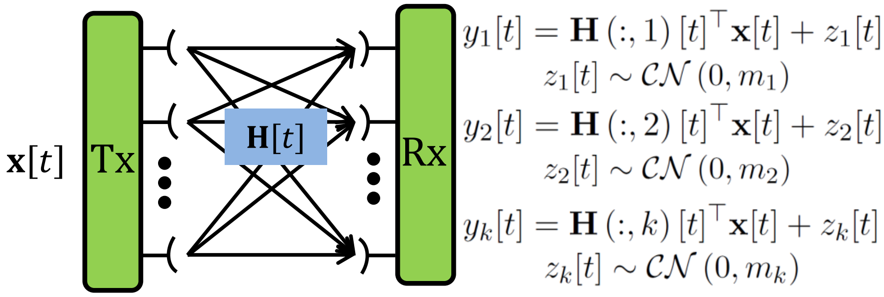


Figure 3.2: Point-to-point complex Gaussian channel with k transmit and receive antennas and no CSIT, where $z_j[t] \sim \mathcal{CN}\left(0, m_j\right)$. Here, $\mathbf{H}[t]$ denotes the channel transfer matrix at time t; and $\mathbf{H}(:, j)[t]$ denotes the channel gains from all transmit antennas to receive antenna j at time t.

where $\bar{k} \triangleq 3 - k$*, and as shown in [54], we have*

$$C_{2\times 1} = \mathbb{E} \log_2 \left[1 + \frac{P}{2} \mathbf{g}^\dagger \mathbf{g}\right], \tag{3.9}$$

where $\mathbf{g}$ *is a* 2 *by* 1 *vector where entries are i.i.d.* $\mathcal{CN}(0, 1)$.

Remark 3.1 *As we will show in Section 3.5,* $\mathcal{C}_k$*,* $k = 1, 2$*, is an outer-bound on the capacity region of a two-user complex Gaussian MIMO BC with no CSIT where* Rx_k *has two antennas and* $\mathsf{Rx}_{\bar{k}}$ *has only one antenna (additive noise processes all having zero-mean and variance* 1*). The corner points of* $\mathcal{C}_1$ *are given by* $(0, C_{2\times 1})$ *and* $(2C_{2\times 1}, 0)$*.*

The following theorem states our contribution for the two-user MISO BC with delayed CSIT.

Theorem 3.1 *The capacity region of the two-user MISO BC with delayed CSIT,* $\mathcal{C}$*, is within* 1.81 *bits/sec/Hz per user of* $(\mathcal{C}_1 \cap \mathcal{C}_2)$, i.e.

$$(\mathcal{C}_1 \cap \mathcal{C}_2) \ominus (1.81, 1.81) \subseteq \mathcal{C} \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2), \tag{3.10}$$

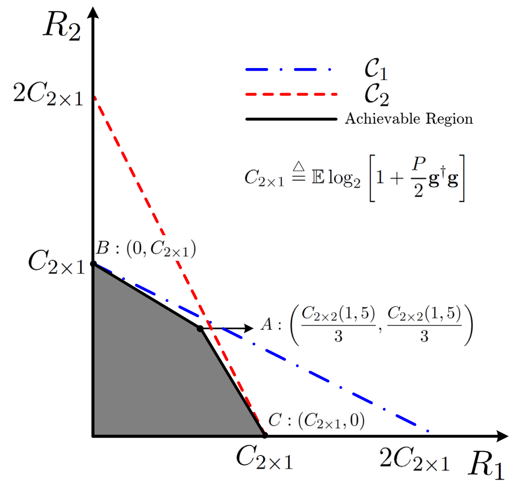


Figure 3.3: The outer-bound on the capacity region of the two-user MISO BC with delayed CSIT is the intersection of $\mathcal{C}_1$ and $\mathcal{C}_2$. We prove that the capacity region is within 1.81 bit/sec/Hz per user of this outer-bound. The achievable rate region is shown by the shaded area. $C_{2\times1}$ is defined in (3.9) and $C_{2\times2}(1,5)$ is given in Definition 3.2.

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are given in Definition 3.3.

Remark 3.2 *Fig. 3.3 pictorially depicts our result for the two-user MISO BC with delayed CSIT. We have defined $C_{2\times1}$ and $C_{2\times2}(1,5)$ in Definition 3.2. We analytically show that the achievable region is within 1.81 bits/sec/Hz per user of the outer-bound. Our numerical analysis shows that the gap is in fact smaller and in the worst case, it is at most 1.1 bits/sec/Hz per user.*

We then consider the K-user MISO BC with delayed CSIT. We only focus on the symmetric capacity defined as follows.

Definition 3.4 *The symmetric capacity of the K-user MISO BC with delayed CSIT, C^K_{SYM}, is given by*

$$C^K_{\text{SYM}} \triangleq \sup\left\{R :\ R \leq R_j, \forall j = 1,2,\ldots,K,\ (R_1, R_2, \ldots, R_K) \in \mathcal{C}^K\right\}, \tag{3.11}$$

where C^K is the capacity region of the K-user MISO BC with delayed CSIT.

The following theorem states our contribution for the K-user setting.

Theorem 3.2 *For the K-user MISO BC with delayed CSIT, C^K_{SYM} is lower-bounded by*

$$\frac{C_{K\times K}\left(1,5,\ldots,1+(j-1)(j+2),\ldots,1+(K-1)(K+2)\right)}{K\sum_{j=1}^{K} j^{-1}}, \tag{3.12}$$

and upper-bounded by $\left(\sum_{j=1}^{K} j^{-1}\right)^{-1} C_{K\times 1}$.

Remark 3.3 *For Theorem 3.2, our numerical analysis shows that the gap is at most* 2.3 *bits/sec/Hz per user for $K \leq 20$ and $P \leq 60$dB. In general, we show that the gap is at most $2\log_2(K+2)$ bits/sec/Hz.*

3.4 Achievability Proof of Theorem 3.1

In this section, we describe the achievability strategy of Theorem 3.1. To characterize the capacity region of the two-user MISO complex Gaussian BC to within 1.81 bits/sec/Hz per user, we need to show that the $(\mathcal{C}_1 \cap \mathcal{C}_2) \ominus (1.81, 1.81)$ is achievable.

We define rate region $\mathcal{R}(D)$ for parameter $D\gamma4$ as follows[1]

$$\mathcal{R}(D) \triangleq \left\{ (R_1, R_2) \left| \begin{array}{l} 0 \leq R_1 + \left(\frac{3C_{2\times 1}}{C_{2\times 2}(1,1+D)} - 1\right) R_2 \leq C_{2\times 1} \\ 0 \leq \left(\frac{3C_{2\times 1}}{C_{2\times 2}(1,1+D)} - 1\right) R_1 + R_2 \leq C_{2\times 1} \end{array} \right. \right\}. \tag{3.13}$$

[1]As we will see later in this section and Section 3.6, we have $\left(\frac{3C_{2\times 1}}{C_{2\times 2}(1,1+D)} - 1\right) > 0$ for $D\gamma4$ and $P > 0$dB.

Later in Section 3.6, we show that $(C_1 \cap C_2) \ominus (1.81, 1.81) \subseteq \mathcal{R}(4)$, and thus, to characterize the capacity region to within 1.81 bit/sec/Hz per user, it suffices to show that $\mathcal{R}(4)$ is achievable.

The shadowed region in Fig. 3.3 corresponds to $\mathcal{R}(4)$, which is a polygon with corner points A, B, and B'. Corner points B and B' of $\mathcal{R}(4)$, are achievable using the result on point-to-point MISO Gaussian channels with no CSIT [54]. Therefore, we only need to describe the achievability strategy for corner point A.

3.4.1 Transmission Strategy for Corner Point A

Here, we present the achievability strategy for corner point A. We show that for any $\epsilon > 0$, we can achieve

$$(R_1, R_2) = \left(\frac{C_{2\times 2}(1,5) - \epsilon}{3}, \frac{C_{2\times 2}(1,5) - \epsilon}{3} \right), \tag{3.14}$$

with vanishing decoding error probability as the communication length goes to infinity.

The achievability strategy is carried on over n blocks, each block consisting of 3 phases each of length n channel uses. Denote by w_k^b, the message of user k in block b, $k \in \{1,2\}$, $b = 1, 2, \ldots, n$. Fix $\epsilon > 0$ and set $R = C_{2\times 2}(1, 1 + D) - \epsilon$. We assume that $w_k^b \in \{1, 2, \ldots, 2^{nR}\}$ and that the messages are distributed uniformly and independently. The encoding is carried on as described below.

• **Encoding:** At the transmitter, the message of user k during block b, *i.e.* w_k^b, is mapped to a codeword of length n denoted by $\mathbf{x}_k^{b,n}$ where any element of this codeword is drawn i.i.d. from $\mathcal{CN}(0, P/2\mathbf{I}_2)$[2].

• Communication during Phase j of block b: During this phase, the transmitter communicates $\mathbf{x}_j^{b,n}$ from its two transmit antennas, $j = 1, 2$, and $b = 1, 2, \ldots, n$. Receiver one obtains $y_{1,j}^{b,n}$ and receiver two obtains $y_{2,j}^{b,n}$.

• Communication during Phase 3 of block b: Using the delayed CSIT, the transmitter creates

$$s^{b,n} = s_{2,1}^{b,n} + s_{1,2}^{b,n}, \tag{3.15}$$

where $s_{2,1}^{b,n}$ is the received signal at Rx_2 during Phase 1 of block b, $y_{2,1}^{b,n}$, minus the noise term as defined in (3.2), and $s_{1,2}^{b,n}$ is the received signal at Rx_1 during Phase 2 of block b, $y_{1,2}^{b,n}$, minus the noise term.

Note that $s_{2,1}^{b,n} + s_{1,2}^{b,n}$ is useful for both receivers since each receiver can subtract its previously received signal to obtain what the other receiver has (up to the noise term). Therefore, the goal in this phase, is to provide $s_{2,1}^{b,n} + s_{1,2}^{b,n}$ to both receivers with distortion $D = 4$.

Remark 3.4 *We note that the idea of creating quantized version of previously received signals has been utilized in prior works for asymptotic degrees of freedom analysis. However for finite SNR regime, we need to take into account the fact that the quantization noise is neither independently distributed over time nor is it independent from the signal; and we need to overcome these challenges.*

The input signal to a lattice quantizer needs to be independently distributed over time (see [76, 77]). Thus, in order to quantize this signal using lattice quan-

[2] $\mathbf{I}_2$ is the 2×2 identity matrix.

tizer, we need

$$\left(s^b_{2,1}[\ell] + s^b_{1,2}[\ell]\right)^n_{\ell=1}$$

to be an independently distributed sequence. However, given message w^b_k, the transmit signal $\mathbf{x}^{b,n}_k$ is correlated across time and as a result, the aforementioned signals are not independent anymore. In order to overcome this issue, we incorporate an interleaving step according to the following mapping which is depicted in Fig. 3.4.

$$\tilde{s}^b[t] = s^t_{2,1}[b] + s^t_{1,2}[b], \tag{3.16}$$

where $b = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, n$. It is important to notice that with this interleaving, the resulting signal at different time instants of a given phase at a given block are independent from each other. This is due to the fact that these signals are created from independent messages.

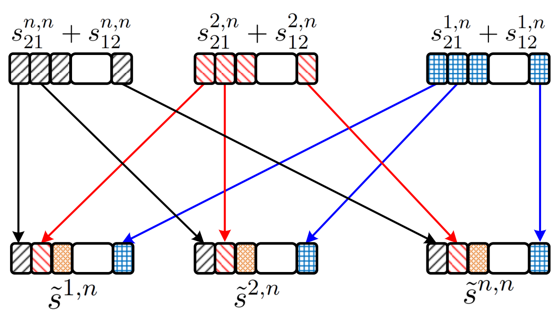


Figure 3.4: The interleaving step: the resulting signal at different time instants of a given phase at a given block are independent from each other.

Note that given the previous channel realizations, the signal in (3.16) at each time has a Gaussian distribution but its variance varies from each time instant to the other. Thus, in order to be able to quantize it, we need a generalization of the rate-distortion function to include non-identically distributed sources. Below, we discuss this issue.

Lemma 3.1 *[Rate distortion for non-identically distributed Gaussian Source] Consider a independently non-identically distributed Gaussian source* $\mathbf{u}$*, where at time instant* t*, it has zero mean and variance* $\sigma^2[t]$*. Assume that* $\sigma^2[t]$ *is drawn from some i.i.d. continuous distribution with* $\mathbb{E}\left[\sigma^2\right] < \infty$*. The sequence of* $\sigma^2[t]$ *is non-causally known by both encoder and decoder. Then, with squared-error distortion, we can quantize the signal at any rate greater than or equal to*

$$\min_{D_\sigma:\mathbb{E}[D_\sigma]\leq D} \mathbb{E}\left[\log_2 \frac{\sigma^2[t]}{D_\sigma}\right]^+, \tag{3.17}$$

and achieve distortion D *(per sample), where the expectation is with respect to the distribution of* σ^2*.*

Proof sketch: Suppose σ^2 could only take m finite values σ_i^2 with probability p_i, $i = 1, 2, \ldots, m$. The problem would then be similar to that of m parallel Gaussian channels where waterfilling gives the optimal solution ((Theorem 10.3.3 [15])). For the time instants where source has variance σ_i^2, we choose a distortion D_i such that $\sum_{i=1}^m p_i D_i \leq D$. Note that in order to derive the optimum answer, we need to optimize over the choice of D_i's. The case where σ^2 take values in a continuous set can be viewed as the limit of the discrete scenario using standard arguments. ■

It is easy to see that any rate greater than or equal to

$$\mathbb{E}\left[\log_2\left(1 + \frac{\sigma^2[t]}{D}\right)\right], \tag{3.18}$$

is also achievable at distortion D (per sample). Basically, we have ignored the optimization over D and added a 1 to remove $\max\{., 0\}$ (or $.^+$).

Moreover, we would like the distortion to be independent of the signals. In order to have a distortion that is independent of the signal and is uncorrelated across time, we can incorporate lattice quantization with "dither" as described in [76]. Basically, dither is a random vector distributed uniformly over the basic Voronoi that is added to the signal before feeding it to the quantizer. At the decoder, this random vector will be subtracted.

From (3.18), we conclude that we can quantize $\tilde{s}^{b,n}$ with squared-error distortion D at rate $R_Q(D)$ (per sample), defined as

$$R_Q(D) \triangleq \mathbb{E}\left[\log_2\left(1 + \frac{P}{2D}\left(\|\mathbf{g}\|_2^2 + \|\mathbf{h}\|_2^2\right)\right)\right], \tag{3.19}$$

where $\mathbf{g}, \mathbf{h} \in \mathbb{C}^{2\times 1}$ with i.i.d. $\mathcal{CN}(0,1)$ entries.

We can reliably communicate the quantization index over

$$\left\lceil \frac{nR_Q(4)}{C_{2\times 1}} \right\rceil \tag{3.20}$$

time instants. Next, we need to show that given the appropriate choice of parameters, receivers can recover the corresponding messages with vanishing error probability as $n \to \infty$.

3.4.2 Decoding

Upon completion of Phase 3 of block b, each receiver decodes the quantized signal. We know that as $n \to \infty$, this could be done with arbitrary small decoding error probability. Therefore, each receiver has access to

$$\tilde{s}^{b,n} + z_Q^{b,n}, \tag{3.21}$$

where $z_Q^{b,n}$ is the quantization noise with variance D which is independent of the transmit signals. Note that $z_Q^b[t_1]$ and $z_Q^b[t_2]$ are uncorrelated but not necessarily independent, $t_1, t_2 = 1, 2, \ldots, n$, $t_1 \neq t_2$.

Receiver 1 at the end of the n^{th} communication block, reconstructs signals by reversing the interleaving procedure described above, and removes $y_{1,2}^{b,n}$ to obtain

$$\tilde{y}_{2,1}^{b,n} = y_{2,1}^{b,n} + \tilde{z}_{Q}^{b,n}, \tag{3.22}$$

here $\tilde{z}_{Q}^{b,n}$ is the quantization noise with variance D which is independent of the transmit signals. Moreover, $\tilde{z}_{Q}^{b}[\ell]_{\ell=1}^{n}$ is an *independent* sequence.

Note that since the messages are encoded at rate $C_{2\times 2}(1, 1+D) - \epsilon$ for $\epsilon > 0$, if receiver one has access to $y_{2,1}^{b,n}$ up to distortion D, it can recover w_1^b with arbitrary small decoding error probability as $\epsilon \to 0$ and communication length goes to infinity. Thus, from $y_{1,1}^{b,n}$ and $\tilde{y}_{2,1}^{b,n}$, receiver one can decode w_1^b, $b = 1, 2, \ldots, n$. Similar argument holds for receiver two.

An error may occur in either of the following steps: (1) if an error occurs in decoding message w_k^b provided required signals to the receiver, $k = 1, 2$; (2) if an error occurs in quantizing $\tilde{s}^{b,n}$; and (3) if an error occurs in decoding $\tilde{s}^{b,n} + z_Q^{b,n}$ at either of the receivers, $b = 1, 2, \ldots, n$. The probability of each one of such errors decreases exponentially in n (see [24, 54] and references therein). Using union bound and given that we have $O\left(n^2\right)$ possible errors and the fact that each error probability decreases exponentially to zero, the total error probability goes to zero as $n \to \infty$.

3.4.3 Achievable Rate

Using the achievable strategy described above, as $n \to \infty$, we can achieve a (symmetric) sum-rate point of

$$(R_1, R_2) = \left(\frac{C_{2\times2}(1,5)}{2 + R_Q(4)/C_{2\times1}}, \frac{C_{2\times2}(1,5)}{2 + R_Q(4)/C_{2\times1}} \right). \tag{3.23}$$

In Appendix B.2, we show that $R_Q(4)/C_{2\times1} \leq 1$ for all values of P. Therefore, Phase 3 of each block has at most n time instants. Thus, we conclude that a (symmetric) sum-rate point of

$$(R_1, R_2) = \left(\frac{C_{2\times2}(1,5)}{3}, \frac{C_{2\times2}(1,5)}{3} \right), \tag{3.24}$$

is achievable.

3.5 Converse Proof of Theorem 3.1

In this section, we provide the converse proof of Theorem 3.1. The converse consists of two main parts. In part 1, we show that the capacity region of the problem is included in the capacity region of a (stochastically) degraded BC, and in part 2, we derive an outer-bound on the capacity region of the degraded BC.

Part 1: We create the stochastically degraded BC as follows. We first provide the received signal of Rx_2, *i.e.* $y_2[t]$, to Rx_1 as depicted in Fig. 3.5. Note that the resulting channel is *physically degraded*, and we know that for a physically degraded broadcast channel, feedback does not enlarge the capacity region [22]. Therefore, we can ignore the delayed knowledge of the channel state information at the transmitter (*i.e.* no CSIT assumption). The resulting channel is stochastically degraded.

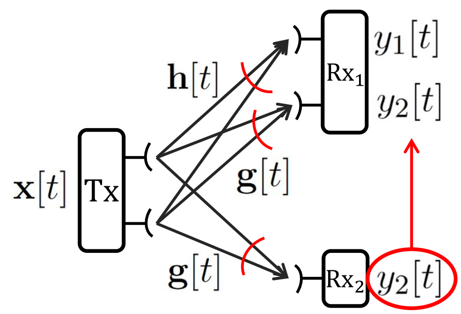


Figure 3.5: By providing y_2^n to Rx_1, we create a physically degraded BC. We then ignore the delayed knowledge of the channel state information at the transmitter because for a physically degraded broadcast channel, feedback does not enlarge the capacity region. The resulting channel is stochastically degraded.

Thus, we form a *stochastically degraded* channel as shown in Fig. 3.5, and we formally define it in Definition 3.5 below. Let us denote an outer-bound on the capacity region of this channel by $\mathcal{C}_1$. The argument above shows that $\mathcal{C} \subseteq \mathcal{C}_1$.

Definition 3.5 *The input-output relationship of a two-user stochastically degraded MIMO BC as depicted in Fig. 3.5, is given by*

$$\mathbf{y}_1[t] = \mathbf{H}[t]\mathbf{x}[t] + \mathbf{z}_1[t], \qquad y_2[t] = \mathbf{g}^\top[t]\mathbf{x}[t] + z_2[t], \tag{3.25}$$

for

$$\mathbf{H}[t] = \begin{bmatrix} h_1[t] & h_2[t] \\ g_1[t] & g_2[t] \end{bmatrix}, \qquad \mathbf{g}[t] = \begin{bmatrix} g_1[t] \\ g_2[t] \end{bmatrix}, \tag{3.26}$$

where $h_j[t]$ and $g_j[t]$ are distributed as i.i.d. (independent over time and from each other) $\mathcal{CN}(0,1)$ as described in Section 3.2 for $j = 1, 2$. We have

$$\mathbf{y}_1[t] = \begin{bmatrix} y_{11}[t] \\ y_{12}[t] \end{bmatrix}. \tag{3.27}$$

We assume $\mathbf{z}_1[t] \in \mathbb{C}^{2\times 1}$ and $z_2[t] \sim \mathcal{CN}(0,1)$, and the transmit signal is subject to average power constraint P. We further assume that the transmitter has no knowledge of the channel gains besides their distributions (no CSIT assumption).

Part 2: In this part, we derive an outer-bound, C_1, on the capacity region of the stochastically degraded broadcast channel. For the stochastically degraded BC defined in Definition 3.5, suppose there exists encoders and decoders at the transmitter and receivers such that each message can be decoded at its corresponding receiver with arbitrary small decoding error probability.

$$\begin{aligned}
n\left(R_1+2R_2-3\epsilon_n\right) &\overset{\text{Fano}}{\leq} I\left(w_1; y_{11}^n, y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) + 2I\left(w_2; y_2^n | \mathbf{H}^n, \mathbf{g}^n\right) \\
&\overset{(a)}{=} I\left(w_1; y_{11}^n, y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) + I\left(w_2; y_{11}^n | \mathbf{H}^n, \mathbf{g}^n\right) + I\left(w_2; y_{12}^n | \mathbf{H}^n, \mathbf{g}^n\right) \\
&= h\left(y_{11}^n, y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) - h\left(y_{11}^n, y_{12}^n | w_1, w_2, \mathbf{H}^n, \mathbf{g}^n\right) \\
&\quad + h\left(y_{11}^n | \mathbf{H}^n, \mathbf{g}^n\right) - h\left(y_{11}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) + h\left(y_{12}^n | \mathbf{H}^n, \mathbf{g}^n\right) - h\left(y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) \\
&\overset{(b)}{=} h\left(y_{11}^n | \mathbf{H}^n, \mathbf{g}^n\right) - h\left(z_{11}^n | \mathbf{H}^n, \mathbf{g}^n\right) + h\left(y_{12}^n | \mathbf{H}^n, \mathbf{g}^n\right) - h\left(z_{12}^n | \mathbf{H}^n, \mathbf{g}^n\right) \\
&\quad + h\left(y_{11}^n, y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) - h\left(y_{11}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) - h\left(y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) \\
&\overset{(c)}{\leq} 2\mathbb{E}\log_2\left[1+\frac{P}{2}\mathbf{g}^\dagger\mathbf{g}\right] - I\left(y_{11}^n; y_{12}^n | w_2, \mathbf{H}^n, \mathbf{g}^n\right) \overset{(d)}{\leq} 2\mathbb{E}\log_2\left[1+\frac{P}{2}\mathbf{g}^\dagger\mathbf{g}\right] \overset{(3.9)}{=} 2C_{2\times 1},
\end{aligned} \tag{3.28}$$

where (a) follows from the fact that due to no CSIT assumption, we have

$$h\left(w_2 | y_{11}^n, \mathbf{H}^n, \mathbf{g}^n\right) \leq n\epsilon_n, \qquad h\left(w_2 | y_{12}^n, \mathbf{H}^n, \mathbf{g}^n\right) \leq n\epsilon_n; \tag{3.29}$$

(b) holds since

$$\begin{aligned}
h\left(\mathbf{y}_1^n | w_1, w_2, \mathbf{H}^n, \mathbf{g}^n\right) &= h\left(\mathbf{y}_1^n | w_1, w_2, \mathbf{x}^n, \mathbf{H}^n, \mathbf{g}^n\right) \\
&= h\left(z_{11}^n, z_{12}^n | w_1, w_2, \mathbf{x}^n, \mathbf{H}^n, \mathbf{g}^n\right) = h\left(z_{11}^n | \mathbf{H}^n, \mathbf{g}^n\right) + h\left(z_{12}^n | \mathbf{H}^n, \mathbf{g}^n\right);
\end{aligned} \tag{3.30}$$

(c) follows from the results in [54]; and (d) follows from fact that mutual information is always positive. Dividing both sides by n and letting $n \to \infty$, we obtain the desired result. This completes the derivation of $\mathcal{C}_1$. Similarly, we can derive $\mathcal{C}_2$, and we have $\mathcal{C} \subseteq \mathcal{C}_1 \cap \mathcal{C}_2$ which completes the converse proof for Theorem 3.1.

3.6 Gap Analysis

In this section, we evaluate the gap between our achievable rate-region and the outer-bound. We analytically prove that the gap is at most 1.81 bits/sec/Hz per user. We show that

$$(\mathcal{C}_1 \cap \mathcal{C}_2) \ominus (1.81, 1.81) \subseteq \mathcal{R}(4). \tag{3.31}$$

Since the achievable rate region and the outer-bound (see Fig. 3.3) are formed by time sharing among the corresponding corner points (and thus, characterized by straight lines), we only need to consider the symmetric capacity, C^2_{SYM}, defined in Definition 3.4.

We evaluate the gap between the inner-bound in (3.24), *i.e.*

$$(R_1, R_2) = \left(\frac{C_{2\times 2}(1,5)}{3}, \frac{C_{2\times 2}(1,5)}{3} \right). \tag{3.32}$$

and the symmetric point $\left(C^2_{\mathrm{SYM}}, C^2_{\mathrm{SYM}}\right)$, obtained from Theorem 3.1.

A numerical evaluation of the gap between the sum-rate inner-bound and outer-bound is plotted in Fig. 3.6.

To analyze the gap between the two bounds, we first study the gap between $C_{2\times 2}(1,5)$ and $2C_{2\times 1}$.

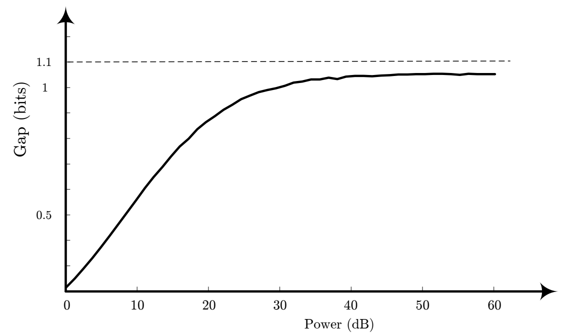


Figure 3.6: Numerical evaluation of the per-user gap between the sum-rate inner-bound and outer-bound.

Corollary 3.3 *Consider a MIMO point-to-point channel with* 2 *transmit antennas and* 2 *receive antennas as described in [54]. The only difference is that the additive noise at one antenna has variance* 1 *while the additive noise at the other antenna has variance* $(1 + D)$*. The ergodic capacity of this channel,* $C_{2\times2}(1, 1 + D)$*, satisfies*

$$C_{2\times2}(1, 1 + D) \geq \left(\mathbb{E} \log_2 \det\left[\mathbf{I}_2 + \frac{P}{2}\mathbf{H}\mathbf{H}^\dagger\right] - \log_2(1 + D)\right)^+ . \tag{3.33}$$

Proof:

$$\begin{aligned}
& C_{2\times2}(1, 1 + D) \\
& \overset{(a)}{\geq} \mathbb{E} \log_2 \det\left[\mathbf{I}_2 + \frac{P}{2}\begin{bmatrix} h_{11} & h_{12} \\ h_{21}/\sqrt{1+D} & h_{22}/\sqrt{1+D} \end{bmatrix}\begin{bmatrix} h_{11}^\dagger & h_{21}^\dagger/\sqrt{1+D} \\ h_{12}^\dagger & h_{22}^\dagger/\sqrt{1+D} \end{bmatrix}\right] \\
& = \mathbb{E} \log_2 \det\left[\mathbf{I}_2 + \frac{P}{2}\begin{bmatrix} |h_{11}|^2 + |h_{12}|^2 & \left(h_{11}h_{21}^\dagger + h_{12}h_{22}^\dagger\right)/\sqrt{1+D} \\ \left(h_{11}^\dagger h_{21} + h_{12}^\dagger h_{22}\right)/\sqrt{1+D} & \left(|h_{21}|^2 + |h_{22}|^2\right)/(1+D) \end{bmatrix}\right] \\
& = \mathbb{E} \log_2 \det\begin{bmatrix} 1 + \frac{P}{2}|h_{11}|^2 + |h_{12}|^2 & \frac{P}{2}\left(h_{11}h_{21}^\dagger + h_{12}h_{22}^\dagger\right)/\sqrt{1+D} \\ \frac{P}{2}\left(h_{11}^\dagger h_{21} + h_{12}^\dagger h_{22}\right)/\sqrt{1+D} & 1 + \frac{P}{2}\left(|h_{21}|^2 + |h_{22}|^2\right)/(1+D) \end{bmatrix} \\
& \geq \mathbb{E} \log_2 \det\begin{bmatrix} 1 + \frac{P}{2}|h_{11}|^2 + |h_{12}|^2 & \frac{P}{2}\left(h_{11}h_{21}^\dagger + h_{12}h_{22}^\dagger\right)/\sqrt{1+D} \\ \frac{P}{2}\left(h_{11}^\dagger h_{21} + h_{12}^\dagger h_{22}\right)/\sqrt{1+D} & \left(1 + \frac{P}{2}\left(|h_{21}|^2 + |h_{22}|^2\right)\right)/(1+D) \end{bmatrix}
\end{aligned}$$

$$= \underbrace{\mathbb{E}\log_2\det\left[\mathbf{I}_2 + \frac{P}{2}\mathbf{H}\mathbf{H}^\dagger\right]}_{\triangleq C_{2\times 2}} - \log_2(1+D), \tag{3.34}$$

where (a) holds since the right hand side is obtained by evaluating the mutual information between the input and output, for a complex Gaussian input with covariance matrix $\mathbb{E}\left[\mathbf{x}^\dagger\mathbf{x}\right] = P/2\mathbf{I}_2$. ■

Therefore, the gap between the sum-rate inner-bound and outer-bound can be upper-bounded by

$$\frac{4C_{2\times 1}}{3} - \frac{2C_{2\times 2}(1, 1+D)}{3} \leq \frac{2\left(2C_{2\times 1} - C_{2\times 2} + \log_2(1+D)\right)}{3}, \tag{3.35}$$

where

$$C_{2\times 2} \triangleq \mathbb{E}\log_2\det\left[\mathbf{I}_2 + \frac{P}{2}\mathbf{H}\mathbf{H}^\dagger\right]. \tag{3.36}$$

Remark 3.5 *While in this work we evaluate the gap for Rayleigh fading channels, our expressions for the inner-bounds and the outer-bounds hold for general i.i.d. channel realizations. The challenege to evaluate the gap for distributions other than Rayleigh fading, asises from determining the optimal covariance matrix for the transmit signal and evaluating the capacity result as discussed in [54] Section 4.1.*

For $P \leq 2$, the sum-rate outer-bound is smaller than 2 bits (smaller than the gap itself). So, we assume $P > 2$. We have

$$\begin{aligned}
2C_{2\times 1} - C_{2\times 2} &= 2\mathbb{E}\log_2\left[1 + \frac{P}{2}\mathbf{g}^\dagger\mathbf{g}\right] - \mathbb{E}\log_2\det\left[\mathbf{I}_2 + \frac{P}{2}\mathbf{H}\mathbf{H}^\dagger\right] \\
&= 2\mathbb{E}\log_2\left[\frac{2}{P} + \mathbf{g}^\dagger\mathbf{g}\right] + 2\log_2\left(\frac{P}{2}\right) - \mathbb{E}\log_2\det\left[\frac{2}{P}\mathbf{I}_2 + \mathbf{H}\mathbf{H}^\dagger\right] - \log_2\left(\frac{P^2}{4}\right) \\
&= 2\mathbb{E}\log_2\left[\frac{2}{P} + \mathbf{g}^\dagger\mathbf{g}\right] - \mathbb{E}\log_2\det\left[\frac{2}{P}\mathbf{I}_2 + \mathbf{H}\mathbf{H}^\dagger\right] \\
&\leq 2\mathbb{E}\log_2\left[1 + \mathbf{g}^\dagger\mathbf{g}\right] - \mathbb{E}\log_2\det\left[\mathbf{H}\mathbf{H}^\dagger\right].
\end{aligned} \tag{3.37}$$

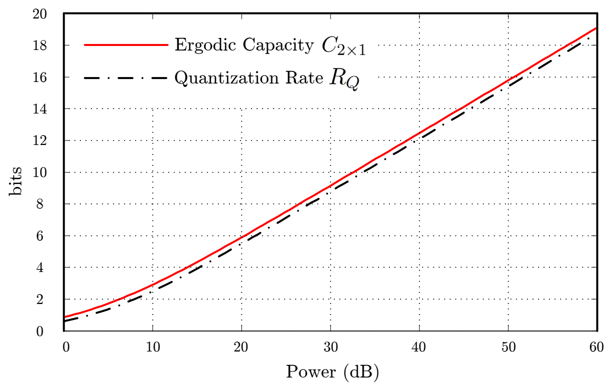


Figure 3.7: Using numerical analysis, we can show that $R_Q(3)/C_{2\times1} \leq 1$.

Thus, we have

$$\begin{aligned}&\frac{4C_{2\times1}}{3} - \frac{2C_{2\times2}(1, 1 + D)}{3} \\ &\qquad \leq \frac{2}{3}\left(2\mathbb{E}\log_2\left[1 + \mathbf{g}^\dagger\mathbf{g}\right] - \mathbb{E}\log_2\det\left[\mathbf{H}\mathbf{H}^\dagger\right] + \log_2(1 + D)\right) \leq 3.62,\end{aligned} \tag{3.38}$$

which implies that the gap is less than 1.81 bits per user independent of power P. We could also use numerical analysis to evaluate the gap. In particular, using numerical analysis, we can show that $R_Q(3)/C_{2\times1} \leq 1$ (see Fig. 3.7). We have plotted

$$\frac{4C_{2\times1}}{3} - \frac{2C_{2\times2}(1, 1 + D)}{3} \tag{3.39}$$

in Fig. 3.6 for $D = 4$, and for P between 0 dB and 60 dB. As we can see from Fig. 3.6, the sum-rate inner-bound and outer-bound are at most 2.2 bits (or 1.1 per user) away from each other for P between 0 dB and 60 dB.

3.7 Extension to K-user MISO BC

Now that we have presented our results for the two-user multiple-input single-output complex Gaussian broadcast channel with delayed CSIT, we consider the

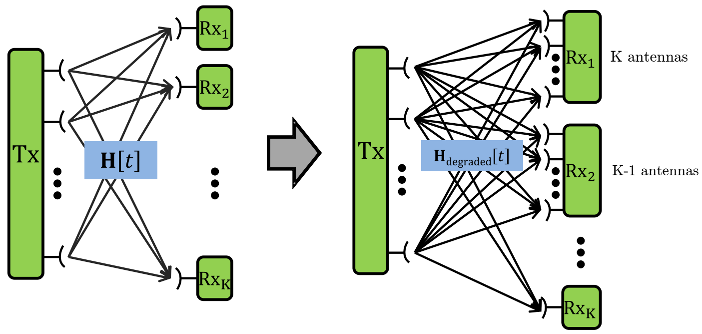


Figure 3.8: Left: K-user Multiple-Input Single-Output (MISO) Complex Gaussian Broadcast Channel; and right: degraded MIMO BC.

K-user setting as depicted in Fig. 3.8. The channel matrix from the transmitter to the receivers is denoted by $\mathbf{H} \in \mathbb{C}^{K \times K}$, where the entries of $\mathbf{H}[t]$ are distributed as i.i.d. $\mathcal{CN}(0, 1)$ (independent across time, antenna, and users). The transmit signal $\mathbf{x}[t] \in \mathbb{C}^{K \times 1}$ is subject to average power constraint P, *i.e.* $\mathbb{E}\left[\mathbf{x}^{\dagger}[t]\mathbf{x}[t]\right] \leq P$ for $P > 0$. The noise processes are independent from the transmit signal and are distributed i.i.d. as $z_k[t] \sim \mathcal{CN}(0, 1)$. The input-output relationship in this channel is given by

$$\begin{bmatrix} y_1[t] \\ y_2[t] \\ \vdots \\ y_K[t] \end{bmatrix} = \mathbf{H}[t]\mathbf{x}[t] + \begin{bmatrix} z_1[t] \\ z_2[t] \\ \vdots \\ z_K[t] \end{bmatrix}. \tag{3.40}$$

3.7.1 Outer-bound

The derivation of the outer-bound in Theorem 3.2 is based on creating a degraded MIMO BC where Rx_ℓ has access to the received signal of Rx_j for $j \gamma \ell$, $\ell = 1, 2, \ldots, K$. This channel is physically degraded, and thus feedback does not enlarge the capacity region [22]. Therefore, we can ignore the delayed knowledge of the channel state information at the transmitter (*i.e.* no CSIT assumption). See Fig. 3.8 for a depiction. For the resulting multiple-input multiple-output (MIMO) broadcast channel denote the channel matrix by $\mathbf{H}^n_{\text{degraded}}$. Then, under no CSIT assumption, the following result is known for the MIMO BC.

Lemma 3.2 ([69])

$$h\left(\mathbf{y}_j^n|w_{j+1},\ldots,w_K,\mathbf{H}^n\right) \geq \frac{K-j+1}{K-j} h\left(\mathbf{y}_{j+1}^n|w_{j+1},\ldots,w_K,\mathbf{H}^n\right), \qquad j=1,\ldots,K-1. \tag{3.41}$$

Using Lemma 3.2, similar to the argument presented in (3.28), we get

$$\begin{aligned}
& n\left(R_1 + \frac{K}{K-1}R_2 + \frac{K}{K-2}R_2 + \ldots + KR_K - \epsilon_n\right) \\
& = I\left(w_1;\mathbf{y}_1^n|w_2,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) + \frac{K}{K-1} I\left(w_2;\mathbf{y}_2^n|w_3,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) \\
& + \ldots + KI\left(w_K; y_K^n|\mathbf{H}^n_{\text{degraded}}\right) \\
& \leq h\left(\mathbf{y}_1^n|w_2,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) - \frac{K}{K-1} h\left(\mathbf{y}_2^n|w_2,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) \\
& + \frac{K}{K-1} h\left(\mathbf{y}_2^n|w_3,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) - \frac{K}{K-2} h\left(\mathbf{y}_3^n|w_3,\ldots,w_K,\mathbf{H}^n_{\text{degraded}}\right) \\
& + \ldots + Kh\left(y_K^n|\mathbf{H}^n_{\text{degraded}}\right) \\
& \overset{Lemma\ 3.2}{\leq} Kh\left(y_K^n|\mathbf{H}^n_{\text{degraded}}\right) \leq K\mathbb{E}\log_2\left[1 + \frac{P}{K}\mathbf{g}^\dagger\mathbf{g}\right] \overset{(3.9)}{=} KC_{K\times 1},
\end{aligned} \tag{3.42}$$

where $\mathbf{g}$ is a K by 1 vector where entries are i.i.d. $\mathcal{CN}(0,1)$. Then, for the symmetric sum-rate point we have $R_1 = R_2 = \ldots = R_K = R$, which gives us

$$\left(\sum_{j=1}^{K} j^{-1}\right) R \leq KC_{K\times 1}, \tag{3.43}$$

which implies that

$$C_{\text{SYM}}^K \leq \frac{C_{K\times 1}}{\sum_{j=1}^K j^{-1}}. \tag{3.44}$$

This completes the converse proof of Theorem 3.2.

3.7.2 Achievability

In this subsection, we focus on achievability and discuss the extension of our achievability results to the K-user MISO BC with delayed CSIT. The transmission strategy follows the steps of [36], but similar to the two-user case, some additional ingredients are needed to make an approximately optimal scheme. We first demonstrate the techniques and the required ingredients for the three-user MISO BC with delayed CSIT. Then, we discuss how the result is extended to the general case.

Transmission strategy for the three-user MISO BC: Each communication block includes eight phases and we focus on a specific block b. Fix $\epsilon > 0$.

• During Phases 1, 2, and 3 of block b, the message of user k (w_k^b) is encoded as $\mathbf{x}_k^{b,n}$ at rate

$$C_{3\times 3}\left(1,5,11\right) - \epsilon, \tag{3.45}$$

where any element of this codeword is drawn i.i.d. from $\mathcal{CN}\left(0, P/3\mathbf{I}_3\right)^3$; and $C_{3\times 3}\left(1,5,11\right)$ is given in Definition 3.2.

Remark 3.6 *The reason for the encoding rate of (3.45) becomes apparent as we describe the achievability strategy. Basically the quantization noise accumulates on top of the previous noises throughout some of the phases and thus, we need to adjust the encoding rate accordingly.*

Definition 3.6 *We define $s_{k,j}^{b,n}$ to be the noiseless version of $y_{k,j}^{b,n}$ for appropriate choice of indices. This is similar to Definition 3.2 for the two-user case.*

[3] $\mathbf{I}_3$ is the 3×3 identity matrix.

Transmitter communicates these codewords and receiver one obtains $y_{1,j}^{b,n}$, receiver two obtains $y_{2,j}^{b,n}$, and receiver three obtains $y_{3,j}^{b,n}$, $j = 1, 2, 3$. Note that at the end of the third phase, transmitter has access to $s_{1,j}^{b,n}$, $s_{2,j}^{b,n}$, and $s_{3,j}^{b,n}$. Due to the rate given in (3.45), if $s_{2,1}^{b,n}$ and $s_{3,1}^{b,n}$ are provided to receiver one with distortions 4 and 8 respectively, then receiver one will be able to decode its message with arbitrary small decoding error probability. Similarly receiver two is interested in $s_{1,2}^{b,n}$ and $s_{3,2}^{b,n}$; and receiver three is interested in $s_{1,3}^{b,n}$ and $s_{2,3}^{b,n}$.

Based on the discussion above and the available signal at each receiver, we observe that $s_{2,1}^{b,n} + s_{1,2}^{b,n}$ is of common interest to receivers one and two. Similarly, $s_{3,1}^{b,n} + s_{1,3}^{b,n}$ is of common interest to receivers one and three; and $s_{2,3}^{b,n} + s_{3,2}^{b,n}$ is of common interest to receivers two and three. Therefore, the goal would be to deliver these signals to their interested receivers during the following phases.

• Communication during Phase 4: Consider the communication during another block b', and the corresponding signal $s_{2,1}^{b',n} + s_{1,2}^{b',n}$. Using Lemma 3.1, we quantize $s_{2,1}^{b,n} + s_{1,2}^{b,n}$ and $s_{2,1}^{b',n} + s_{1,2}^{b',n}$ at distortion 4, and we create the XOR of the resulting bits[4]. Then these bits will be encoded as $\mathbf{v}_4^{b,n}$ at rate $C_{3\times 2}(1, 5) - \epsilon$, and will be transmitted during Phase 4. $C_{3\times 2}(1, 5)$ is given in Definition 3.2.

• Communication during Phase 5: Consider the communication during another block b', and the corresponding signal $s_{3,1}^{b',n} + s_{1,3}^{b',n}$. Using Lemma 3.1, we quantize $s_{3,1}^{b,n} + s_{1,3}^{b,n}$ and $s_{3,1}^{b',n} + s_{1,3}^{b',n}$ at distortion 4 and we create the XOR of the resulting bits. Then these bits will be encoded at rate $C_{3\times 2}(1, 5) - \epsilon$ denoted by $\mathbf{v}_5^{b,n}$ and will be transmitted during Phase 5. Receiver two obtains $y_{2,5}^{b,n}$ that is of interest of users one and three.

[4]We note that we need these signals to be distributed independently, to handle this issue, we can incorporate an interleaving step similar to Section 3.4.

• Communication during Phase 6: Consider the communication during another block b', and the corresponding signal $s_{2,3}^{b',n} + s_{3,2}^{b',n}$. Using Lemma 3.1, we quantize $s_{2,3}^{b,n} + s_{3,2}^{b,n}$ and $s_{2,3}^{b',n} + s_{3,2}^{b',n}$ at distortion 4 and we create the XOR of the resulting bits. Then these bits will be encoded at rate $C_{3\times 2}(1,5) - \epsilon$ denoted by $\mathbf{v}_6^{b,n}$ and will be transmitted during Phase 5. Receiver one obtains $y_{1,6}^{b,n}$ that is of interest of users two and three.

We now create two signals that are of interest of all three receivers:

$$\frac{1}{\sqrt{2}} s_{3,4}^{b,n} + \frac{1}{\sqrt{3}} s_{2,5}^{b,n} + \frac{1}{\sqrt{6}} s_{1,6}^{b,n}, \qquad \text{and} \qquad \frac{1}{\sqrt{6}} s_{3,4}^{b,n} + \frac{-1}{\sqrt{3}} s_{2,5}^{b,n} + \frac{1}{\sqrt{2}} s_{1,6}^{b,n}. \tag{3.46}$$

Remark 3.7 *The choice of coefficients in (3.46) is such that all users are assigned equal powers and the linear combinations at the receivers remain independent. We also note that using such coefficients results in a* $1/3$ *power loss for each user.*

Note that any receiver has access to both signals in (3.46), then it can recursively recover the signals it is interested in, and decode the intended message.

• Communication during Phase 7: Using Lemma 3.1, we quantize $\left(\frac{1}{\sqrt{2}} s_{3,4}^{b,n} + \frac{1}{\sqrt{3}} s_{2,5}^{b,n} + \frac{1}{\sqrt{6}} s_{1,6}^{b,n}\right)$ at distortion 5. Then these quantized bits will be encoded at rate $C_{3\times 1} - \epsilon$ and transmitted during Phase 7.

• Communication during Phase 8: Using Lemma 3.1, we quantize $\left(\frac{1}{\sqrt{6}} s_{3,4}^{b,n} + \frac{-1}{\sqrt{3}} s_{2,5}^{b,n} + \frac{1}{\sqrt{2}} s_{1,6}^{b,n}\right)$ at distortion 5. Then these quantized bits will be encoded at rate $C_{3\times 1} - \epsilon$ and transmitted during Phase 8.

Using the achievability strategy described above and for $n \to \infty$ and $\epsilon \to 0$, it can be shown that a per-user rate of $\frac{2}{11} C_{3\times 3}(1,5,11)$ is achievable.

Transmission strategy for the K-user MISO BC: Now that we have described the transmission strategy for the 3-user MISO BC with delayed CSIT, we explain the transmission strategy for the general case $K > 3$. As mentioned before, the transmission strategy follows the steps of [36]. However, we highlight the key differences that are needed to derive near optimal results.

The scheme includes K phases. For simplicity, we do not go into details of the interleaving process. At the beginning of Phase j, transmitter has access to signals that are of interest of j receivers, $j = 1, 2, \ldots, K$. There are a total of $(K - j + 1)\binom{K}{j}$ such signals. Phase j has $\binom{K}{j}$ time-slots. Consider a subset $\mathcal{S}$ of the receivers where $|\mathcal{S}| = j$. Transmitter accumulates a total of $(K - j + 1)$ signals that are of interest of all receivers in $\mathcal{S}$ using $(K - j + 1)$ different blocks (this is similar to Phase 4 for the 3-user MISO BC with delayed CSIT). Similar to Lemma 3.1, transmitter quantizes these signals at distortion $(j + 2)$, and creates the XOR of the resulting bits. The resulting bits are encoded at rate $C_{K\times j}(1, 5, \ldots, 1 + (j - 1)(j + 2)) - \epsilon$ and communicated during time slot $t_{\mathcal{S}}$ of Phase j.

Remark 3.8 *The noise variance* $1+(j - 1)(j + 2)$ *results from the fact that each receiver has to solve* $(j - 1)$ *equations of the signals that he is interested in. This step results in boosting up the noise variance.*

Consider any subset $\mathcal{S}'$ of the receivers where $|\mathcal{S}'| = j + 1$. Upon completion of Phase j, we observe that any receiver in $\mathcal{S}'$ has a signal that is simultaneously of common interest of all other receivers in $\mathcal{S}'$. Transmitter has access to this signal (up to noise) using delayed knowledge of the channel state information. Transmitter forms j random linear combinations of such signals for each subset $\mathcal{S}'$ where $|\mathcal{S}'| = j + 1$. Then transmitter creates $j\binom{K}{j+1}$ signals that are simultaneously of interest of $j + 1$ receivers. These signals (after being quantized) will be delivered in Phases $j+1, j+2, \ldots, K$. The rest of the scheme is identical to that of [36] and is omitted due to space limitations.

Recursively solving the achievable rate over K phases, we can show that a per-user rate of

$$\frac{C_{K\times K}\left(1, 5, \ldots, 1+(j-1)(j+2), \ldots, 1+(K-1)(K+2)\right)}{K\sum_{j=1}^{K} j^{-1}} \tag{3.47}$$

is achievable.

3.7.3 Gap Analysis

In Fig. 3.9, we have plotted the numerical analysis of the per user gap between the inner-bound and the outer-bound of Theorem 3.2 for P between 0 dB and 60 dB and $K = 2, 3, 5, 10, 20$.

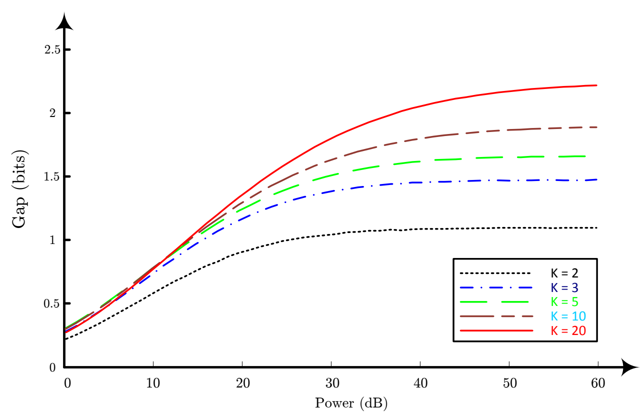


Figure 3.9: Numerical evaluation of the per user gap between the inner-bound and the outer-bound in Theorem 3.2 for P between 0 dB and 60 dB and $K = 2, 3, 5, 10, 20$.

For the per user gap, we analytically show that the outer-bound and lower-bound given in Theorem 3.2 are at most $2\log_2(K+2)$ away from each other. However as depicted in Fig. 3.9, we can see that the per-user gap is in fact smaller than $2\log_2(K+2)$. The argument is similar to the one we presented in (3.34). To derive the gap, we first increase the noise term at all receive antennas to $1+(K-1)(K+2)$. Then similar to (3.34), we obtain

$$
\begin{aligned}
& C_{K\times K}\left(1, 5, \ldots, 1+(j-1)(j+2), \ldots, 1+(K-1)(K+2)\right) \\
& \gamma C_{K\times K} - K\log_2\left(1+(K-1)(K+2)\right). \qquad (3.48)
\end{aligned}
$$

Using (3.48) and considering the gap between $C_{K\times K}$ and $KC_{K\times 1}$, we can show that the gap between the achievable symmetric rate and the symmetric capacity is bounded by $2\log_2(K+2)$.

3.8 Conclusion and Future Directions

In this paper, we studied the capacity region of the multiple-input single-output complex Gaussian Broadcast Channels with delayed CSIT. We showed that a modification of the scheme introduced in [36], can be applied in the finite SNR regime to obtain an inner-bound that is within $2\log_2(K+2)$ bits of the outer-bound. Therefore the gap scales as the logarithm of the number of users. This happens due to noise accumulation during the transmission strategy. Thus, an interesting future direction would be to figure out whether there exists a transmission strategy that results in constant gap (independent of channel parameters, transmission power and number of users) approximation of the capacity region.

Another direction is to consider a two-user MISO BC with delayed CSIT where the noise processes and the channel gains are not distributed as i.i.d. random variables. For example, consider the scenario where the noise processes have different variances. This model captures the scenario where users are located at different distances to the base station. For this setting, even the (generalized) DoF region is not known.

CHAPTER 4

COMMUNICATION WITH LOCAL NETWORK VIEWS

4.1 Introduction

In dynamic wireless networks, channels between any subset of nodes undergo constant changes. As a result, the network state information consisting of channel state information between different nodes is constantly changing. In large networks, nodes often measure channels in their immediate neighborhood but may have limited or no information about channels between nodes many hops away. One reason of only acquiring local information at a node is to ensure scalability of the overall measurement architecture, such that measurement overhead is independent of network size. The local information architecture implies that nodes are only aware of local network state. Then the key question is, how do optimal decentralized decisions perform in comparison to the optimal centralized decisions which rely on full channel state information.

In this paper, we consider multi-source multi-destination two-layer wireless networks and seek fundamental limits of communications when sources have access to limited local CSI. To model local views at the nodes in this sequel, we consider the case where each source-destination (S-D) pair has enough information to perform optimally when other pairs do not interfere. Beyond that, the only other information available at each node is the global network connectivity. We refer to this model of local network knowledge as "local view." The motivation for this model stems from coordination protocols like routing which are often employed in multi-hop networks to discover S-D routes in the network.

In the absence of *global* channel state information, the only feasible solution seems to be the orthogonalization of interfering information sessions. However, surprisingly we show that with inter-session coding, we can outperform the interference avoidance (orthogonalization) techniques. In fact, we show that the gain we obtain from inter-session coding over interference avoidance techniques can be unbounded.

Our main contributions then are as follows. We propose an algebraic framework for inter-session coding that only requires local view at the nodes and combines coding with interference avoidance scheduling. The scheme is a combination of three main techniques: (1) per layer interference avoidance; (2) repetition coding to allow overhearing of interference; and (3) network coding to allow interference neutralization. Our work reveals the important role of relays in interference management over two-layer wireless networks with local view.

We characterize the achievable normalized sum-rate of our proposed scheme. We note that the transmission strategy solely relies on the local view assumption and guarantees the achievability of a fraction of the network capacity with *full* channel state knowledge, regardless of the actual values of the unknown parameters at each node. We then analyze the optimality of the proposed scheme for some classes of networks. We consider two-layer networks: (1) with two relays and any number of source-destination pairs; (2) with three source-destination pairs and three relays; and (3) with folded-chain structure (defined in Section 4.5). We also show that the gain from inter-session coding over interference avoidance scheduling can be unbounded in L-nested folded-chain networks (defined in Section 4.5). To derive information-theoretic outer-bounds in these networks, we seek a worst case channel realization in which achieving beyond a certain fraction of the network capacity would be infeasible with local view.

It is worth noting that since each channel gain can range from zero to a maximum value, our formulation is similar to compound channels [10, 44] with one major difference. In the multi-terminal compound network formulations, *all* nodes are missing identical information about the channels in the network, whereas in our formulation, the local view results in *asymmetric* information about channels at different nodes.

In the literature, many models for imprecise network information have been considered for interference networks. These models range from having no channel state information at the sources, delayed channel state information, or rate-limited feedback links. Most of these works assume fully connected network topology or a small number of users. A study to understand the role of limited network knowledge, was first initiated in [5, 6] for general single-layer networks with *arbitrary* connectivity, where the authors used a message-passing abstraction of network protocols to formalize the notion of local view of the network at each node. The key result was that local-view-based (decentralized) decisions can be either sum-rate optimal or can be arbitrarily worse than the global-view (centralized) sum-capacity. In this work, we focus on two-layer setting and we show that several important additional ingredients are needed compared to the single-layer scenario.

The rest of the paper is organized as follows. In section 4.2, we introduce our network model and the new model for partial network knowledge. In Section 4.3, via a number of examples, we motivate our transmission strategies and the algebraic framework. In Section 4.4, we formally define the algebraic framework and we characterize the performance of the transmission strategy based on this framework. In Section 4.5, we prove the optimality of our strategies for several network topologies. Finally, Section 4.6 concludes the paper.

4.2 Problem Formulation

In this section, we introduce our model for channel and network knowledge at nodes. We also define the notion of normalized sum-capacity which will be used as the performance metric for strategies with local network knowledge.

4.2.1 Network Model and Notations

A network is represented by a directed graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{h_{ij}\}_{(i,j)\in\mathcal{E}}) \tag{4.1}$$

where $\mathcal{V}$ is the set of vertices representing nodes in the network, $\mathcal{E}$ is the set of directed edges representing links among the nodes, and $\{h_{ij}\}_{(i,j)\in\mathcal{E}}$ represents the channel gains associated with the edges and we have $h_{ij} \in \mathbb{C}$.

We focus on two-layer networks where out of $|\mathcal{V}|$ nodes in the network, K are denoted as sources and K are destinations. We label these source and destination nodes by S_i's and D_i's respectively, $i = 1, 2, \ldots, K$. The remaining $|\mathcal{V}| - 2K$ nodes are relay nodes. The layered structure of the network imposes that a source can be only connected to a subset of relays, and a relay can be only connected to a subset of destinations.

The channel input at node V_i at time t is denoted by $X_{\mathsf{V}_i}[t] \in \mathbb{C}$, and the received signal at node V_j at time t is denoted by $Y_{\mathsf{V}_j}[t] \in \mathbb{C}$ given by

$$Y_{\mathsf{V}_j}[t] = \sum_i h_{ij} X_{\mathsf{V}_i}[t] + Z_j[t], \tag{4.2}$$

where $Z_j[t]$ is the additive white complex Gaussian noise with unit variance. The noise processes are distributed independently across space and time, and are independent from the transmit signals. We also assume a power constraint of 1 at all nodes, *i.e.*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\sum_{t=1}^{n} |X_{\mathsf{V}_i}[t]|^2 \right) \leq 1. \tag{4.3}$$

A *route* from a source S_i to a destination D_j is a set of nodes such that there exists an ordering of these nodes where the first one is S_i, last one is D_j, and any two consecutive nodes in this ordering are connected by an edge in the graph.

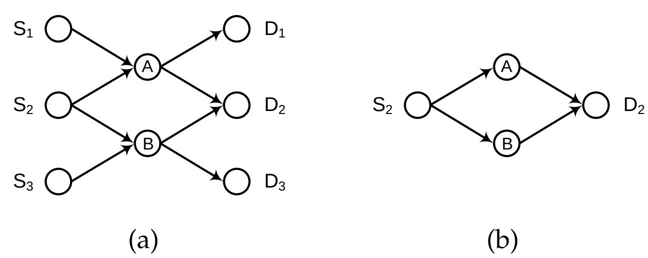


Figure 4.1: *(a) A two-layer wireless network with three S-D pairs and two relays, and (b) the induced subgraph $\mathcal{G}_{22}$.*

Definition 4.1 *Induced subgraph $\mathcal{G}_{ij}$ is a subgraph of $\mathcal{G}$ with its vertex set being the union of all routes from source S_i to a destination D_j, and its edge set being the subset of all edges in $\mathcal{G}$ between the vertices of $\mathcal{G}_{ij}$. Fig. 4.1(b) depicts the induced subgraph $\mathcal{G}_{22}$ for the network in Fig. 4.1(a).*

We say that S-D pair i and S-D j are *non-interfering* if $\mathcal{G}_{ii}$ and $\mathcal{G}_{jj}$ have disjoint vertex sets.

The in-degree function $d_{\text{in}}(\mathsf{V}_i)$ is the number of in-coming edges connected to node V_i. Similarly, the out-degree function $d_{\text{out}}(\mathsf{V}_i)$ is the number of out-going edges connected to node V_i. Note that the in-degree of a source and the out-degree of a destination are both equal to 0. The maximum degree of the nodes in $\mathcal{G}$ is defined as

$$d_{\max} \triangleq \max_{i\in\{1,\ldots,|\mathcal{V}|\}} \left(d_{\text{in}}(\mathsf{V}_i), d_{\text{out}}(\mathsf{V}_i)\right). \tag{4.4}$$

4.2.2 Model for Partial Network Knowledge

We now describe our model for the partial network knowledge that we refer to as *local view*.

- All nodes have full knowledge of the network topology, $(\mathcal{V}, \mathcal{E})$ (*i.e.* they know which links are in $\mathcal{G}$, but not necessarily their channel gains). The network topology knowledge is denoted by side information SI;
- Each source, S_i, knows the gains of all the channels that are in $\mathcal{G}_{ii}$, $i = 1, 2, \ldots, K$. This channel knowledge at a source is denoted by L_{S_i};
- Each node V_i (which is not a source) has the union of the information of all those sources that have a route to it, and this knowledge at node is denoted by L_{V_i}.

For a depiction, consider the network in Fig. 4.2(a). Source S_1 has the knowledge of the channel gains of the links that are denoted by solid red arrows in Fig. 4.2(a). On the other hand, relay A has the union of the information of sources S_1 and S_2. The partial knowledge of relay A is denoted by solid red arrows in Fig. 4.2(b).

Remark 4.1 *Our formulation is a general version of compound channel formulation where nodes have mismatched asymmetric lack of knowledge.*

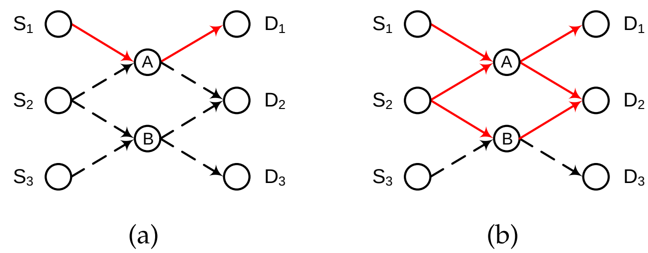


Figure 4.2: *(a) The partial knowledge of the channel gains available at source* S_1 *is denoted by solid red arrows, and (b) the partial knowledge of the channel gains available at relay* A *denoted by solid red arrows.*

4.2.3 Performance Metric

As mentioned before, our goal is to find out the minimum number of time-slots such that each S-D pair is able to reconstruct any transmission snapshot in their induced subgraph over the original network. It turns out that the aforementioned objective is closely related to the notion of normalized sum-capacity introduced in [3, 4]. The connection between the two concepts will be made clear in Section 4.3. Here, we define the notion of normalized sum-capacity which is going to be our metric for evaluating network capacity with local view. Normalized sum-capacity represents the maximum fraction of the sum-capacity with full knowledge that can be always achieved when nodes only have partial knowledge about the network and is defined as follows.

Consider the scenario in which source S_i wishes to reliably communicate message $\mathsf{W}_i \in \{1, 2, \ldots, 2^{NR_i}\}$ to destination D_i during N uses of the channel, $i = 1, 2, \ldots, K$. We assume that the messages are independent and chosen uniformly.

For each source S_i, let message W_i be encoded as $X_{\mathsf{S}_i}^N$ using the encoding function $e_i(\mathrm{W}_i|L_{\mathsf{S}_i}, \mathsf{SI})$, which depends on the available local network knowledge, L_{S_i}, and the global side information, SI.

Each relay in the network creates its input to the channel X_{V_i}, using the encoding function $f_{\mathsf{V}_i}[t](Y_{\mathsf{V}_i}^{(t-1)}|L_{\mathsf{V}_i}, \mathsf{SI})$, which depends on the available network knowledge, L_{V_i}, and the side information, SI, and all the previous received signals at the relay $Y_{\mathsf{V}_i}^{(t-1)} = [Y_{\mathsf{V}_i}[1], Y_{\mathsf{V}_i}[2], \ldots, Y_{\mathsf{V}_i}[t-1]]$. A relay strategy is defined as the union of all encoding functions used by the relays, $\{f_{\mathsf{V}_i}[t](Y_{\mathsf{V}_i}^{(t-1)}|L_{\mathsf{V}_i}, \mathsf{SI})\}$, $t = 1, 2, \ldots, N$.

Destination D_i is only interested in decoding W_i and it will decode the message using the decoding function $\widehat{\mathrm{W}}_i = d_i(Y_{\mathsf{D}_i}^N|L_{\mathsf{D}_i}, \mathsf{SI})$, where L_{D_i} is the destination D_i's network knowledge. Note that the local view can be different from node to node.

Definition 4.2 *A* Strategy $\mathcal{S}_N$ *is defined as the set of: (1) all encoding functions at the sources; (2) all decoding functions at the destinations; and (3) the relay strategy,* i.e.

$$\mathcal{S}_N = \left\{ \begin{array}{l} e_i(W_i|L_{\mathsf{S}_i}, \mathsf{SI}) \\ f_{\mathsf{V}_i}[t](Y_{\mathsf{V}_i}^{(t-1)}|L_{\mathsf{V}_i}, \mathsf{SI}) \\ d_i(Y_{\mathsf{D}_i}^N|L_{\mathsf{D}_i}, \mathsf{SI}) \end{array} \right\} \tag{4.5}$$

for $t = 1, 2, \ldots, N$*, and* $i = 1, 2, \ldots, K$*.*

An error occurs when $\widehat{\mathrm{W}}_i \neq \mathrm{W}_i$ and we define the decoding error probability, λ_i, to be equal to $P(\widehat{\mathrm{W}}_i \neq \mathrm{W}_i)$. A rate tuple $(R_1, R_2, \ldots, R_K)$ is said to be achievable, if there exists a set of strategies $\{\mathcal{S}_j\}_{j=1}^N$ such that the decoding error probabilities $\lambda_1, \lambda_2, \ldots, \lambda_K$ go to zero as $N \to \infty$ for all network states consistent with the side information. Moreover, for any S-D pair i, denote the maximum achievable rate R_i with full network knowledge by C_i. The sum-capacity C_{sum}, is the supremum of $\sum_{i=1}^{K} R_i$ over all possible encoding and decoding functions with full network knowledge.

We will now define the normalized sum-rate and the normalized sum-capacity.

Definition 4.3 ([4]) Normalized sum-rate *of α is said to be achievable, if there exists a set of strategies $\{\mathcal{S}_j\}_{j=1}^{N}$ such that following holds. As N goes to infinity, strategy $\mathcal{S}_N$ yields a sequence of codes having rates R_i at the source S_i, $i = 1, \ldots, K$, such that the error probabilities at the destinations, $\lambda_1, \cdots \lambda_K$, go to zero, satisfying*

$$\sum_{i=1}^{K} R_i \gamma \alpha C_{\text{sum}} - \tau$$

for all the network states consistent with the side information, and for a constant τ that is independent of the channel gains.

Definition 4.4 ([4]) Normalized sum-capacity *α^*, is defined as the supremum of all achievable normalized sum-rates α. Note that $\alpha^* \in [0, 1]$.*

Remark 4.2 *While it may appear that the notion of normalized sum-capacity is pessimistic, the results presented here are much like any other compound channel analysis that aims to optimize the worst case scenario. We adopted the normalized capacity as our metric which opens the door for exact (or near-exact) results for several cases as we will see in this paper. We view our current formulation as a step towards the general case, much like the commonly-accepted methodology where channel state information is assumed known perfectly in first steps of a new problem (*e.g.*, MIMO, interference* *channels, scaling laws), even though the assumption of perfectly known channel state information is almost never possible.*

4.3 Motivating Example

Before diving into the main results, we use an example to illustrate the mechanisms that allow us outperform interference avoidance with only local view. Consider the two-layer network depicted in Fig. 4.3(a). For simplicity, we refer to the relays as relay A, relay B, and relay C. It is straightforward to see that interference avoidance requires the three information sessions to be orthogonalized. Thus, it takes three time-slots to reconstruct the induced subgraphs of Fig. 4.3(b), 4.3(c), and 4.3(d), which results in a normalized sum-rate of $\alpha = \frac{1}{3}$.

However, we now show that with inter-session coding, it is possible to achieve $\alpha = \frac{1}{2}$ and reconstruct any transmission snapshot over the three induced subgraphs in only two time-slots with local view. Therefore, a normalized sum-rate of $\frac{1}{2}$ is achievable[1]. In a sense, inter-session coding allows interfering information sessions to co-exist. The key ingredient is the mixing of signals at the relays as we describe shortly.

Consider any strategy for S-D pairs 1, 2, and 3 as illustrated in Fig. 4.3(b), 4.3(c), and 4.3(d). We split the communication block into two time-

[1]The reason for achieving a normalized sum-rate of $\frac{1}{2}$ if we can reconstruct any transmission snapshot over the three induced subgraphs in only two time-slots, is as follows. Since any transmission strategy for the induced subgraphs can be implemented in the original network by using only two time-slots, we can implement the strategies that achieve the capacity for any S-D pair i with full network knowledge, *i.e.* C_i, over two time-slots up to a constant term which is due to the difference in noise variances. Hence, we can achieve $\frac{1}{2}(C_1 + C_2 + C_3) - \tau$. On the other hand, we have $C_{\text{sum}} \leq C_1 + C_2 + C_3$. As a result, we can achieve a set of rates such that $\sum_{i=1}^{3} R_i \geq \frac{1}{2} C_{\text{sum}} - \tau$, and by the definition of normalized sum-rate, we achieve $\alpha = \frac{1}{2}$.

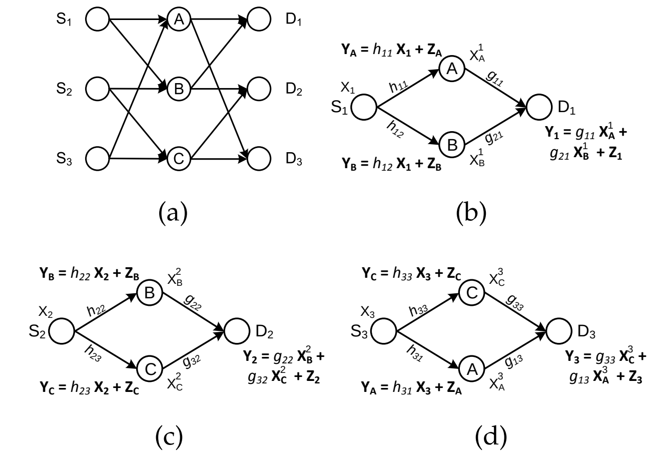


Figure 4.3: *(a) a two-layer network in which we need to incorporate network coding to achieve the normalized sum-capacity, (b), (c) and (d) the induced subgraphs of S-D pairs 1,2,and 3 respectively.*

slots of equal length and we describe the scheme for each layer separately.

Transmission scheme for the first layer: Over the first time-slot, source S_1 transmits the same codeword as if it was in the induced subgraph, and S_2 transmits the same codeword as if it was in the induced subgraph multiplied by -1.

Over the second time-slot, S_3 transmits the same codeword as if it was in the induced subgraph and S_2 repeats its transmitted signal from the previous time-slot.

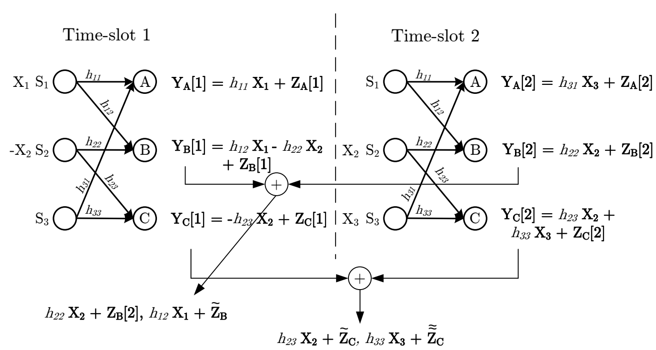


Figure 4.4: *Achievability strategy for the first layer of the network in Figure 4.3(a).*

Transmission scheme for the first layer: During first time-slot, relays A and B transmit X_{A}^1 and X_{B}^1 respectively, whereas relay C transmits $\frac{1}{\sqrt{2}}\left(X_{\mathsf{C}}^3 - X_{\mathsf{C}}^2\right)$ as depicted in Fig. 4.5.

During second time-slot , relays B and C transmit X_{B}^2 and X_{C}^2 respectively, whereas relay A transmits $\frac{1}{\sqrt{2}}\left(X_{\mathsf{A}}^3 - X_{\mathsf{A}}^1\right)$.

Recovering signals at relays: At the end of the first time-slot, relay A receives the signal from source S_1, and relay C receives the signal from source S_2. Relay B receives the summation of the signals from sources S_1 and S_2.

Upon completion of the second time-slot, relay B receives the signal from source S_2 and it can use its received signals over the two time-slots to recover the signal of source S_1 as well. Now, if relay C adds its received signals over the two time-slots, it also recovers the signal[2] of source S_3 as depicted in Fig. 4.4.

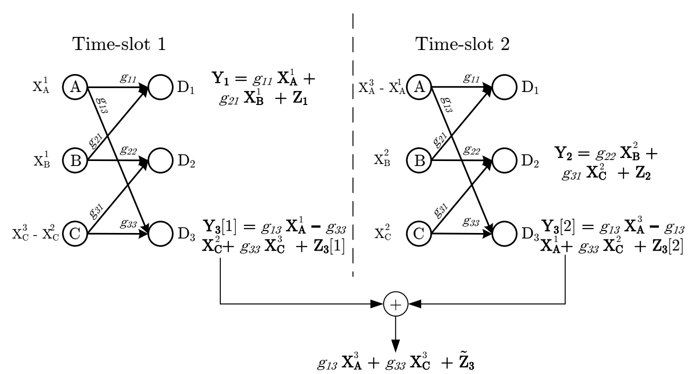


Figure 4.5: *Achievability strategy for the second layer of the network in Figure 4.3(a).*

Decoding at destinations: At the end of the first time-slot, destination D_1 receives the same signal as in Fig. 4.3(b) (with different noise variance).

[2]We note that the effective noise variance at relays B and C would be twice the noise variance of its induced subgraph for sources S_1 and S_3 respectively. However, by the definition of normalized sum-capacity, this does not affect the achievable normalized sum-rate

Finally, at the end of the second time-slot, destination D_2 receives the same signal as in Fig. 4.3(c) (with different noise variance). If destination D_3 adds its received signals over the two time-slots, it recovers the same signal as in Fig. 4.3(d) (up to noise term). Therefore, each destination receives the same signal as if it was only in its corresponding diamond network over two time-slots. Hence, the normalized sum-rate of $\alpha = \frac{1}{2}$ is achievable.

In the following section, we generalize the ideas presented in this section by introducing an algebraic framework for inter-session coding.

4.4 An Algebraic Framework for Inter-Session Coding with Local View

In this section, we present an algebraic framework for inter-session coding with local view in two-layer wireless networks over $T \in \mathbb{N}$ time-slots. We first go over some notations and definitions. Consider a two-layer wireless network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{h_{ij}\}_{(i,j)\in\mathcal{E}})$ with K source-destination pairs and $|\mathcal{V}| - 2K$ relay nodes.

We label the sources as $\mathsf{S}_1, \ldots, \mathsf{S}_K$, the destinations as $\mathsf{D}_1, \ldots, \mathsf{D}_K$, and the relays as $\mathsf{V}_1, \ldots, \mathsf{V}_{|\mathcal{V}|-2K}$.

Definition 4.5 *We assign a number N_{V_j} to relay V_j defined as*

$$N_{\mathsf{V}_j} \triangleq \left|\left\{i|\mathsf{V}_j \in \mathcal{G}_{ii}\right\}\right|, \tag{4.6}$$

and we assign a number N_{D_j} to destination D_j defined as

$$N_{\mathsf{D}_j} \triangleq \sum_{i=1}^{K} \Big(\textit{number of routes from } \mathsf{S}_i \textit{ to } \mathsf{D}_j\Big). \tag{4.7}$$

We now describe the assignment of transmit and receive matrices to each node in the network where all entries to these matrices are in $\{-1, 0, 1\}$. These matrices form the basis for inter-session coding as described later in this section.

- We assign a transmit vector $\mathbf{T}_{\mathsf{S}_i}$ of size $T \times 1$ to source S_i where each row corresponds to a time-slot;
- To each relay V_j, we assign a receive matrix $\mathbf{R}_{\mathsf{V}_j}$ of size $T \times d_{\text{in}}(\mathsf{V}_j)$, and we label each column with a source that has a route to that relay. The entry at row t and column labeled by S_i in $\mathbf{R}_{\mathsf{V}_j}$ is equal to the entry at row t of $\mathbf{T}_{\mathsf{S}_i}$;
- We assign a transmit matrix $\mathbf{T}_{\mathsf{V}_j}$ of size $T \times N_{\mathsf{V}_j}$ to each relay where each column corresponds to a S-D pair i where the relay belongs to $\mathcal{G}_{ii}$;
- Finally, to each destination D_i, we assign a receive matrix $\mathbf{R}_{\mathsf{D}_i}$ of size $T \times N_{\mathsf{D}_i}$ where each column corresponds to a route from a source S_j through a specific relay $\mathsf{V}_{j'}$ and is labeled as $\mathsf{S}_j : \mathsf{V}_{j'}$. The entry at row t and column labeled by $\mathsf{S}_j : \mathsf{V}_{j'}$ in $\mathbf{R}_{\mathsf{D}_i}$ is equal to the entry at row t and column labeled by S_j of $\mathbf{T}_{\mathsf{V}_{j'}}$.

Remark 4.3 *Since nodes have only local view, they are unaware of the interfering channel gains. Thus, the entries to the vectors and matrices are only chosen to be* $-1, 0,$ *or* 1*. This way they do not rely on the actual value of the interfering channel gains but rather the topology.*

We say that an assignment of transmit and receive matrices to the nodes in the network is *valid* if the following conditions are satisfied:

C.1 [Source Condition] For $i = 1, 2, \ldots, K$,

$$\text{Rank}\left(\mathbf{T}_{\mathsf{S}_i}\right) = 1. \tag{4.8}$$

Intuitively, condition **C.1** guarantees that every transmitter communicates its signal at least once during the T time-slots.

C.2 [Relay Condition] For any relay V_j, if $\mathsf{V}_j \in \mathcal{G}_{ii}$, then

$$\begin{array}{c} \mathsf{S}_i \qquad\qquad\qquad \\ \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \end{array} \in \operatorname{rowspan}\left(\mathbf{R}_{\mathsf{V}_j}\right). \tag{4.9}$$

Intuitively, if condition **C.2** is satisfied, then each relay has access to the signal of any S-D pair that has a route through that relay.

C.3 [Destination Condition] For destination D_i, if relays $\mathsf{V}_{j_1}, \dots, \mathsf{V}_{j_{N_{\mathsf{D}_i}}}$ are on routes from S_i to D_i, then

$$\begin{array}{c} \begin{array}{ccc} \mathsf{S}_i : \mathsf{V}_{j_1} & \dots & \mathsf{S}_i : \mathsf{V}_{j_{N_{\mathsf{D}_i}}} \end{array} \qquad\qquad \\ \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \end{array}$$
$$\in \operatorname{rowspan}\left(\mathbf{R}_{\mathsf{D}_i}\right). \tag{4.10}$$

Finally, condition **C.3** guarantees the feasibility of interference neutralization at each destination.

Furthermore, we limit all linear operations in conditions **C.2** and **C.3** required to obtain (4.9) and (4.10), to simply be a summation of a subset of rows.

Remark 4.4 *It is straightforward to show that a necessary condition for* **C.3** *is for each* $\mathbf{T}_{\mathsf{V}_j}$ *to have full column rank. This gives us a lower bound on* T *which is* $\max_{\mathsf{V}_j} \left|\left\{ i | \mathsf{V}_j \in \mathcal{G}_{ii} \right\}\right|$*. Moreover, TDMA is always a lower bound on the performance of any strategy and thus, provides us with an upper bound on* T*. In other words, we have*

$$\max_{\mathsf{V}_j} \left|\left\{ i | \mathsf{V}_j \in \mathcal{G}_{ii} \right\}\right| \leq T \leq K, \tag{4.11}$$

which implies that for two-layer wireless networks, we have

$$\frac{1}{K} \leq \alpha^* \leq \frac{1}{\max_{\mathsf{V}_j} \left|\left\{i | \mathsf{V}_j \in \mathcal{G}_{ii}\right\}\right|}. \tag{4.12}$$

Consider a transmission snapshot in each of the K induced subgraphs in the absence of all other S-D pairs where

- Node V_i (similarly a source) in the induced subgraph $\mathcal{G}_{jj}$ transmits $X^j_{V_i}$,
- Node V_i (similarly a destination) in the induced subgraph $\mathcal{G}_{jj}$ receives

$$Y^j_{\mathsf{V}_i} = \sum_{i':(i',i)\in\mathcal{E}} h_{i'i} X^j_{\mathsf{V}_{i'}} + Z^j_{\mathsf{V}_i}. \tag{4.13}$$

We define a transmission strategy based on the assigned transmit and receive matrices as follows. At any time-slot t:

- Source S_i transmits

$$X_{\mathsf{S}_i}[t] = \mathbf{T}_{\mathsf{S}_i}(t) X^i_{\mathsf{S}_i}, \tag{4.14}$$

- Each relay V_j transmits

$$X_{\mathsf{V}_j}[t] = \frac{1}{\sqrt{T}} \sum_{i=1}^{K} \mathbf{T}_{\mathsf{V}_j}(t, \mathsf{S}_i) X^i_{\mathsf{V}_j} \tag{4.15}$$

  where $t = 1, \ldots, T$, and the coefficient $\frac{1}{T}$ is to guarantee the power constraint at nodes.

Before stating our main results, we show how to interpret the transmission strategy of Section 4.3 in the algebraic framework presented above. We assign a transmit vector $\mathbf{T}_{\mathsf{S}_i}$ of size 2×1 to source S_i.

$$\mathbf{T}_{\mathsf{S}_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{T}_{\mathsf{S}_2} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \mathbf{T}_{\mathsf{S}_3} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.16}$$

To each relay, we assign a receive matrix of size 2×2 where each row corresponds to a time-slot, and each column corresponds to a source that has a route to that relay. This receive matrix is formed by concatenating the transmit vectors of the sources that are connected to the relay. In our example, we have

$$\mathbf{R}_{\mathsf{A}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_1 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}, \quad \mathbf{R}_{\mathsf{B}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_1 & \mathsf{S}_2 \end{array} \\ \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \end{array},$$

$$\mathbf{R}_{\mathsf{C}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_2 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix} \end{array}. \tag{4.17}$$

From the receive matrices in (4.17), we can easily check whether each relay has access to the same received signals as if it was in the diamond networks of Figures 4.3(b), 4.3(c), and 4.3(d). In fact all three matrices in (4.17) are full rank.

We also assign a transmit matrix of size 2×2 to each relay where each column corresponds to a S-D pair i such that the relay belongs to $\mathcal{G}_{ii}$. During time-slot t, the relay communicates the signal it has for S-D pair i multiplied by the entry at row t and column S_i of its transmit matrix. In our example, we have

$$\mathbf{T}_{\mathsf{A}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_1 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \end{array}, \quad \mathbf{T}_{\mathsf{B}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_1 & \mathsf{S}_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array},$$

$$\mathbf{T}_{\mathsf{C}} = \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cc} \mathsf{S}_2 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \end{array}. \tag{4.18}$$

Here, if in row t more than one non-zero entry appears, then the relay creates the normalized linear combination of the signals it has for the S-D pairs that have a non-zero entry in their column and transmits it.

Now, to each destination D_i, we assign a receive matrix $\mathbf{R}_{\mathsf{D}_i}$ of size 2×4 where each row corresponds to a time-slot, and each column corresponds to a route from a source through a specific relay (*e.g.*, a route from S_1 through relay V_1 is denoted by $\mathsf{S}_1 : \mathsf{A}$). In fact, $\mathbf{R}_{\mathsf{D}_i}$ is formed by concatenating (and reordering columns of) the transmit matrices of the relays that are connected to D_i. In our example, we have

$$
\begin{aligned}
\mathbf{R}_{\mathsf{D}_1} &= \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cccc} \mathsf{S}_1:\mathsf{A} & \mathsf{S}_1:\mathsf{B} & \mathsf{S}_2:\mathsf{B} & \mathsf{S}_3:\mathsf{A} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \end{pmatrix} \end{array}, \\
\mathbf{R}_{\mathsf{D}_2} &= \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cccc} \mathsf{S}_1:\mathsf{B} & \mathsf{S}_2:\mathsf{B} & \mathsf{S}_2:\mathsf{C} & \mathsf{S}_3:\mathsf{C} \end{array} \\ \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}, \\
\mathbf{R}_{\mathsf{D}_3} &= \begin{array}{c} \\ t=1 \\ t=2 \end{array} \begin{array}{c} \begin{array}{cccc} \mathsf{S}_1:\mathsf{A} & \mathsf{S}_2:\mathsf{C} & \mathsf{S}_3:\mathsf{A} & \mathsf{S}_3:\mathsf{C} \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \end{array}.
\end{aligned} \tag{4.19}
$$

From the receive matrices in (4.19), we can verify whether each destination can recover its corresponding signals or not. For instance, from $\mathbf{R}_{\mathsf{D}_1}$, we know that in time-slot 1, the signals from S_1 through relays A and B are received interference-free. From $\mathbf{R}_{\mathsf{D}_3}$, we see that at each time-slot one of the two signals that D_3 is interested in is received. However with a linear row-operation, we can create a row where 1's only appear in the columns associated with S_3.

More precisely,

$$\mathbf{R}_{\mathsf{D}_3}(1,:) + \mathbf{R}_{\mathsf{D}_3}(2,:) \tag{4.20}$$
$$= \begin{array}{c} \begin{array}{cccc} \mathsf{S}_1:\mathsf{A} & \mathsf{S}_2:\mathsf{C} & \mathsf{S}_3:\mathsf{A} & \mathsf{S}_3:\mathsf{C} \end{array} \\ \left(\begin{array}{cccc} 0 & 0 & 1 & 1 \end{array}\right) \end{array},$$

or equivalently

$$\begin{array}{c} \begin{array}{cccc} \mathsf{S}_1:\mathsf{A} & \mathsf{S}_2:\mathsf{C} & \mathsf{S}_3:\mathsf{A} & \mathsf{S}_3:\mathsf{C} \end{array} \\ \left(\begin{array}{cccc} 0 & 0 & 1 & 1 \end{array}\right) \end{array} \in \mathrm{rowspan}\left(\mathbf{R}_{\mathsf{D}_3}\right). \tag{4.21}$$

Thus, each destination can recover its intended signal interference-free (up to noise term).

The following theorem characterizes the achievable normalized sum-rate of transmission strategy described above for a *valid* choice of transmit and receive matrices.

Theorem 4.1 *For a K-user two-layer network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{h_{ij}\}_{(i,j)\in\mathcal{E}})$ with local view, if there exists a valid assignment of transmit and receive matrices, then a normalized sum-rate of $\alpha = \frac{1}{T}$ is achievable.*

Proof: Network $\mathcal{G}$ has K S-D pairs. We need to show that any transmission snapshot over these induced subgraphs can be implemented in $\mathcal{G}$ over T time-slots such that all nodes receive the same signal as if they were in the induced subgraphs (up to noise terms) using the transmission strategy described above.

At any time instant $t \in \{1, 2, \ldots, T\}$, relay V_i (similarly a destination) receives

$$Y_{\mathsf{V}_i}[t] = \sum_{i':(i',i)\in\mathcal{E}} h_{i'i} X_{\mathsf{V}_{i'}}[t] + Z_{\mathsf{V}_i}[t], \quad t = 1, \ldots, T. \tag{4.22}$$

Reconstructing the received signals: Based on the transmission strategy described above, we need to show that at any relay V_i (similarly a destination), the received signal $Y^j_{\mathsf{V}_i}$ can be obtained up to the noise term.

Based on **Relay Condition C.2**, we know that there exists a matrix $\mathbf{G}_{\mathsf{V}_i}$ of size $N_{\mathsf{V}_i} \times T$ with entries in $\{0, 1\}$, and a choice of ℓ, such that if $\mathsf{V}_i \in \mathcal{G}_{jj}$ then

$$\mathbf{G}_{\mathsf{V}_i}(\ell, :)\mathbf{R}_{\mathsf{V}_i} = \begin{array}{c} \mathsf{S}_i \\ \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \end{array}. \tag{4.23}$$

Therefore, we have

$$\sum_{t=1}^{T} \mathbf{G}_{\mathsf{V}_i}(\ell, t) Y_{\mathsf{V}_i}[t] = Y^j_{\mathsf{V}_i} + \tilde{Z}^j_{\mathsf{V}_i}, \tag{4.24}$$

where $\tilde{Z}^j_{\mathsf{V}_i}$ is a zero-mean noise term that is independent form the signals and has variance $\left(\sum_{t=1}^{T} \mathbf{G}(\ell, t)\right) - 1$. It is easy to see that the variance of $\tilde{Z}^j_{\mathsf{V}_i}$ is bounded by $T - 1$. This implies that V_i is able to cancel out all interference by appropriately summing the received signals at different time-slots (at the cost of a higher noise variance).

Similar argument holds for any destination D_i. Based on **Destination Condition C.3**, we know that there exists a matrix $\mathbf{G}_{\mathsf{D}_i}$ of size $N_{\mathsf{D}_i} \times T$ with entries in $\{0, 1\}$, and a choice of ℓ, such that if relays $\mathsf{V}_{j_1}, \dots, \mathsf{V}_{j_{N_{\mathsf{D}_i}}}$ are on routes from S_i to D_i, then

$$\mathbf{G}_{\mathsf{D}_i}(\ell, :)\mathbf{R}_{\mathsf{V}_i} = \begin{array}{cccccc} \mathsf{S}_i:\mathsf{V}_{j_1} & \dots & \mathsf{S}_i:\mathsf{V}_{j_{N_{\mathsf{D}_i}}} & & & \\ \multicolumn{6}{c}{\begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}} \end{array}. \tag{4.25}$$

Therefore, we have

$$\sqrt{T} \sum_{t=1}^{T} \mathbf{G}_{\mathsf{D}_i}(\ell, t) Y_{\mathsf{D}_i}[t] = Y^i_{\mathsf{V}_i} + \tilde{Z}^i_{\mathsf{D}_i}, \tag{4.26}$$

where $\tilde{Z}^i_{\mathsf{V}_i}$ is a zero-mean noise term that is independent form the signals and has variance less than or equal to $T^2 - 1$. In a sense, each destination is able to cancel out all interference by appropriately summing the received signals at different time-slots (at the cost of a higher noise variance).

As we have seen in (4.24) and (4.26), due to the row operations, the resulting signal may have a noise variance that is different from the original noise process. However, the resulting variance is independent of the channel gains. Thus, we need to show that with the row operations performed on transmit and received signals, the capacity for each S-D pair remains within a constant of the capacity of the induced subgraphs when there is no interference. The following lemma which is proved in Appendix C.2 provides us with this result.

Lemma 4.1 *Consider a complex multi-hop Gaussian relay network with one source* S *and one destination* D *represented by a directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{h_{ij}\}_{(i,j)\in\mathcal{E}})$ *where* $\{h_{ij}\}_{(i,j)\in\mathcal{E}}$ *represent the channel gains associated with the edges.*

We assume that at each receive node the additive white complex Gaussian noise has variance σ^2*. We also assume a power constraint of* P *at all nodes,* i.e. $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(\sum_{t=1}^n |X_{\mathsf{V}_i}[t]|^2) \leq P$*. Denote the capacity of this network by* $C(\sigma^2, P)$*. Then for any* $\kappa \geq 1$*,* $\kappa \in \mathbb{R}$*, we have*

$$C(\sigma^2, P) - \tau \leq C(\kappa\sigma^2, P/\kappa) \leq C(\sigma^2, P), \tag{4.27}$$

where $\tau = |\mathcal{V}|\,(2\log\kappa + 17)$ *is a constant independent of channel gains,* P*, and* σ^2*.*

Note that if we increase the noise variance at all nodes to $\left(T^2 - 1\right)$, we can only decrease the capacity. Next, applying Lemma 4.1 (with $\kappa = T^2 - 1$), we conclude that the achievable rate for each S-D pair after using inter-session coding, remains within a constant of the capacity of the induced subgraphs when there is no interference. The constant is independent of the channel gains and the transmit power as described in Lemma 4.1. This completes the proof of Theorem 4.1. ■

In the following section, we analyze the optimality of the proposed scheme for some classes of networks. We consider two-layer networks: (1) with two relays and any number of source-destination pairs; (2) with three source-destination pairs and three relays; and (3) with folded-chain structure. We also show that the gain from inter-session coding over interference avoidance scheduling can be unbounded in L-nested folded-chain networks.

4.5 Optimality of the Strategies

In this section, we consider several classes of networks and derive the minimum number of time-slots such that any transmission snapshot over induced subgraphs can be reconstructed as if there is no interference present. In other words, we characterize the normalized sum-capacity of such networks. We derive information-theoretic outer-bounds on the normalized sum-capacity in these networks by providing a worst case channel realization in which achieving beyond a certain fraction of the network capacity would be infeasible with local view.

As mentioned before, a key ingredient in inter-session coding is the role of relays and the mixing of the signals that takes place at the relays. To gain a better perspective of the role of relays in two-layer wireless networks with local view, we consider several network topologies with different number of relays and gradually increase this number. We first provide an overview of our results.

4.5.1 Overview

We start by considering a class of two-layer networks with only two relays in Section 4.5.2. The normalized sum-capacity of such networks is one over the maximum node degree in the network $d_{\max}$ given by

$$d_{\max} \triangleq \max_{i \in \{1,\ldots,|\mathcal{V}|\}} \left(d_{\text{in}}(\mathsf{V}_i), d_{\text{out}}(\mathsf{V}_i) \right). \tag{4.28}$$

It turns out that to achieve this normalized sum-capacity, relays have to smartly schedule the information flows and no coding is necessary. However, we point out that the performance of a naive orthogonalization scheme can be very poor. In fact, the scheduling of information sessions in the two layers can be different and this task is carried on by the relays.

We then increase the number of relays to three and consider networks with three information sessions in Section 4.5.3. As we already know from Section 4.3, there exists a network topology in this family where without inter-session coding, achieving the normalized sum-capacity is not possible. We show that for this class of networks, the normalized sum-capacity can only take values in $\{1, 1/2, 1/3\}$.

Beyond three relays, we consider a class of two-layer wireless networks in Section 4.5.4 that we refer to as two-layer (K, m) folded-chain networks. In such networks, we have k information sessions and k relays connected in a specific ordering such that the maximum degree is m. For this class, we show that the normalized sum-capacity is equal to $1/m$. To achieve this normalized sum-capacity, inter-session coding is required and the performance of orthogonalization schemes can be far from optimal. This class can be considered as a generalization of the network we considered in Section 4.3.

Finally, in order to show that the gain of inter-session coding over interference avoidance techniques can be unbounded, we consider a single-layer topology that we refer to as L-nested folded-chain network in Section 4.5.5. We show that the gain of inter-session coding over interference avoidance techniques grows exponentially with parameter L which depends on the number of users in the network.

4.5.2 $K \times 2 \times K$ Networks

We now move to two-layer networks and start with a special class of networks with two relays as defined below.

Definition 4.6 *A $K \times 2 \times K$ network is a two-layer network (as defined in Section 4.2.1) with $|\mathcal{V}| - 2K = 2$.*

We establish the normalized sum-capacity of such networks in the following theorem.

Theorem 4.2 *The normalized sum-capacity of a $K \times 2 \times K$ network with local view, is*

$$\alpha^* = \frac{1}{d_{\max}}, \tag{4.29}$$

where $d_{\max}$ is defined in (4.4).

Proof:

The result for $K = 1$ is trivial, so we assume $K > 1$. We refer to the two relays as A_1 and A_2, see Fig. 4.6.

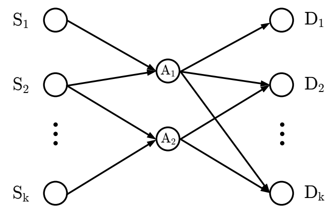


Figure 4.6: *Illustration of a $K \times 2 \times K$ network.*

Achievability: We divide the S-D pair ID's into 3 disjoint subsets as follows:

- $\mathcal{J}_i$ is the set of all the S-D pair ID's such that the corresponding source is connected *only* to relay A_i, $i = 1, 2$;
- $\mathcal{J}_{12}$ is the set of all the other S-D pair ID's. In other words, $\mathcal{J}_{12}$ is the set of all the S-D pair ID's where the corresponding source is connected to both relays.

Without loss of generality assume that $d_{\text{in}}(\mathsf{A}_2) \geq d_{\text{in}}(\mathsf{A}_1)$, and rearrange sources such that

$$\begin{aligned}
\mathcal{J}_1 &= \{1, 2, \ldots, d_{\text{in}}(\mathsf{A}_1)\}, \\
\mathcal{J}_2 &= \{d_{\text{in}}(\mathsf{A}_1) + 1, \ldots, d_{\text{in}}(\mathsf{A}_1) + d_{\text{in}}(\mathsf{A}_2)\}, \\
\mathcal{J}_{12} &= \{d_{\text{in}}(\mathsf{A}_1) + d_{\text{in}}(\mathsf{A}_2) + 1, \ldots, K\}.
\end{aligned} \tag{4.30}$$

We pick the smallest member of $\mathcal{J}_1$ and the smallest member of $\mathcal{J}_2$, and we set the first entry of the corresponding transmit vectors equal to 1 and all other entries equal to zero. We remove these members from $\mathcal{J}_1$ and $\mathcal{J}_2$. Then, we pick two smallest members of $\mathcal{J}_1$ and $\mathcal{J}_2$ and we set the second entry of the corresponding transmit vectors equal to 1 and all other entries equal to zero. We continue this process until $\mathcal{J}_1$ is empty. For any remaining S-D pair ID j, we set the j^{th} entry of the corresponding transmit vector equal to 1 and all other entries equal to zero.

In the second layer, we divide S-D pair ID's based on the connection of destinations to relays, *i.e.* $\mathcal{J}'_i$ is the set of all the S-D pair ID's such that the corresponding destination is connected to relay A_i, $i = 1, 2$, and $\mathcal{J}'_{12}$ is the set of all the other S-D pair ID's. To A_i, we assign a transmit matrix of size $T \times \left(\mathcal{J}'_i + \mathcal{J}'_{12}\right)$, $i = 1, 2$, as described below.

Without loss of generality assume that $d_{\text{out}}(\mathsf{A}_2) \geq d_{\text{out}}(\mathsf{A}_1)$. We pick one member of $\mathcal{J}'_1$ and one member of $\mathcal{J}'_2$ randomly, and at the first row of $\mathbf{T}_{\mathsf{A}_1}$ and $\mathbf{T}_{\mathsf{A}_2}$, we set the entry at the column corresponding to the picked indices equal to 1 and all other entries at those rows equal to zero. We remove these members from $\mathcal{J}'_1$ and $\mathcal{J}'_2$. We then pick one member of $\mathcal{J}'_1$ and one member of $\mathcal{J}'_2$ randomly, and at the second row of $\mathbf{T}_{\mathsf{A}_1}$ and $\mathbf{T}_{\mathsf{A}_2}$, we set the entry at the column corresponding to the picked indices equal to 1 and all other entries at those rows equal to zero. We continue this process till $\mathcal{J}'_1$ is empty.

We then pick one of the remaining S-D pair IDs (members of $\mathcal{J}'_{12}$ and the remaining members of $\mathcal{J}'_2$) and assign a 1 at the next available row and to the column corresponding to the picked index in the corresponding transmit matrix and all other entries at those rows equal to zero. We continue the process until no S-D pair ID is left.

Condition **C.1** is trivially satisfied. The corresponding transmission strategy in this case would be a "per layer" interference avoidance, *i.e.* if in the first hop two sources are connected to the same relay, they do not transmit simultaneously, and if in the second hop two destinations are connected to the same relay, they are not going to be served simultaneously. Therefore, since the scheme does not allow any interference to be created, no row operations on the receive matrices is required and conditions **C.2** and **C.3** are satisfied.

Note that according to the assignment of the vectors and matrices, we require

$$T = \max\{d_{\text{in}}(\mathsf{A}_2), d_{\text{out}}(\mathsf{A}_2)\} = d_{\max}. \tag{4.31}$$

Hence from Theorem 4.1, we know that a normalized sum-rate of $\alpha = \frac{1}{d_{\max}}$ is achievable.

Converse: Assume that a normalized sum-rate of α is achievable, we show that $\alpha \leq \frac{1}{d_{\max}}$. It is sufficient to consider two cases: (1) $d_{\max} = d_{\text{in}}(\mathsf{A}_1)$, and (2) $d_{\max} = d_{\text{out}}(\mathsf{A}_1)$. Here, we provide the proof for case (1) and we postpone the proof for case (2) to Appendix C.1.

The proof is based on finding a worst case scenario. Thus to derive the upper bound, we use specific assignment of channel gains. Consider D_j for $j \in \mathcal{J}_1$; any such destination is either connected to relay A_1 or to both relays. If it is connected to both, then set the channel gain from relay A_2 equal to 0. Follow similar steps for the members of $\mathcal{J}_2$ (if the destination is connected to both, then set the channel gain from relay A_1 equal to 0).

Now, consider D_j for $j \in \mathcal{J}_{12}$; such destination is either connected to only one relay or to both relays. If such destination is connected to both relays, assign the channel gain of 0 to one of the links connecting it to a relay (pick this link at random). For all other links in the network, assign a channel gain of $h \in \mathbb{C}$.

With the channel gain assignment described above, it is straight forward to see that

$$\forall\ j\ \in\ \mathcal{J}_1:\ H\left(W_j|Y^n_{\mathsf{A}_1}, L_{\mathsf{A}_1}, \mathsf{SI}\right) \leq n\epsilon_n, \tag{4.32}$$

where $\epsilon_n \to 0$ as $n \to \infty$. Thus, relay A_1 is able to decode all messages coming from sources corresponding to the members of $\mathcal{J}_1$.

A similar claim holds for relay A_2 and all messages coming from $\mathcal{J}_2$. Relays A_1 and A_2 decode the messages coming from members of $\mathcal{J}_1$ and $\mathcal{J}_2$ respectively, and remove their contributions from the received signals.

Now, it is easy to see that

$$\forall\ \ell\ \in\ \mathcal{J}_{12}:\ H\left(W_\ell|Y^n_{\mathsf{A}_1}, X^n_{j\in\mathcal{J}_1}, L_{\mathsf{A}_1}, \mathsf{SI}\right) \leq n\epsilon_n. \tag{4.33}$$

Thus each relay is able to decode the rest of the messages (they receive same signals with different noise terms). This means that relay A_1 is able to decode all the messages from $\mathcal{J}_1$ and $\mathcal{J}_{12}$, *i.e.*

$$\sum_{j\in(\mathcal{J}_1\cup\mathcal{J}_{12})} H\left(W_j|Y^n_{\mathsf{A}_1}, L_{\mathsf{A}_1}, \mathsf{SI}\right) \leq n\epsilon_n, \tag{4.34}$$

which in turn implies that

$$n\left(\sum_{j\in(\mathcal{J}_1\cup\mathcal{J}_{12})} R_j - \epsilon_n\right) \leq h\left(Y^n_{\mathsf{A}_1}|L_{\mathsf{A}_1}, \mathsf{SI}\right). \tag{4.35}$$

Note that $d_{\text{in}}(\mathsf{A}_1) = |\mathcal{J}_1| + |\mathcal{J}_{12}|$.

Given the assumption of local view, in order to achieve a normalized sum-rate of α, each source should transmit at a rate greater than or equal to $\alpha\log(1 + |h|^2) - \tau$ since from each source's point of view, it is possible that the other S-D pairs have capacity 0. From the MAC capacity at relay A_1, we get

$$d_{\text{in}}(\mathsf{A}_1)(\alpha\log(1 + |h|^2) - \tau) \leq \log(1 + d_{\text{in}}(\mathsf{A}_1) \times |h|^2), \tag{4.36}$$

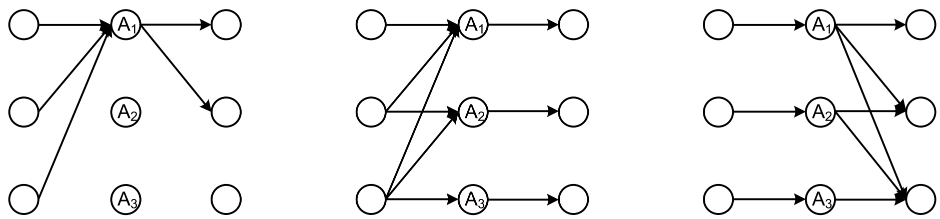


Figure 4.7: *The normalized sum-capacity of a* $3 \times 3 \times 3$ *network with* 1*-local view,* α^**, is equal to* $1/3$ *if and only if one of the graphs in this figure is a subgraph of* $\mathcal{G}$*.*

which results in

$$d_{\text{in}}(\mathsf{A}_1)(\alpha \log(1 + |h|^2) - \tau) \leq \log(d_{\text{in}}(\mathsf{A}_1)) + \log(1 + \times|h|^2). \tag{4.37}$$

Hence, we have

$$(d_{\text{in}}(\mathsf{A}_1)\alpha - 1) \log(1 + |h|^2) \leq \log(d_{\text{in}}(\mathsf{A}_1)) + d_{\text{in}}(\mathsf{A}_1)\tau. \tag{4.38}$$

Since this has to hold for all values of h, and α and τ are independent of h, we get $\alpha \leq \frac{1}{d_{\text{in}}(\mathsf{A}_1)}$.

Combining the argument presented above with the result in Appendix C.1, we get

$$\alpha \leq \frac{1}{d_{\max}}. \tag{4.39}$$

This completes the proof of the converse.

■

4.5.3 $3 \times 3 \times 3$ Networks

We then consider two-layer networks with three source-destination pairs and three relays. Interestingly, we face networks in which we need to incorporate network coding techniques in order to achieve the normalized sum-capacity with local view. The coding comes in the form of repetition coding at sources and a combination of repetition and network coding at relays.

Definition 4.7 *A* $3\times3\times3$ *network is a two-layer network (as defined in Section 4.2.1) with* $K = 3$ *and* $|\mathcal{V}| - 2K = 3$.

Theorem 4.3 *The normalized sum-capacity of a* $3 \times 3 \times 3$ *network with local view,* α^*, *is equal to*

(a) 1 if and only if $\mathcal{G}_{ii} \cap \mathcal{G}_{jj} = \emptyset$ *for* $i \neq j$;

(b) 1/3 if and only if *one of the graphs in Fig. 4.7 is a subgraph of the network connectivity graph* $\mathcal{G}$;

(c) 1/2 *otherwise.*

As we show in this section, the transmission strategy is a combination of three main techniques:

(a) per layer interference avoidance;

(b) repetition coding to allow overhearing of the interference;

(c) network coding to allow interference neutralization.

Remark 4.5 *From Theorem 4.2 and Theorem 4.3, we conclude that for all single-layer networks (not the main focus of this work),* $K\times2\times K$, *and* $3\times3\times3$ *networks with* 1*-local view, the normalized sum-capacity* $\alpha^* = 1/K$ (i.e. *TDMA is optimal),* if and only if *when all channel gains are equal and non-zero, then there exists a node that can decode* all *messages. We refer to such node as an "omniscient" node. We conjecture that this observation holds for much more general network connectivities with local view or in fading networks with no channel state information at the transmitters. However, this is a different line of research and it is beyond the scope of this paper.*

Proof:

Achievability: The achievability proof for networks in category (a) is trivial as there is no interference present in such networks. For networks in category (b), a simple TDMA achieves a normalized sum-rate of 1/3. Thus, we only need to prove the result for networks in category (c).

Suppose none of the graphs in Fig. 4.7 is a subgraph of the network connectivity graph $\mathcal{G}$ and the network does not fall in category (a). This immediately implies that

$$d_{\max} = 2. \tag{4.40}$$

Definition 4.8 *The first layer conflict graph of a* $3 \times 3 \times 3$ *network has three vertices each corresponding to a source. Two vertices are connected by an undirected edge if and only if the corresponding sources are both connected to a relay. Similarly, the second layer conflict graph of a* $3 \times 3 \times 3$ *network has three vertices each corresponding to a destination. Two vertices are connected by an undirected edge if and only if the corresponding destinations are both connected to a relay.*

We have the following claim for a $3 \times 3 \times 3$ network with $d_{\max} = 2$.

Claim 4.1 *For a* $3 \times 3 \times 3$ *network with* $d_{\max} = 2$*, the only connectivity that results in a per layer fully connected conflict graph, is the one shown in Fig. 4.8.*

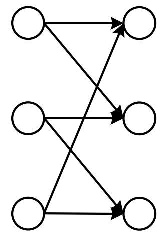


Figure 4.8: *The only connectivity that results in a per layer fully connected conflict graph in a* $3 \times 3 \times 3$ *network with* $d_{\max} = 2$.

Proof of Claim 4.1 is straightforward and can be obtained by contradiction.

If neither of the layer conflict graphs is fully connected and the network does not fall in category (a), then a normalized sum-rate of $\alpha = 1/2$ is easily achievable using per layer interference avoidance. Moreover from Claim 4.1, we know that with $d_{\max} = 2$, the folded-chain structure of Fig. 4.8 exists in at least one of the layers and in Section 4.3, we showed that a normalized sum-rate of $\alpha = 1/2$ is achievable.

We note that these cases can be easily described within the algebraic framework of Section 4.4. In fact, if the folded-chain structure of Fig. 4.8 exists in at least one of the layers, as shown in Section 4.3, the transmission can be expressed as a valid assignment of transmit and receive matrices; and if the per layer conflict graphs are not fully connected and the network does not fall in category (a), then the scheme is a per layer interference avoidance which can be easily expressed in terms of a valid assignment of transmit and receive matrices.

Converse: The forward direction of the proof for networks in category (a) is trivial as there is no interference present in such networks. For the reverse direction, as shown in Lemma 4.2 below, for any network that does not fall into category (a), an upper bound of 1/2 on the normalized sum-capacity can be easily obtained. In fact, Lemma 4.2 also provides the outer-bound for networks in category (c). Thus, we only need to consider $3 \times 3 \times 3$ networks in category (b).

Lemma 4.2 *In a $3 \times 3 \times 3$ network with local view, if there exists a path from S_i to D_j, for some $i \neq j$, then the normalized sum-capacity is upper-bounded by $\alpha = 1/2$.*

Proof of Lemma 4.2 is presented in Appendix C.3.

For $3 \times 3 \times 3$ networks in category (b), we first consider the forward direction. One of the graphs in Fig. 4.7 is a subgraph of the network connectivity graph $\mathcal{G}$, say the graph in Fig. 4.7(b). Assign channel gain of $h \in \mathbb{C}$ to the links of the subgraph, and channel gain of 0 to the links that are not in the graph of Fig. 4.7(b).

With this assignment of the channel gains, it is straightforward to see that

$$H\left(W_i | Y^n_{\mathsf{A}_i}, L_{\mathsf{A}_i}, \mathsf{SI}\right) \leq n\epsilon_n, \qquad i = 1, 2, 3. \tag{4.41}$$

Basically, each destination D_i is *only* connected to relay A_i, and each relay A_i has all the information that destination D_i requires in order to decode its message, $i = 1, 2, 3$.

Thus, Relay A_1 can decode W_1. After removing the contribution of S_1, relay A_1 is able to decode W_2. Continuing this argument, we conclude that

$$n\left(\sum_{j=1}^{3} R_j - \epsilon_n\right) \leq h\left(Y^n_{\mathsf{A}_1} | L_{\mathsf{A}_1}, \mathsf{SI}\right). \tag{4.42}$$

The MAC capacity at relay A_1, gives us

$$3(\alpha \log(1 + |h|^2) - \tau) \leq \log(1 + 3 \times |h|^2), \tag{4.43}$$

which results in

$$(3\alpha - 1)\log(1 + |h|^2) \leq \log(3) + 3\tau. \tag{4.44}$$

Since this has to hold for all values of h, and α and τ are independent of h, we get $\alpha \leq \frac{1}{3}$. The proof for the graphs in Fig. 4.7(a) and Fig. 4.7(c) is very similar.

For the reverse direction, as shown above, if none of the graphs in Fig. 4.7 is a subgraph of the network connectivity graph $\mathcal{G}$, a normalized sum-rate of $\alpha = \frac{1}{2}$ is achievable. This completes the converse proof. ■

4.5.4 Folded-Chain Networks

We now consider a class of networks for which we need to incorporate network coding in order to achieve the normalized sum-capacity with local view.

Definition 4.9 *A two-layer (K, m) folded-chain network ($1 \leq m \leq K$) is a two-layer network with K S-D pairs and K relays in the middle. Each S-D pair i has m disjoint paths, through relays with indices $1 + [\{(i-1)^+ + (j-1)\}$ **mod** $K]$ where $i = 1, \ldots, K$, $j = 1, \ldots, m$.*

Theorem 4.4 *The normalized sum-capacity of a two-layer (K, m) folded-chain network with local view is $\alpha^* = \frac{1}{m}$.*

Remark 4.6 *In order to achieve the normalized sum-capacity of a two-layer (K, m) folded-chain network with local view, we need to incorporate repetition coding at sources and network coding at relays. Also, we note that a two-layer (K, K) folded-chain network is basically the fully-connected structure and in that case with local view, interference avoidance achieves the normalized sum-capacity of $1/K$.*

Proof:

Achievability: The result is trivial for $m = 1$. For $m = K$, the upper bound of $\alpha = \frac{1}{K}$ can be achieved simply by using TDMA. Suppose $m < K < 2m$ (we will later generalize the achievability scheme for arbitrary K). Let $m' = K - m$. To each source S_i, $i = 1, \ldots, K$, we assign a transmit vector $\mathbf{T}_{\mathsf{S}_i}$ as follows.

$$\mathbf{T}_{\mathsf{S}_i}(j) = 1 \Leftrightarrow j \leq i \leq j + m', \quad j = 1, \ldots, m. \tag{4.45}$$

It is straight forward to verify that this assignment satisfies conditions **C.1** and **C.2**:

- **C.1** is trivially satisfied since for any source S_i, there exists at least one value of j such that $\mathbf{T}_{\mathsf{S}_i}(j) = 1$.
- **C.2** is satisfied since for $1 \leq i \leq m$, $\mathbf{R}_{\mathsf{V}_i}$ has a single 1 in row i and the column labeled as S_i; and for $m < i < 2m$, the matrix is in is in echelon form.

For the second layer, we have $K + 1$ steps.

• Steps 1 through m: During step i, $1 \leq i \leq m$, our goal is to provide destination D_i with its desired signal without any interference. Therefore, row i of any relay connected to D_i has a single 1 in the column associated with S-D pair i and 0's elsewhere;

• Steps $m + 1$ through K: During step i, $m + 1 \leq i \leq K$, our goal is to provide destination D_i with its desired signal (interference will be handled later). To do so, consider the transmit matrix of any relay connected to D_i; we place a 1 in the column associated with S-D pair i and the row with the least number of 1's and

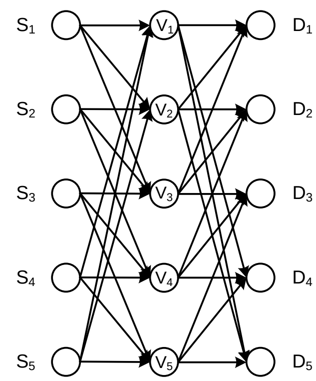


Figure 4.9: *A* $(5, 3)$ *two-layer folded-chain network.*

the smallest index (we first find the rows with the least number of 1's and then among these rows, we pick the row that has the smallest index);

• Step $K+1$: During this step our goal is to resolve interference and goes through the following loop:

(a) Let $\mathcal{L}_j$ denote the set of row indices for which there exists at least a 1 in the column associated with S-D pair j of the transmit matrix of a relay connected to D_j, $m+1 \leq j \leq K$;

(b) Set $j = m+1$. For any $j' \neq j$, $j' \in \{1, 2, \ldots, K\}$, if

$$\sum_{\mathsf{V}_p \text{ connected to } \mathsf{D}_j} \sum_{\ell \in \mathcal{L}_j} \mathbf{T}_{\mathsf{V}_p}\left(\ell, \mathsf{S}_{j'}\right) \neq 0, \tag{4.47}$$

then make this summation 0 by making any of the $\mathbf{T}_{\mathsf{V}_p}\left(\ell, \mathsf{S}_{j'}\right)$'s that is not previously assigned, equal to 1 or -1 as needed[3]. Note that since each route contributes at most once to any receive matrix, the summation in (4.47) can only take values in $\{-1, 0, 1\}$;

[3]If $\mathbf{T}_{\mathsf{V}_p}\left(\ell, \mathsf{S}_{j'}\right)$ is not yet assigned a value, treat it as 0 in the summation.

$$\begin{array}{llll}
\text{End of Step:} & \quad m & \quad K & \quad K+1
\end{array}$$

$$\mathbf{T}_{\mathsf{V}_1}: \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ - & - & - \\ - & - & - \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ - & 1 & - \\ - & - & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{T}_{\mathsf{V}_2}: \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ - & - & - \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{T}_{\mathsf{V}_3}: \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_1 & \mathsf{S}_2 & \mathsf{S}_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{T}_{\mathsf{V}_4}: \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_2 & \mathsf{S}_3 & \mathsf{S}_4 \end{array} \\ \begin{pmatrix} - & - & - \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_2 & \mathsf{S}_3 & \mathsf{S}_4 \end{array} \\ \begin{pmatrix} - & - & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_2 & \mathsf{S}_3 & \mathsf{S}_4 \end{array} \\ \begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

$$\mathbf{T}_{\mathsf{V}_5}: \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_3 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} - & - & - \\ - & - & - \\ 1 & 0 & 0 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_3 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} - & 1 & - \\ - & - & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array} \quad \begin{array}{c} \begin{array}{ccc} \mathsf{S}_3 & \mathsf{S}_4 & \mathsf{S}_5 \end{array} \\ \begin{pmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 4.10: *The evolution of the relays' transmit matrices at the end of steps m, K, and $K + 1$. In this example, the loop in Step $K + 1$ is repeated three times.*

(c) Set $j = j+1$; If $j > K$ and the during previous loop no change has occurred, then set all entries that are not yet defined equal to zero and terminate; otherwise, go to line 2 of the loop.

$$
\begin{array}{ccccccccc}
\mathsf{S}_1:\mathsf{V}_1 & \mathsf{S}_2:\mathsf{V}_4 & \mathsf{S}_3:\mathsf{V}_4 & \mathsf{S}_3:\mathsf{V}_5 & \mathsf{S}_4:\mathsf{V}_1 & \mathsf{S}_4:\mathsf{V}_4 & \mathsf{S}_4:\mathsf{V}_5 & \mathsf{S}_5:\mathsf{V}_1 & \mathsf{S}_5:\mathsf{V}_5
\end{array}
$$
$$
\begin{pmatrix}
1 & -1 & 0 & 1 & 0 & 1 & 1 & 0 & -1 \\
-1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}
\tag{4.46}
$$

We need to show that the assignment of transmit and receive matrices described above satisfies Conditions **C.1**, **C.2** and **C.3**. Conditions **C.1** and **C.2** are satisfied as shown before. Due to the construction described above, any interfering signal will be provided to the receive node such that it can be canceled out later. This is guaranteed by making sure the summation in (4.47) remains equal to zero. As a result, Condition **C.3** is also satisfied. Thus, we only need to show that the algorithm terminates in finite time. Note that each destination is only served once via a relay that is connected to it and if that signal is retransmitted, it is for interference neutralization purposes. Thus, if the summation in (4.47) is not zero, then there exists an unassigned entry such that the summation can be made zero. Since there are a total of mK^2 entries, the algorithm terminates in finite time. This completes the description of transmit and receive matrices. Using Theorem 4.1, we know that a normalized sum-rate of $\alpha = \frac{1}{m}$ is achievable.

For general K the achievability works as follows. Suppose, $K = c(2m-1)+r$, where $c \geq 1$ and $0 \leq r < (2m-1)$, we implement the scheme for S-D pairs $1, 2, \ldots, 2m-1$ as if they are the only pairs in the network. The same for source-destination pairs $2m, 2m+1, \ldots, 4m-2$, etc. Finally, for the last r S-D pairs, we implement the scheme with $m' = \max\{r-m+1, 1\}$. This completes the proof of achievability.

We now describe the $K+1$ steps via an example of $(5,3)$ two-layer folded-chain network of Fig. 4.9. In Fig. 4.10, we have demonstrated the evolution of the relays' transmit matrices at the end of steps m, K, and $K+1$. For this example, the loop in Step $K+1$ is repeated three times.

It is easy to verify Condition **C.3** for destination D_1, D_2 and D_3. We have provided $\mathbf{R}_{\mathsf{D}_4}$ in (4.46), and as we can see by adding the first and the second row, we can have a row that has only 1's in the columns corresponding to source S_4. Similarly, we can show that the condition holds for D_5.

Converse: Assume that a normalized sum-rate of α is achievable, *i.e.* there exists a transmission strategy with local view, such that for all channel realizations, it achieves a sum-rate satisfying

$$\sum_{i=1}^{K} R_i \gamma \alpha C_{\text{sum}} - \tau, \tag{4.48}$$

with error probabilities going to zero as $N \to \infty$ and for some constant $\tau \in R$ independent of the channel gains.

Consider the first layer of the two-layer (K,m) folded-chain network, where the channel gain of a link from source i to relays $i, i+1, \ldots, m$ is equal to $h \in \mathbb{C}$, $i = 1, 2, \ldots, m$, and all the other channel gains are equal to zero. In the second layer, we set the channel gain from relay i to destination i equal to h, and all other channel gains equal to 0, $i = 1, 2, \ldots, m$. See Figure 4.11 for a depiction.

With this configuration, each destination D_i is *only* connected to relay A_i, and each relay A_i has all the information that destination D_i requires in order to decode its message, $i = 1, \ldots, m$. We have

$$H\left(W_i | Y_{\mathsf{A}_i}^n, L_{\mathsf{A}_i}, \mathsf{SI}\right) \leq n\epsilon_n, \qquad i = 1, \ldots, m. \tag{4.49}$$

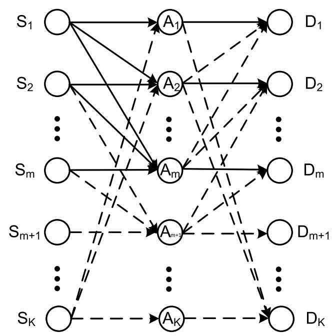


Figure 4.11: *Channel gain assignment in a two-layer* (K, m) *folded-chain network. All solid links have capacity* h*, and all dashed links have capacity* 0*.*

At relay A_m after decoding and removing the contribution of S_m, relay A_m is able to decode W_{m-1}. With a recursive argument, we conclude that

$$n\left(\sum_{j=1,2,\ldots,m} R_j - \epsilon_n\right) \leq h\left(Y^n_{\mathsf{A}_m} | L_{\mathsf{A}_m}, \mathsf{SI}\right). \tag{4.50}$$

The MAC capacity at relay A_m, gives us

$$m(\alpha \log(1 + |h|^2) - \tau) \leq \log(1 + m \times |h|^2), \tag{4.51}$$

which results in

$$(m\alpha - 1)\log(1 + |h|^2) \leq \log(m) + m\tau. \tag{4.52}$$

Since this has to hold for all values of h, and α and τ are independent of h, we get $\alpha \leq \frac{1}{m}$. ■

4.5.5 Gain of Coding over Interference Avoidance: Nested Folded-Chain Networks

In this subsection, we show that the gain from using coding over interference avoidance techniques can be unbounded. To do so, we first define the following class of networks.

Definition 4.10 *An L-nested folded-chain network is a single-layer network with $K = 3^L$ S-D pairs, $\{\mathsf{S}_1, \ldots, \mathsf{S}_{3^L}\}$ and $\{\mathsf{D}_1, \ldots, \mathsf{D}_{3^L}\}$. For $L = 1$, an L-nested folded-chain network is the same as a single-layer $(3, 2)$ folded-chain network. For $L > 1$, an L-nested folded-chain network is formed by first creating 3 copies of an $(L-1)$-nested folded-chain network. Then,*

- *The i-th source in the first copy is connected to the i-th destination in the second copy, $i = 1, \ldots, 3^{L-1}$,*
- *The i-th source in the second copy is connected to the i-th destination in the third copy, $i = 1, \ldots, 3^{L-1}$,*
- *The i-th source in the third copy is connected to the i-th destination in the first copy, $i = 1, \ldots, 3^{L-1}$.*

Fig. 4.12 illustrates a 2-nested folded-chain network.

Consider an L-nested folded-chain network. The conflict graph of this network is fully connected, and as a result, interference avoidance techniques can only achieve a normalized sum-rate of $\left(\frac{1}{3}\right)^L$. However, we know that for a single-layer $(3, 2)$ folded-chain network, a normalized sum-rate of $\frac{1}{2}$ is achievable.

Hence, applying our scheme to an L-nested folded-chain network, a normalized sum-rate of $\left(\frac{1}{2}\right)^L$ is achievable. For instance, consider the 2-nested folded-chain network in Fig. 4.12. We show that any transmission strategy over the induced subgraphs can be implemented in the original network by using only four time-slots, such that all nodes receive the same signal as if they were in the induced subgraphs.

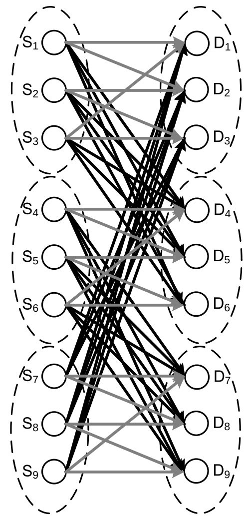
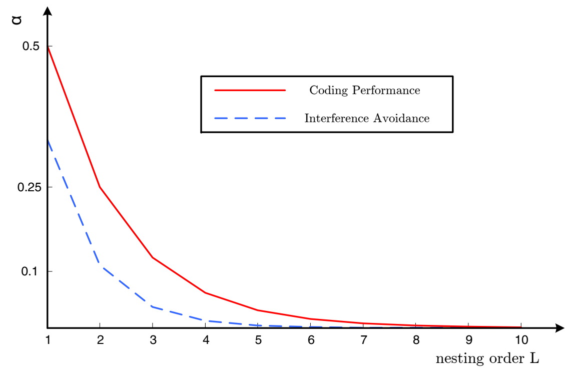


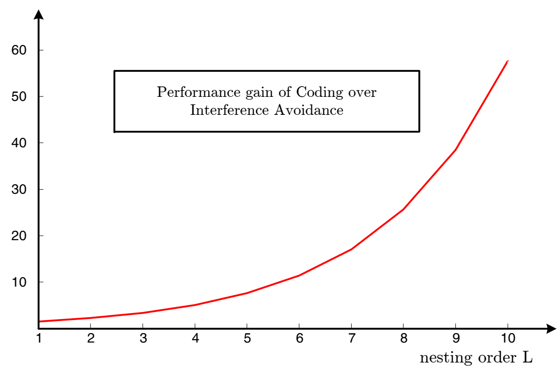
Figure 4.12: *A 2-nested folded-chain network.*

To achieve a normalized sum-rate of $\alpha = \left(\frac{1}{2}\right)^2$, we split the communication block into 4 time-slots of equal length. During time-slot 1, sources 1, 2, 4, and 5 transmit the same codewords as if they are in the induced subgraphs. During time-slot 2, sources 3 and 6 transmit the same codewords as if they are in the induced subgraphs and sources 2 and 5 repeat their transmit signal from the first time-slot. During time-slot 3, sources 7 and 8 transmit the same codewords as if they are in the induced subgraphs and sources 4 and 5 repeat their transmit signal from the first time-slot. During time-slot 4, source 9 transmits the same codewords as if it is in the induced subgraph and sources 5, 6, and 8 repeat their transmit signal.

It is straight forward to verify that with this scheme, all destinations receive the same signal as if they were in the induced subgraphs. Hence, a normalized sum-rate of $\alpha = \left(\frac{1}{2}\right)^2$ is achievable for the network in Figure 4.12. Therefore, the gain of using coding over interference avoidance is $\left(\frac{3}{2}\right)^L$ which goes to infinity as $L \to \infty$. See Fig.4.13 for a depiction. As a result, we can state the following



(a)



(b)

Figure 4.13: *(a) Achievable normalized sum-rate of inter-session coding and interference avoidance in an L-nested folded-chain network; and (b) the performance gain of inter-session coding scheme over interference avoidance.*

lemma.

Lemma 4.3 *Consider an L-nested folded-chain network. The gain of using MCL scheduling over MIL scheduling is* $\left(\frac{3}{2}\right)^L$ *which goes to infinity as* $L \to \infty$.

The scheme required to achieve a normalized sum-rate of $\alpha = \left(\frac{1}{2}\right)^L$ for an L-nested folded-chain network, can be viewed as a simple extension of the results presented in Section 4.4. In a sense instead of reconstructing a single snapshot, we reconstruct L snapshots of the network. The following discussion is just for the completion of the results.

To each source S_i, we assign a transmit vector $\mathbf{T}_{\mathsf{S}_i}$ of size $(LT) \times 1$ where each row corresponds to a time-slot. If we denote the transmit signal of node S_i in the ℓ^{th} snapshot by $X^{\ell}_{\mathsf{S}_i}$, then if $\mathbf{T}_{\mathsf{S}_i}(j) = 1$ and $cT + 1 \leq j < (c+1)T$ for $c = 0, 1, \ldots, L-1$, then S_i communicates $X^{c+1}_{\mathsf{S}_i}$. The other transmit and receive matrices can be described similarly. Conditions **C.2** and **C.3** have to be satisfies for submatrices of $\mathbf{R}_{\mathsf{V}_j}$ corresponding to rows $cT+1, cT+2, \ldots, (c+1)T - 1$ for $c = 0, 1, \ldots, L-1$.

4.6 Concluding Remarks

In this paper, we studied the fundamental limits of communications over two-layer wireless networks where each node has only limited knowledge of the channel state information. We proposed an algebraic framework for inter-session coding in such networks. We developed new transmission strategies based on the algebraic framework that combines multiple ideas including interference avoidance and network coding. We established the optimality of our proposed strategy for several classes of networks in terms of achieving the normalized sum-capacity. We also demonstrated several connections between network topology, normalized sum-capacity, and the achievability strategies.

One major direction is to characterize the increase in normalized sum-capacity as nodes learn more and more about the channel state information. We have also focused on the case that the nodes know the network-connectivity globally, but the actual values of the channel gains are only known for a subset of flows. Another important direction would be to understand the effects of local knowledge about network connectivity on the capacity and develop distributed strategies to optimally route information with partial knowledge about network connectivity.

CHAPTER 5

COMMUNICATION WITH RATE-LIMITED FEEDBACK LINKS

5.1 Introduction

The history of feedback in communication systems traces back to Shannon. It is well-known that feedback does not increase the capacity of discrete memoryless point-to-point channels. However, feedback can enlarge the capacity region of multi-user networks, even in the most basic case of the two-user memoryless multiple-access channel [20, 39]. Hence, there has been a growing interest in developing feedback strategies and understanding the fundamental limits of communication over multi-user networks with feedback, in particular the two-user interference channel (IC). See [23, 31–33, 45, 50, 53] for example.

Especially in [50], the infinite-rate feedback capacity of the two-user Gaussian IC has been characterized to within a 2-bit gap. One consequence of this result is that interestingly feedback can provide an unbounded capacity increase. This is in contrast to point-to-point and multiple-access channels where feedback provides no gain and bounded gain respectively.

While the feedback links are assumed to have *infinite* capacity in [50], a more realistic feedback model is one where feedback links are *rate-limited*. In this paper, we study the impact of the rate-limited feedback in the context of the two-user IC. We focus on two fundamental questions: (1) what is the maximum capacity gain that can be obtained with access to feedback links at a specific rate of C_{FB}? (2) what are the transmission strategies that exploit the available feedback links efficiently? Specifically, we address these questions under three channel models: the El Gamal-Costa deterministic model [16], the linear deterministic model of [7], and the Gaussian model.

Under the El Gamal-Costa deterministic model, we derive inner-bounds and outer-bounds on the capacity region. As a result, we show that the capacity region can be enlarged using feedback by at most the amount of available feedback, *i.e.*, "one bit of feedback is at most worth one bit of capacity". Our achievable scheme employs three techniques: (1) Han-Kobayashi message splitting; (2) quantize-and-binning; and (3) decode-and-forward. Unlike the infinite-rate feedback case [50], in the rate-limited feedback case, a receiver cannot provide its *exact* received signal to its corresponding transmitter; therefore, the main challenge is how to smartly decide what to send back through the available rate-limited feedback links. We overcome this challenge as follows. We first split each transmitter's message into three parts: the cooperative common, the non-cooperative common, and the private message. Next, each receiver quantizes its received signal and then generates a binning index so as to capture part of the other user's common message (that we call the cooperative common message) which causes interference to its own message. The receiver will then send back this binning index to its intended transmitter through the rate-limited feedback links. With this feedback, each transmitter decodes the other user's cooperative common message by exploiting its own message as side information. This way transmitters will be able to cooperate by means of the feedback links, thereby enhancing the achievable rates. This result will be described in Section 5.4.

We then study the problem under the linear deterministic model [7] which captures the key properties of the wireless channel, and thus provides insights that can lead to an approximate capacity of Gaussian networks [7, 11, 12, 46, 50].

We show that our inner-bounds and outer-bounds match under this linear deterministic model, thus establishing the capacity region. While this model is a special case of the El Gamal-Costa model, it has a significant role to play in motivating a generic achievable scheme for the El Gamal-Costa model. Moreover, the explicit achievable scheme in this model provides a concrete guideline to the Gaussian case. We will explain this result in Section 5.5.

Inspired by the results in the deterministic models, we develop an achievable scheme and also derive new outer-bounds for the Gaussian channel. In order to translate the main ideas in our achievability strategy for the deterministic models into the Gaussian case, we employ lattice coding which enables receivers to decode superposition of codewords. Specifically at each transmitter, we employ lattice codes for cooperative messages. By appropriate power assignment of the codewords, we make the desired lattice codes arrive at the same power level, hence, receivers being able to decode the superposition of codewords. Each receiver will then decode the index of the lattice code corresponding to the superposition and sends it back to its corresponding transmitter where the cooperative common message of the other user will be decoded. For symmetric channel gains, we show that the gap between the achievable sum-rate and the outer-bounds can be bounded by a constant, independent of the channel gains. This will be explained in Section 5.6.

5.2 Problem Formulation and Network Model

We consider a two-user interference channel (IC) where a noiseless rate-limited feedback link is available from each receiver to its corresponding transmitter.

See Figure 5.1. The feedback link from receiver k to transmitter k is assumed to have a capacity of C_{FBk}, $k = 1, 2$. on.

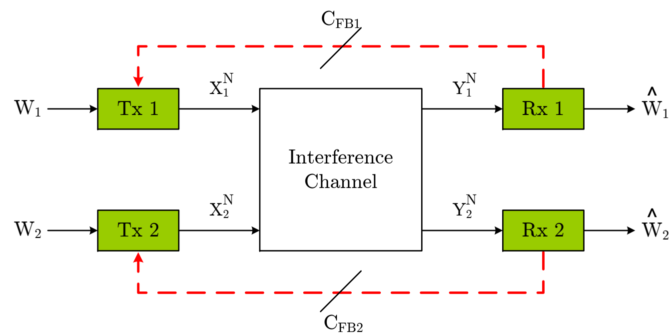


Figure 5.1: Two-user interference channel with rate-limited feedback.

Transmitters 1 and 2 wish to reliably communicate independent and uniformly distributed messages $W_1 \in \{1, 2, \ldots, 2^{NR_1}\}$ and $W_2 \in \{1, 2, \ldots, 2^{NR_2}\}$ to receivers 1 and 2 respectively, during N uses of the channel. The transmitted signal of transmitter k, $k = 1, 2$, at time i, $1 \leq i \leq N$, and the received signal of receiver k, $k = 1, 2$, at time i, $1 \leq i \leq N$, are respectively denoted by $X_{k,i}$ and $Y_{k,i}$. There are two feedback encoders at the receivers that create the feedback signals from the received signals:

$$\tilde{Y}_{k,i} = \tilde{e}_{k,i}(Y_{k,1}, \ldots, Y_{k,i-1}) = \tilde{e}_{k,i}(Y_k^{(i-1)}), \quad k = 1, 2. \tag{5.1}$$

where we use shorthand notation to indicate the sequence up to $i - 1$.

Due to the presence of feedback, the encoded signal $X_{k,i}$ of user k at time i is a function of both its own message and previous outputs of the corresponding feedback encoder:

$$X_{k,i} = e_{k,i}(W_k, \tilde{Y}_k^{(i-1)}), \qquad k = 1, 2. \tag{5.2}$$

Each receiver k, $k = 1, 2$, uses a decoding function $d_{k,N}$ to get the estimate $\hat{W}_k$, from the channel outputs $\{Y_{k,i} : i = 1, \ldots, N\}$. An error occurs whenever $\hat{W}_k \neq W_k$.

The average probability of error is given by

$$\lambda_{k,N} = \mathbb{E}[P(\hat{W}_k \neq W_k)], \qquad k = 1, 2, \tag{5.3}$$

where the expectation is taken with respect to the random choice of the transmitted messages W_1 and W_2.

We say that a rate pair (R_1, R_2) is achievable, if there exists a block encoder at each transmitter, a block encoder at each receiver that creates the feedback signals, and a block decoder at each receiver as described above, such that the average error probability of decoding the desired message at each receiver goes to zero as the block length N goes to infinity. The capacity region C is the closure of the set of the achievable rate pairs.

We will consider the following three channel models to investigate this problem.

1- El Gamal-Costa Deterministic Interference Channel:

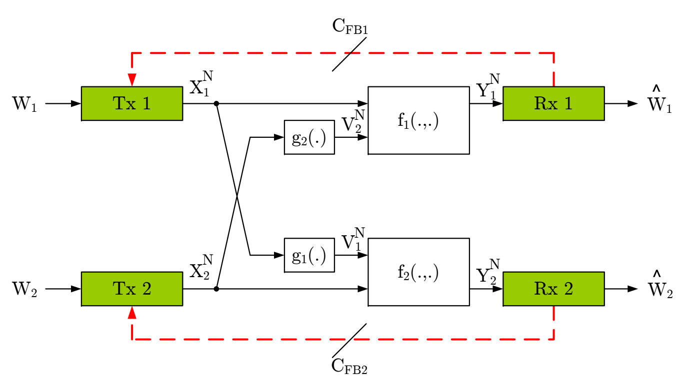


Figure 5.2: The El Gamal-Costa deterministic IC with rate-limited feedback.

Figure 5.2 illustrates the El Gamal-Costa deterministic IC [16] with rate-limited feedback. In this model the outputs Y_1 and Y_2 and the interferences

V_1 and V_2 are (deterministic) functions of inputs X_1 and X_2 [16]:

$$\begin{aligned} Y_{1,i} &= f_1(X_{1,i}, V_{2,i}), \\ Y_{2,i} &= f_2(X_{2,i}, V_{1,i}), \\ V_{1,i} &= g_1(X_{1,i}), \\ V_{2,i} &= g_2(X_{2,i}), \end{aligned} \tag{5.4}$$

where $f_1(.,.)$ and $f_2(.,.)$ are such that

$$\begin{aligned} H(V_{2,i}|Y_{1,i}, X_{1,i}) = 0, \\ H(V_{1,i}|Y_{2,i}, X_{2,i}) = 0. \end{aligned} \tag{5.5}$$

Here V_k is a part of X_k ($k = 1, 2$), visible to the unintended receiver. This implies that in any system where each decoder can decode its message with arbitrary small error probability, V_1 and V_2 are completely determined at receivers 2 and 1, respectively, *i.e.*, these are common signals.

2- Linear Deterministic Interference Channel:

This model, which was introduced in [7], captures the effect of broadcast and superposition in wireless networks. We study this model to bridge from general deterministic networks into Gaussian networks. In this model, there is a non-negative integer n_{kj} representing channel gain from transmitter k to receiver j, $k = 1, 2$, and $j = 1, 2$. In the linear deterministic IC, we can write the channel input to the transmitter k at time i as $X_{k,i} = [X_{k,i}^1 \; X_{k,i}^2 \ldots X_{k,i}^q]^T \in \mathbb{F}_2^q$, $k = 1, 2$, such that $X_{k,i}^1$ and $X_{k,i}^q$ represent the most and the least significant bits of the transmitted signal respectively. Also, q is the maximum of the channel gains in the network, *i.e.*, $q = \max_{k,j}\left(n_{kj}\right)$. At each time i, the received signals are given by

$$
\begin{aligned}
Y_{1,i} &= \mathbf{S}^{q-n_{11}} X_{1,i} \oplus \mathbf{S}^{q-n_{21}} X_{2,i}, \\
Y_{2,i} &= \mathbf{S}^{q-n_{12}} X_{1,i} \oplus \mathbf{S}^{q-n_{22}} X_{2,i},
\end{aligned}
\tag{5.6}
$$

where $\mathbf{S}$ is the $q \times q$ shift matrix and operations are performed in $\mathbb{F}_2$ (*i.e.*, modulo two). See Figure 5.3 for an example.

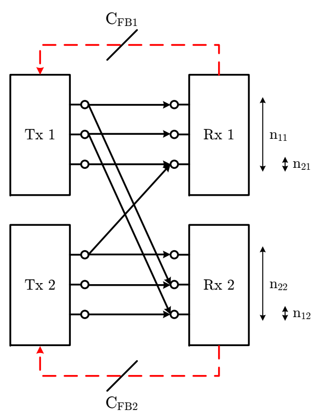


Figure 5.3: An example of a linear deterministic IC with rate-limited feedback, where $n_{11} = n_{22} = 3$, $n_{12} = 2$, $n_{21} = 1$, and $q = 3$.

It is easy to see that this model also satisfies the conditions of (5.5), hence it is a special class of the El Gamal-Costa deterministic IC.

3- Gaussian Interference Channel:

In this model, there is a complex number h_{kj} representing the channel from transmitter k to receiver j, $k = 1, 2$, and $j = 1, 2$. The received signals are

$$
\begin{aligned}
Y_{1,i} &= h_{11} X_{1,i} + h_{21} X_{2,i} + Z_{1,i}, \\
Y_{2,i} &= h_{12} X_{1,i} + h_{22} X_{2,i} + Z_{2,i},
\end{aligned}
\tag{5.7}
$$

where $\{Z_{j,i}\}_{i=1}^{N}$ is the additive white complex Gaussian noise process with zero mean and unit variance at receiver j, $j = 1, 2$. Without loss of generality, we assume a power constraint of 1 at all nodes, *i.e.*,

$$\frac{1}{N}\mathbb{E}(\sum_{i=1}^{N}|X_{k,i}|^2) \leq 1 \qquad k = 1, 2, \tag{5.8}$$

where N is the block length. We will use the following notations:

$$\begin{aligned} \mathsf{SNR}_1 &= |h_{11}|^2, \quad \mathsf{SNR}_2 = |h_{22}|^2, \\ \mathsf{INR}_{12} &= |h_{12}|^2, \quad \mathsf{INR}_{21} = |h_{21}|^2. \end{aligned} \tag{5.9}$$

5.3 Motivating Example

We start by analyzing a motivating example. Consider the linear deterministic IC with rate-limited feedback as depicted in Figure 5.4(a). As we will see in Section 5.5, the capacity region of this network is given by the region shown in Figure 5.4(b). Our goal in this section is to demonstrate how feedback can help increase the capacity. In particular, we describe the achievability strategy for one of the corner points, *i.e.*, $(R_1, R_2) = (4, 1)$. From this example, we will make important observations that will later provide insights into the achievable scheme.

The achievability strategy works as follows. In the first time slot, transmitter 1 sends four bits $a_1, \dots, a_4$ and transmitter 2 sends only one bit b_1 at the third level, see Figure 5.4(a). This way receiver 1 can decode its intended four bits interference free, while receiver 2 has access to only $a_1 \oplus b_1$ and a_2. In the second time slot, through the feedback link, receiver 2 feeds $a_1 \oplus b_1$ back to transmitter 2 who can remove b_1 from it to decode a_1. Also during the second time slot, transmitter 1 sends four fresh bits $a_5, \dots, a_8$, whereas transmitter 2 sends one new bit b_2. In the third time slot, through the feedback link, receiver 2 feeds $b_2 \oplus a_5$ back to transmitter 2 who can remove b_2 from it to decode a_5. Moreover,

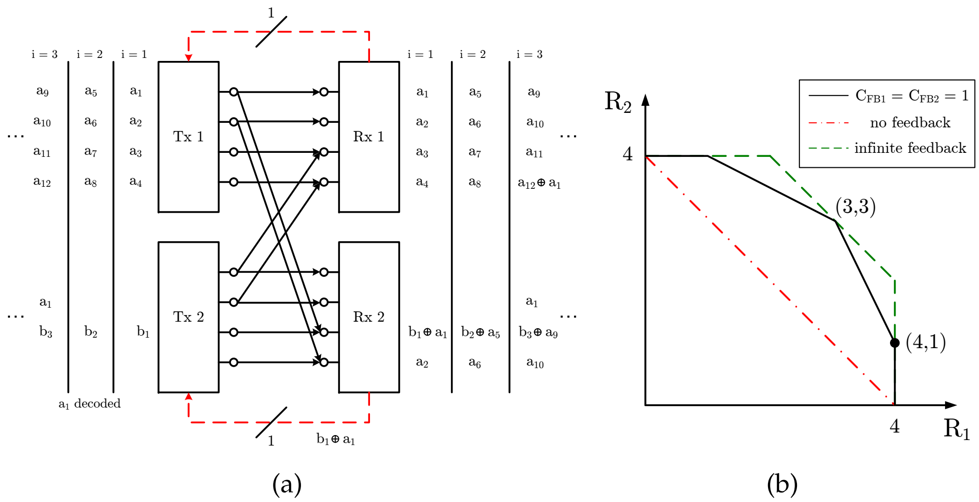


Figure 5.4: *(a) A two-user linear deterministic IC with channel gains $n_{11} = n_{22} = 4$, $n_{12} = n_{21} = 2$ and feedback rates $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = 1$, and (b) its capacity region.*

during the third time slot, transmitter 1 sends four new bits $a_9, \ldots, a_{12}$, while transmitter 2 sends one new bit b_3 and at the level shown in Figure 5.4(a), sends the other user's information bit a_1 decoded in the second time slot with the help of feedback. With this strategy receiver 2 has now access to a_1 and can use it to decode b_1. Note that receiver 1 already knows a_1 and hence can decode a_{12}. This procedure will be repeated over the next time slots. During the last two time slots, only transmitter 2 sends the other user's information decoded before, while transmitter 1 sends nothing. Therefore, after B time slots, we achieve a rate of $(R_1, R_2) = \frac{B-2}{B}(4, 1)$, which converges to $(4, 1)$ as B goes to infinity.

Based on this simple capacity-achieving strategy, we can now make several observations:

- The messages coming from transmitter 1 can be split into three parts: (1) "cooperative common": this message is visible to both receivers, while interfering with the other user's signals (e.g., a_1 at transmitter 1 in the first time slot).

$$
\begin{aligned}
R_1 &\leq I(U, V_2, X_1; Y_1) && \text{(5.10a)}\\
R_1 &\leq I(X_1; Y_1|U, U_1, V_2) + \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) && \text{(5.10b)}\\
R_2 &\leq I(U, V_1, X_2; Y_2) && \text{(5.10c)}\\
R_2 &\leq I(X_2; Y_2|U, U_2, V_1) + \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) && \text{(5.10d)}\\
R_1 + R_2 &\leq I(X_1; Y_1|U, V_1, V_2) + I(U, V_1, X_2; Y_2) && \text{(5.10e)}\\
R_1 + R_2 &\leq I(X_2; Y_2|U, V_1, V_2) + I(U, V_2, X_1; Y_1) && \text{(5.10f)}\\
R_1 + R_2 &\leq \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) + \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) &&\\
&\quad + I(X_1, V_2; Y_1|U, U_1, U_2) + I(X_2; Y_2|U, V_1, V_2) && \text{(5.10g)}\\
R_1 + R_2 &\leq \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) + \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) &&\\
&\quad + I(X_2, V_1; Y_2|U, U_1, U_2) + I(X_1; Y_1|U, V_1, V_2) && \text{(5.10h)}\\
R_1 + R_2 &\leq \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) + \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) &&\\
&\quad + I(X_1, V_2; Y_1|U, V_1, U_2) + I(X_2, V_1; Y_2|U, V_2, U_1) && \text{(5.10i)}\\
2R_1 + R_2 &\leq I(U, V_2, X_1; Y_1) + I(X_1; Y_1|U, V_1, V_2) + I(X_2, V_1; Y_2|U, U_1, V_2) && \text{(5.10j)}\\
&+ \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) &&\\
2R_1 + R_2 &\leq 2\min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) && \text{(5.10k)}\\
&\quad + \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) &&\\
&\quad + I(X_1, V_2; Y_1|U, U_1, U_2) + I(X_1; Y_1|U, V_1, V_2) + I(X_2, V_1; Y_2|U, U_1, V_2) &&\\
R_1 + 2R_2 &\leq I(U, V_1, X_2; Y_2) + I(X_2; Y_2|U, V_2, V_1) + I(X_1, V_2; Y_1|U, U_2, V_1) && \text{(5.10l)}\\
&+ \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) &&\\
R_1 + 2R_2 &\leq 2\min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) && \text{(5.10m)}\\
&\quad + \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) &&\\
&\quad + I(X_2, V_1; Y_2|U, U_1, U_2) + I(X_2; Y_2|U, V_1, V_2) + I(X_1, V_2; Y_1|U, U_2, V_1) &&
\end{aligned}
$$

over all joint distributions

$$p(u)p(u_1|u)p(u_2|u)p(v_1|u,u_1)p(v_2|u,u_2)p(x_1|u,u_1,v_1)p(x_2|u,u_2,v_2)p(\hat{y}_1|y_1)p(\hat{y}_2|y_2),$$

where

$$
\begin{aligned}
\delta_1 &= I(\hat{Y}_1; Y_1|U, U_2, X_1),\\
\delta_2 &= I(\hat{Y}_2; Y_2|U, U_1, X_2).
\end{aligned}
$$

This should be fed back to the transmitter so that it can be used later in refining the desired signals corrupted by the interfering signal; (2) "non-cooperative common": this message is visible to both receivers, however it does not cause any interference (e.g., a_2 at transmitter 1 in the first time slot); (3) "private": this message is visible only to the intended receiver (e.g., a_3 and a_4 at transmitter 1 in the first time sot). Denote these messages by w_{kcc}, w_{knc}, and w_{kp}, respectively, where $k = 1, 2$ is the transmitter index.

• To refine the desired signal corrupted by the cooperative common signal of transmitter 1 (*i.e.*, a_1), receiver 2 utilizes the feedback link to send the interfered signal (*i.e.*, $a_1 \oplus b_1$) back to transmitter 2. Transmitter 2 then employs a partial decode-and-forward to help receiver 1 decode its messages, *i.e.*, the cooperative message of transmitter 1 is decoded at transmitter 2 and it will be forwarded to receiver 2 during another time slot.

• As we can see in this example, encoding operations at each time slot depend on previous operations, thereby motivating us to employ block Markov encoding. As for the decoding, we implement backward decoding at receivers. Each receiver waits until the last time B and we use the last received signal to decode the message received at time $B-2$. We then decode the message received at time $B-3$ and all the way down to the message received at time 1.

These observations will form the basis for our achievable schemes in the following sections.

5.4 Deterministic Interference Channel

In this section, we consider the El Gamal-Costa deterministic IC with rate-limited feedback, described in Section 4.2. The motivating example in the previous section leads us to develop a generic achievable scheme based on three ideas: (1) Han-Kobayashi message splitting [27]; (2) quantize-and-binning; and (3) decode-and-forward [14]. As mentioned earlier, we split the message into three parts: the cooperative common message; non-cooperative common message; and private message. We employ quantize-and-binning to feed back part of the interfered signals through the rate-limited feedback link. With feedback, each transmitter decodes part of the other user's common information (cooperative common) that interfered with its desired signals. We accomplish this by using the partial decode-and-forward scheme. We also derive a new outer bound based on the genie-aided argument [16] and the dependence-balance-bound technique [28, 53, 74].

5.4.1 Achievable Rate Region

Theorem 5.1 *The capacity region of the El Gamal-Costa deterministic IC with rate-limited feedback includes the set $\mathcal{R}$ of (R_1, R_2) satisfying inequalities (5.10a)–(5.10m).*

Proof: We first provide an outline of our achievable scheme. We employ block Markov encoding with a total size B of blocks. In block 1, transmitter 1 splits its own message into cooperative common, non-cooperative common and private parts and then sends a codeword superimposing all of these messages. The cooperative common message is sent via the codeword $u_1^{N,(1)}$. The non-cooperative common message is added to this, being sent via $v_1^{N,(1)}$. The private message is then superimposed on top of the previous messages, being sent via $x_1^{N,(1)}$. Similarly, transmitter 2 sends $x_2^{N,(1)}$. In block 2, receiver 1 quantizes its received signal $y_1^{N,(1)}$ into $\hat{y}_1^{N,(1)}$ with the rate of $\hat{R}_1$. Next it generates a bin index by considering the capacity of its feedback link and then feeds the bin index back to its corresponding transmitter. Similarly, receiver 2 feeds back the corresponding bin index. In block 3, with feedback, each transmitter decodes the other user's cooperative common message (sent in block 1) that interfered with its desired signals. The following messages are then available at the transmitter: (1) its own message; and (2) the other user's cooperative common message decoded with the help of feedback. Using its own cooperative common message as well as the other user's counterpart, each transmitter generates the codeword $u^{N,(3)}$. This captures the correlation between the two transmitters that might induce the cooperative gain. Conditioned on these previous cooperative common messages, each transmitter generates new cooperative common, non-cooperative common, and private messages. It then sends the corresponding codeword. This procedure is repeated until block $B-2$. In the last two blocks $B-1$ and B, to facilitate backward decoding, each transmitter sends the predetermined common messages and a new private message. Each receiver waits until total B blocks have been received and then performs backward decoding.

Codebook Generation: Fix a joint distribution

$$p(u)p(u_1|u)p(u_2|u)p(x_1|u_1,u)p(x_2|u_2,u)p(\hat{y}_1|y_1)p(\hat{y}_2|y_2).$$

We will first show that $p(x_1|u,u_1,v_1)$ and $p(x_2|u,u_2,v_2)$ are functions of the above distributions. To see this, let us write a joint distribution $p(u,u_1,u_2,v_1,v_2,x_1,x_2)$ in two different ways:

$$\begin{aligned} & p(u,u_1,u_2,v_1,v_2,x_1,x_2) \\ &= p(u)p(u_1|u)p(u_2|u)p(x_1|u,u_1)p(x_2|u,u_2)\delta(v_1-g_1(x_1)) \\ &\quad \times \delta(v_2-g_2(x_2)) \qquad (5.11) \\ &= p(u)p(u_1|u)p(u_2|u)p(v_1|u,u_1)p(v_2|u,u_2)p(x_1|u,u_1,v_1) \\ &\quad \times p(x_2|u,u_2,v_2), \end{aligned}$$

where $\delta(\cdot)$ indicates the Kronecker delta function. Notice that by the El Gamal-Costa model assumption (5.4), $p(v_1|u, u_1, x_1) = \delta(v_1 - g_1(x_1))$ and $p(v_2|u, u_2, x_2) = \delta(v_2 - g_2(x_2))$. From this, we can easily see that

$$p(x_1|u, u_1, v_1) = \frac{p(x_1|u, u_1)\delta(v_1 - g_1(x_1))}{p(v_1|u, u_1)},$$

$$p(x_2|u, u_2, v_2) = \frac{p(x_2|u, u_2)\delta(v_2 - g_2(x_1))}{p(v_2|u, u_2)}. \tag{5.12}$$

We now generate codewords as follows. Transmitter 1 generates $2^{N(R_{1cc}+R_{2cc})}$ independent codewords $u^N(i, j)$ according to $\prod_{i=1}^{N} p(u_i)$[1], where $i \in \{1, \cdots, 2^{NR_{1cc}}\}$ and $j \in \{1, \cdots, 2^{NR_{2cc}}\}$. For each codeword $u^N(i, j)$, it generates $2^{NR_{1cc}}$ independent codewords $u_1^N((i, j), k)$ according to $\prod_{i=1}^{N} p(u_{1i}|u_i)$ where $k \in \{1, \cdots, 2^{NR_{1cc}}\}$. Subsequently, for each pair of codewords $\left(u^N(i, j), u_1^N((i, j), k)\right)$, generate $2^{NR_{1nc}}$ independent codewords $v_1^N((i, j), k, l)$ according to $\prod_{i=1}^{N} p(v_{1i}|u_i, u_{1i})$ where $l \in \{1, \cdots, 2^{NR_{1nc}}\}$. Lastly, for each triple of codewords $\left(u^N(i, j), u_1^N((i, j), k), v_1^N((i, j), k, l)\right)$, generate $2^{NR_{1p}}$ independent codewords $x_1^N((i, j), k, l, m)$ according to $\prod_{i=1}^{N} p(x_{1i}|u_i, u_{1i}, v_{1i})$ where $m \in \{1, \cdots, 2^{NR_{1p}}\}$. On the other hand, receiver 1 generates $2^{N\hat{R}_1}$ sequences $\hat{y}_1^N(q)$ according to $\prod_{i=1}^{N} p(\hat{y}_{1i})$ where $q \in \{1, \cdots, 2^{N\hat{R}_1}\}$. In the feedback strategy (to be described shortly), we will see how this codebook generation leads to the joint distribution $p(\hat{y}_1|y_1)$. Similarly, receiver 2 generates $\hat{y}_2^N$.

As it will be clarified later, for a given block b, indices i and j in $u^N(i, j)$ correspond to the cooperative common message of transmitter 1 and transmitter 2 sent during block $(b-2)$ respectively. Then $\prod_{i=1}^{N} p(u_{1i}|u_i)$ is used to create $2^{NR_{1cc}}$ independent codewords corresponding to the cooperative common message of transmitter 1 in $\{1, \cdots, 2^{NR_{1cc}}\}$. Similarly, $2^{NR_{1nc}}$ independent codewords are cre-

[1]With a slight abuse of notation, we use the same index i to represent time.

ated according to $\prod_{i=1}^{N} p(v_{1i}|u_i, u_{1i})$, corresponding to the non-cooperative common message of transmitter 1 in $\{1, \cdots, 2^{NR_{1nc}}\}$. Finally, $\prod_{i=1}^{N} p(x_{1i}|u_i, u_{1i}, v_{1i})$ is used to create $2^{NR_{1p}}$ independent codewords corresponding to the private message of transmitter 1 in $\{1, \cdots, 2^{NR_{1p}}\}$.

Notation: Notations are independently used only for this section. The index k indicates the cooperative common message of user 1 instead of user index. The index i is used for both purposes: (1) indicating the previous cooperative common message of user 1; (2) indicating time index. It could be easily differentiated from contexts.

Feedback Strategy (Quantize-and-Binning): Focus on the b-th transmission block. First receiver 1 quantizes its received signal $y_1^{N,(b)}$ into $\hat{y}_1^{N,(b+1)}$ with the rate of $\hat{R}$. Next it finds an index q such that $\left(\hat{y}_1^{N,(b+1)}(q), y_1^{N,(b)}\right) \in \mathcal{T}_\epsilon^{(N)}$, where $q \in [1 : 2^{N\hat{R}_1}]$ and $\mathcal{T}_\epsilon^{(N)}$ indicates a jointly typical set. The quantization rate $\hat{R}_1$ is chosen so as to ensure the existence of such an index with probability 1. The covering lemma in [21] guides us to choose $\hat{R}_1$ such that

$$\hat{R}_1 \geq I(\hat{Y}_1; Y_1), \tag{5.13}$$

since under the above constraint, the probability that there is no such an index becomes arbitrarily small as N goes to infinity. Notice that with this choice of $\hat{R}_1$, the codebook $\hat{y}_1^N(q)$ according to $\prod_{i=1}^{N} p(\hat{y}_{1i})$ would match the codeword according to $\prod_{i=1}^{N} p(\hat{y}_{1i}|y_{1i})$.

We then partition the set of indices $q \in [1 : 2^{N\hat{R}_1}]$ into the number $2^{NC_{\mathsf{FB1}}}$ of equal-size subsets (that we call *bins*):

$$\mathcal{B}(r) = \left[(r-1)2^{N(\hat{R}_1 - C_{\mathsf{FB1}})} + 1 : r2^{N(\hat{R}_1 - C_{\mathsf{FB1}})}\right],$$
$$r \in [1 : 2^{NC_{\mathsf{FB1}}}].$$

Now the idea is to feed back the bin index r such that $q \in \mathcal{B}(r)$. This helps transmitter 1 to decode the quantized signal $\hat{y}_1^{N,(b)}$. Specifically, using the bin index r, transmitter 1 finds a unique index $q \in \mathcal{B}(r)$ such that $\left(\hat{y}_1^{N,(b+1)}(q), x_1^{N,(b)}, u^{N,(b)}\right) \in \mathcal{T}_\epsilon^{(N)}$. Notice that by the packing lemma in [21], the decoding error probability goes to zero if

$$\hat{R}_1 - C_{\mathsf{FB1}} \leq I(\hat{Y}_1; X_1, U). \tag{5.14}$$

Using (5.13) and (5.14), transmitter 1 can now decode the quantized signal as long as

$$I(\hat{Y}_1; Y_1 | X_1, U) \leq C_{\mathsf{FB1}}, \tag{5.15}$$

Similarly, transmitter 2 can decode $\hat{y}_2^{N,(b+1)}(q)$ if

$$I(\hat{Y}_2; Y_2 | X_2, U) \leq C_{\mathsf{FB2}}. \tag{5.16}$$

Encoding: Given $\hat{y}_1^{N,(b-1)}$ (decoded with the help of feedback), transmitter 1 finds a unique index $\hat{w}_{2cc}^{(b-2)} = \hat{k}$ (sent from transmitter 2 in the $(b-2)$-th block) such that

$$\left(u^N(\cdot), u_1^N(\cdot), v_1^N(\cdot), x_1^N(\cdot), u_2^N(\cdot, \hat{k}), \hat{y}_1^{N,(b-1)}\right) \in \mathcal{T}_\epsilon^{(N)},$$

where $(\cdot)$ indicates the known messages $(w_{1cc}^{(b-4)}, \hat{w}_{2cc}^{(b-4)}, w_{1nc}^{(b-2)}, w_{1p}^{(b-2)})$. Notice that due to the feedback delay, the fed back signal contains information of the $(b-2)$-th block. We assume that $\hat{w}_{2cc}^{(b-2)}$ is correctly decoded from the previous block.

By the packing lemma [21], the decoding error probability becomes arbitrarily small (as N goes to infinity) if

$$\begin{aligned} R_{2cc} &\le I(U_2; \hat{Y}_1 | X_1, U) \\ &= I(\hat{Y}_1; Y_1 | U, X_1) - I(\hat{Y}_1; Y_1 | U, U_2, X_1) \\ &\le \min(C_{\mathsf{FB1}} - \delta_1, I(U_2; Y_1 | X_1, U)), \end{aligned} \tag{5.17}$$

where the last inequality follows from (5.15), $\delta_1 := I(\hat{Y}_1; Y_1 | U, U_2, X_1)$ and $I(U_2; \hat{Y}_1 | X_1, U) \le I(U_2; Y_1 | X_1, U)$.

Based on $(w_{1cc}^{(b-2)}, \hat{w}_{2cc}^{(b-2)})$, transmitter 1 generates a new cooperative-common message $w_{1cc}^{(b)}$, a non-cooperative-common message $w_{1nc}^{(b)}$ and a private message $w_{1p}^{(b)}$. It then sends x_1^N. Similarly transmitter 2 decodes $\hat{w}_{1cc}^{(b-2)}$, generates $(w_{2cc}^{(b)}, w_{2nc}^{(b)}, w_{2p}^{(b)})$ and then sends x_2^N.

Decoding: Each receiver waits until total B blocks have been received and then does backward decoding. Notice that a block index b starts from the last B and ends to 1. For block b, receiver 1 finds the unique indices $(\hat{i}, \hat{j}, \hat{k}, \hat{l})$ such that for some $m \in [1 : 2^{NR_{2nc}}]$

$$\begin{aligned} \Big(& u^N(\hat{i}, \hat{j}), u_1^N((\hat{i}, \hat{j}), \hat{w}_{1cc}^{(b)}), v_1^N((\hat{i}, \hat{j}), \hat{w}_{1cc}^{(b)}, \hat{k}), x_1^N((\hat{i}, \hat{j}), \hat{w}_{1cc}^{(b)} \\ & , \hat{k}, \hat{l}), u_2^N((\hat{i}, \hat{j}), \hat{w}_{2cc}^{(b)}), v_2^N((\hat{i}, \hat{j}), \hat{w}_{2cc}^{(b)}, m), y_1^{N,(b)} \Big) \in \mathcal{T}_\epsilon^{(N)}, \end{aligned}$$

where we assumed that a pair of messages $(\hat{w}_{1cc}^{(b)}, \hat{w}_{2cc}^{(b)})$ was successively decoded from the future blocks. Similarly receiver 2 decodes $(\hat{w}_{1cc}^{(b-2)}, \hat{w}_{2cc}^{(b-2)}, \hat{w}_{2nc}^{(b)}, \hat{w}_{2p}^{(b)})$.

Analysis of Probability of Error: By symmetry, we consider the probability of error only for block b and for a pair of transmitter 1 and receiver 1. We assume that $(w_{1cc}^{(b-2)}, w_{2cc}^{(b-2)}, w_{1nc}^{(b)}, w_{1p}^{(b)}) = (1, 1, 1, 1)$ was sent through the blocks; and there was no backward decoding error from the future blocks, *i.e.*, $(\hat{w}_{1cc}^{(b)}, \hat{w}_{2cc}^{(b)})$ are successfully decoded.

Define an event:

$$E_{ijklm} = \Big\{ \Big(u^N(i,j), u_1^N((i,j), \hat{w}_{1cc}^{(b)}), v_1^N((i,j), \hat{w}_{1cc}^{(b)}, k) \\ , x_1^N((i,j), \hat{w}_{1cc}^{(b)}, k, l), u_2^N((i,j), \hat{w}_{2cc}^{(b)}), v_2^N((i,j), \hat{w}_{2cc}^{(b)}, m) \\ , y_1^{N,(b)} \Big) \in \mathcal{T}_\epsilon^{(N)} \Big\}.$$

Let E_{1111m}^c be the complement of the set E_{1111m}. Then, by AEP, $\Pr(E_{1111m}^c) \to 0$ as N goes to infinity. Hence, we focus only on the following error event.

$$\begin{aligned} &\Pr\left(\bigcup_{(i,j,k,l)\neq(1,1,1,1),m} E_{ijklm} \right) \\ &\leq \underbrace{\sum_{(i,j)\neq(1,1)} \Pr\left(\bigcup_{k,l,m} E_{ijklm} \right)}_{\triangleq \Pr(E_1)} \\ &+ \underbrace{\sum_{k\neq 1} \Pr\left(\bigcup_{l} E_{11kl1} \right)}_{\triangleq \Pr(E_2)} + \underbrace{\sum_{k\neq 1, m\neq 1} \Pr\left(\bigcup_{l} E_{11klm} \right)}_{\triangleq \Pr(E_3)} \\ &+ \underbrace{\sum_{l\neq 1} \Pr\left(E_{111l1} \right)}_{\triangleq \Pr(E_4)} + \underbrace{\sum_{l\neq 1, m\neq 1} \Pr\left(E_{111lm} \right)}_{\triangleq \Pr(E_5)}. \end{aligned} \tag{5.18}$$

Here we have:

$$\begin{aligned} \Pr(E_1) &\leq 24 \\ &\quad \times 2^{N(R_{1cc}+R_{2cc}+R_{1nc}+R_{2nc}+R_{1p}-I(U,X_1,V_2;Y_1)+5\epsilon)} \\ \Pr(E_2) &\leq 2 \times 2^{N(R_{1nc}+R_{1p}-I(X_1;Y_1|U,U_1,V_2)+2\epsilon)} \\ \Pr(E_3) &\leq 2 \times 2^{N(R_{1nc}+R_{2nc}+R_{1p}-I(X_1,V_2;Y_1|U,U_1,U_2)+3\epsilon)} \\ \Pr(E_4) &\leq 2^{N(R_{1p}-I(X_1;Y_1|U,V_1,V_2)+\epsilon)} \\ \Pr(E_5) &\leq 2^{N(R_{2nc}+R_{1p}-I(X_1,V_2;Y_1|U,V_1,U_2)+2\epsilon)}. \end{aligned}$$

Notice in $\Pr(E_1)$ that as long as $(i, j) \neq (1, 1)$, all of the cases (decided depending on whether or not $k \neq 1$, $l \neq 1$ and $m \neq 1$) are dominated by the worst case bound that occurs when $k \neq 1$, $l \neq 1$ and $m \neq 1$. Since $(i, j) \neq (1, 1)$ covers three different cases and we have eight different cases depending on the values of (k, l, m), we have 24 cases in total. This number reflects the constant 24 in the above first inequality. Similarly, we get the other four inequalities as above.

Hence, the probability of error can be made arbitrarily small if

$$\begin{cases} R_{2cc} \leq \min(I(U_2; Y_1|U, X_1), C_{\mathsf{FB1}} - \delta_1) \\ R_{1p} \leq I(X_1; Y_1|U, V_1, V_2) \\ R_{1p} + R_{2nc} \leq I(X_1, V_2; Y_1|U, V_1, U_2) \\ R_{1p} + R_{1nc} \leq I(X_1; Y_1|U, U_1, V_2) \\ R_{1p} + R_{1nc} + R_{2nc} \leq I(X_1, V_2; Y_1|U, U_1, U_2) \\ R_{1p} + R_{1cc} + R_{2cc} + R_{1nc} + R_{2nc} \leq I(U, V_2, X_1; Y_1), \end{cases} \tag{5.19}$$

$$\begin{cases} R_{1cc} \leq \min(I(U_1; Y_2|U, X_2), C_{\mathsf{FB2}} - \delta_2) \\ R_{2p} \leq I(X_2; Y_2|U, V_1, V_2) \\ R_{2p} + R_{1nc} \leq I(X_2, V_1; Y_2|U, V_2, U_1) \\ R_{2p} + R_{2nc} \leq I(X_2; Y_2|U, U_2, V_1) \\ R_{2p} + R_{2nc} + R_{1nc} \leq I(X_2, V_1; Y_2|U, U_1, U_2) \\ R_{2p} + R_{2cc} + R_{1cc} + R_{1nc} + R_{2nc} \leq I(U, V_1, X_2; Y_2). \end{cases} \tag{5.20}$$

Employing Fourier-Motzkin-Elimination, we finally get the bounds of (5.10a)–(5.10m). ■

Remark 5.1 (Connection to Related Work [43, 56]) *The three-fold message splitting in our achievable scheme is a special case of the more-than-three-fold message splitting introduced in [43, 56]. Also our scheme has similarity to the schemes in [43, 56] in a sense that the three techniques (message-splitting, block Markov encoding and back-*

ward decoding) are jointly employed. However, due to a fundamental difference between our rate-limited feedback problem and the conferencing encoder problem in [43, 56], a new scheme is required for feedback strategy and this is reflected as the quantize-and-binning scheme in the El Gamal-Costa deterministic model. It turns out this distinction leads to a new lattice-code-based scheme in the Gaussian case, as will be explained in Section VI.

5.4.2 Outer-bound

Theorem 5.2 *The capacity region of the two-user El Gamal-Costa deterministic IC with rate limited feedback (as described in Section 4.2) is included by the set $\bar{C}$ of (R_1, R_2) such that*

$$
\begin{aligned}
R_1 &\leq \min\{H(Y_1), H(Y_1|V_1, V_2, U_1) && (5.21a)\\
&+ H(Y_2|X_2, U_1)\} \\
R_1 &\leq H(Y_1|X_2, U_1) + C_{\mathsf{FB2}} && (5.21b)\\
R_2 &\leq \min\{H(Y_2), H(Y_2|V_1, V_2, U_2) && (5.21c)\\
&+ H(Y_1|X_1, U_2)\} \\
R_2 &\leq H(Y_2|X_1, U_2) + C_{\mathsf{FB1}} && (5.21d)\\
R_1 + R_2 &\leq H(Y_1|V_1, V_2, U_1) + H(Y_2) && (5.21e)\\
R_1 + R_2 &\leq H(Y_2|V_1, V_2, U_2) + H(Y_1) && (5.21f)\\
R_1 + R_2 &\leq H(Y_1|V_1) + H(Y_2|V_2) && (5.21g)\\
&+ C_{\mathsf{FB1}} + C_{\mathsf{FB2}} \\
2R_1 + R_2 &\leq H(Y_1) + H(Y_1|V_1, V_2, U_1) && (5.21h)\\
&+ H(Y_2|V_2) + C_{\mathsf{FB2}}
\end{aligned}
$$

$$R_1 + 2R_2 \leq H(Y_2) + H(Y_2|V_1, V_2, U_2) + H(Y_1|V_1) + C_{\mathsf{FB1}}, \tag{5.21i}$$

for joint distributions $p(u_1, u_2)p(x_1|u_1, u_2)p(x_2|u_1, u_2)$*. As depicted in Figure 5.2,* C_{FB1} *and* C_{FB2} *indicate the capacity of each feedback link.*

Remark 5.2 *In the non-feedback case,* i.e., $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = 0$*, by setting* $U_1 = U_2 = \emptyset$*, we recover the outer-bounds of Theorem* 1 *in [16]. Note that in this case,* $H(Y_1|X_2) = H(Y_1|V_2)$ *and* $H(Y_2|X_1) = H(Y_2|V_1)$*. In fact, our achievable region of Theorem 5.1 matches the outer-bound under this model, thereby achieving the non-feedback capacity region.*

Remark 5.3 (Feedback gain under symmetric feedback cost) *Notice from (5.21g) that the sum-rate capacity can be at most increased by the rate of available feedback,* i.e.*, one bit of feedback provides a capacity increase of at most one bit. Therefore, if the cost of using the feedback link is the same as that of using the forward link, there is no feedback gain under the feedback cost. However, it turns out that there is indeed feedback gain when the costs are* asymmetric*. This will be discussed in more details in Remark 5.5 of Section 5.5.*

Proof: By symmetry, it suffices to prove the bounds of (5.21a), (5.21b), (5.21e), (5.21g), and (5.21h). The bounds of (5.21a) and (5.21b) are nothing but the cut-set bounds (see Appendix D.1 for details). Also (5.21e) is the bound when the feedback link has infinite capacity [50]. Hence, proving the bounds of (5.21g) and (5.21h) is the main focus of the proof. We will present the proof of (5.21g) here and for completeness, the proof for all other bounds is provided in Appendix D.1.

Proof of (5.21g):

$$
\begin{aligned}
N(R_1 + R_2 - \epsilon_N) &\leq I(W_1; Y_1^N) + I(W_2; Y_2^N) \\
&= H(Y_1^N) - H(Y_1^N|W_1) + H(Y_2^N) - H(Y_2^N|W_2) \\
&\overset{(a)}{=} H(Y_1^N) - H(V_1^N|W_2) + H(Y_2^N) - H(V_2^N|W_1) \\
&= H(Y_1^N) - [H(V_1^N) - I(V_1^N; W_2)] + H(Y_2^N) \\
&\quad - [H(V_2^N) - I(V_2^N; W_1)] \\
&\overset{(b)}{\leq} I(V_1^N; W_2) + I(V_2^N; W_1) + H(Y_1^N, V_1^N) - H(V_1^N) \\
&\quad + H(Y_2^N, V_2^N) - H(V_2^N) \\
&= I(V_1^N; W_2) + I(V_2^N; W_1) + H(Y_1^N|V_1^N) + H(Y_2^N|V_2^N) \\
&\overset{(c)}{\leq} I(V_1^N, W_1, \tilde{Y}_1^N; W_2) + I(V_2^N, W_2, \tilde{Y}_2^N; W_1) \\
&\quad + H(Y_1^N|V_1^N) + H(Y_2^N|V_2^N) \\
&\overset{(d)}{=} I(W_1, \tilde{Y}_1^N; W_2) + I(W_2, \tilde{Y}_2^N; W_1) + H(Y_1^N|V_1^N) \\
&\quad + H(Y_2^N|V_2^N) \\
&= I(\tilde{Y}_1^N; W_2|W_1) + I(\tilde{Y}_2^N; W_1|W_2) + H(Y_1^N|V_1^N) + \\
&\quad H(Y_2^N|V_2^N) \\
&= H(\tilde{Y}_1^N|W_1) + H(\tilde{Y}_2^N|W_2) + H(Y_1^N|V_1^N) + H(Y_2^N|V_2^N) \\
&\overset{(e)}{\leq} N(C_{\mathsf{FB1}} + C_{\mathsf{FB2}}) + \sum H(Y_{1i}|V_{1i}) + \sum H(Y_{2i}|V_{2i}),
\end{aligned}
$$

where (a) follows from $H(Y_1^N|W_1) = H(V_2^N|W_1)$ and $H(Y_2^N|W_2) = H(V_1^N|W_2)$ (see Claim 5.1 below); (b) follows from providing V_1^N and V_2^N to receiver 1 and 2, respectively; (c) follows from the fact that adding information increases mutual information; (d) follows from the fact that V_k^N is a function of $(W_k, \tilde{Y}_k^{N-1})$; (e) follows from the fact that $H(\tilde{Y}_k^N|W_k) \leq NC_{\mathsf{FBk}}$ and conditioning reduces entropy.

Claim 5.1 $H(Y_1^N|W_1) = H(V_2^N|W_1)$ *and* $H(Y_2^N|W_2) = H(V_1^N|W_2)$.

Proof: By symmetry, it suffices to prove the first one.

$$
\begin{aligned}
H(Y_1^N|W_1) &= \sum H(Y_{1i}|Y_1^{i-1}, W_1) \\
&\overset{(a)}{=} \sum H(V_{2i}|Y_1^{i-1}, W_1) \\
&\overset{(b)}{=} \sum H(V_{2i}|Y_1^{i-1}, W_1, X_1^i, V_2^{i-1}) \\
&\overset{(c)}{=} \sum H(V_{2i}|W_1, V_2^{i-1}) = H(V_2^N|W_1),
\end{aligned}
$$

where (a) follows from the fact that Y_{1i} is a function of (X_{1i}, V_{2i}) and X_{1i} is a function of (W_1, Y_1^{i-1}); (b) follows from the fact that X_1^i is a function of (W_1, Y_1^{i-1}) and V_2^{i-1} is a function of (Y_1^{i-1}, X_1^{i-1}); (c) follows from the fact that Y_1^{i-1} is a function of (X_1^{i-1}, V_2^{i-1}) and X_1^i is a function of (W_1, V_2^{i-1}) (see Claim 5.2 below). ■

Claim 5.2 *For $i \geq 1$, X_1^i is a function of (W_1, V_2^{i-1}). Similarly, X_2^i is a function of (W_2, V_1^{i-1}).*

Proof: By symmetry, it is enough to prove the first one. Since the channel is deterministic, X_1^i is a function of (W_1, W_2). In Figure 5.2, we see that information of W_2 to the first link pair must pass through V_{2i}. Also note that X_{1i} depends on the past output sequences until $i-1$. Therefore, X_1^i is a function of (W_1, V_2^{i-1}). ■

■

5.5 Linear Deterministic Interference Channel

In this section, we consider the linear deterministic IC with rate-limited feedback described in Section 4.2. Since this model is a special case of the El Gamal-Costa model, our inner and outer bounds derived in the previous section also apply to this model. We show that the inner-bound and the outer bound derived in Theorem 5.1 and 5.2 respectively, coincide under this linear deterministic model, thus establishing the capacity region.

Theorem 5.3 *The capacity region of the linear deterministic IC with rate-limited feedback is the set of non-negative* (R_1, R_2) *satisfying*

$$\begin{aligned}
R_1 &\leq \min\{\max(n_{11}, n_{21}), \max(n_{11}, n_{12})\} && \text{(5.22a)}\\
R_1 &\leq n_{11} + C_{\mathsf{FB2}} && \text{(5.22b)}\\
R_2 &\leq \min\{\max(n_{22}, n_{12}), \max(n_{22}, n_{21})\} && \text{(5.22c)}\\
R_2 &\leq n_{22} + C_{\mathsf{FB1}} && \text{(5.22d)}\\
R_1 + R_2 &\leq (n_{11} - n_{12})^+ + \max(n_{22}, n_{12}) && \text{(5.22e)}\\
R_1 + R_2 &\leq (n_{22} - n_{21})^+ + \max(n_{11}, n_{21}) && \text{(5.22f)}\\
R_1 + R_2 &\leq \max\{n_{21}, (n_{11} - n_{12})^+\} && \text{(5.22g)}\\
&+ \max\{n_{12}, (n_{22} - n_{21})^+\} + C_{\mathsf{FB1}} + C_{\mathsf{FB2}} &&\\
2R_1 + R_2 &\leq (n_{11} - n_{12})^+ + \max(n_{11}, n_{21}) && \text{(5.22h)}\\
&+ \max\{n_{12}, (n_{22} - n_{21})^+\} + C_{\mathsf{FB2}} &&\\
R_1 + 2R_2 &\leq (n_{22} - n_{21})^+ + \max(n_{22}, n_{12}) && \text{(5.22i)}\\
&+ \max\{n_{21}, (n_{11} - n_{12})^+\} + C_{\mathsf{FB1}}. &&
\end{aligned}$$

Remark 5.4 *In the non-feedback case,* i.e., $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = 0$, *this theorem recovers the result of [12, 16]. In the infinite feedback case,* i.e., $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = \infty$, *this recovers the result of [50, 51]. Considering the sum-rate capacity under symmetric setting,* i.e., $n_{11} = n_{22} = n$, $n_{12} = n_{21} = m$, $C_{\mathsf{FB1}} = C_{\mathsf{FB2}}$, *this recovers the result of [57].*

Proof: The converse proof is trivial due to Theorem 5.2. For achievability, we will use the result in Theorem 5.1. By choosing the following input distribution, we will show the tightness of the outer bound. $\forall k \in \{1, 2\}$ and $j \neq k$, we choose

$$\begin{aligned}
U &= \varnothing, \\
U_k &= U \oplus X_{kcc}, \\
V_k &= U_k \oplus X_{knc}, \\
X_k &= V_k \oplus X_{kp}, \\
\hat{Y}_k &= \mathsf{LSB}_{\min(n_{kj}, C_{\mathsf{FBj}})}(Y_k),
\end{aligned} \tag{5.23}$$

where for any column vector A, $\mathsf{LSB}_n(A)$ takes the bottom n ($n \leq |A|$) entries of A while returning zeros for the remaining part; and X_{kcc}, X_{knc}, and X_{kp} are independent random vectors of size $\max\{n_{kk}, n_{kj}\}$, such that

- The random vector X_{kp} consists of $(n_{kk} - n_{kj})^+$ i.i.d. $\mathsf{Ber}\left(\frac{1}{2}\right)$ random variables at the bottom, denoted by $*$ in (5.24), corresponding to the number of private signal levels of transmitter k.
- The random vector X_{kcc} consists of $(n_{kk} - n_{kj})^+$ i.i.d. $\mathsf{Ber}\left(\frac{1}{2}\right)$ random variables in the middle (above the private signal levels), denoted by $*$ in (5.24), corresponding to the number of common signal levels that will be re-sent cooperatively through the other commuication link with the help of feedback.
- The random vector X_{knc} consists of $(n_{kj} - C_{\mathsf{FBj}})^+$ i.i.d. $\mathsf{Ber}\left(\frac{1}{2}\right)$ random variables at the top, denoted by $*$ in (5.24), corresponding to the number of non-cooperative common signal levels.

As we show in Appendix D.2, with this choice of random variables, the achievable region of Theorem 5.1 matches the outer-bounds in Theorem 5.2.

$$X_k = \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \\ \hline * \\ \vdots \\ * \end{bmatrix}}_{X_{kp}} \oplus \underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline * \\ \vdots \\ * \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}}_{X_{kcc}} \oplus \underbrace{\begin{bmatrix} * \\ \vdots \\ * \\ \hline 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}}_{X_{knc}}, \quad k = 1, 2. \tag{5.24}$$

■

It is worth utilizing Theorem 5.3 to illustrate the impact of feedback on the sum-rate capacity of the linear deterministic IC. Consider a symmetric case where $n_{11} = n_{22} = n$, $n_{12} = n_{21} = \alpha n$, and $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = \beta n$. Using Theorem 5.3, we can derive the sum-rate capacity of this network (normalized by n)

$$\frac{C_{\mathsf{sum}}}{n} = \begin{cases} \min(2 - 2\alpha + 2\beta, 2 - \alpha) & \text{for } \alpha \in [0, 0.5] \\ \min(2\alpha + 2\beta, 2 - \alpha) & \text{for } \alpha \in [0.5, \frac{2}{3}] \\ 2 - \alpha & \text{for } \alpha \in [\frac{2}{3}, 1] \\ \alpha & \text{for } \alpha \in [1, 2 + 2\beta] \\ 2 + 2\beta & \text{for } \alpha \in [2 + 2\beta, \infty) \end{cases} \tag{5.25}$$

Figure 5.5 illustrates the (normalized) sum-rate capacity as a function of α, for different values of $\beta = 0$ (*i.e.*, no feedback), $\beta = \infty$ (*i.e.*, infinite feedback), and $\beta = 0.125$. We note the following cases:

- Case 1 ($\alpha \in \left[0, \frac{1}{2}\right]$): In this regime the sum-rate capacity is increased by the total amount of feedback rates and saturates at $2-\alpha$ once the rate of each feedback link is larger than $\frac{\alpha n}{2}$.

- Case 2 ($\alpha \in \left[\frac{1}{2}, \frac{2}{3}\right]$): In this regime the sum-rate capacity is increased by the total amount of feedback rates and saturates at 2α once the rate of each feedback link is larger than $\frac{(2n-3\alpha n)}{2}$.
- Case 3 ($\alpha \in \left[\frac{2}{3}, 2\right]$): In this regime feedback does not increase the capacity.
- Case 4 ($\alpha \in [2+2\beta, \infty)$): In this regime the sum-rate capacity is increased by at most the total amount of feedback rates.

Remark 5.5 (Feedback gain under asymmetric feedback cost) *As it can be seen in (5.25) the sum-rate capacity is increased by at most the total amount of feedback rates. Let the* cost *be the amount of resources (e.g., time, frequency) paid for sending one bit. With this cost in mind, let us consider the effective gain of using feedback which counts the cost. Notice that by Case 1,2, and 4, there are many channel parameter scenarios where one bit of feedback can provide a capacity increase of exactly one bit. This implies that the effective feedback gain depends on the cost difference between feedback and forward links. So if the feedback cost is cheaper than that of using forward link, then there is indeed feedback gain. The cellular network may be this case. Suppose that downlink is used for feedback purpose, while uplink is used as a forward link. Then, this is the scenario where the feedback cost is cheaper than the cost of using the forward link, as downlink power is typically larger than uplink power, thus inducing cheaper feedback cost.*

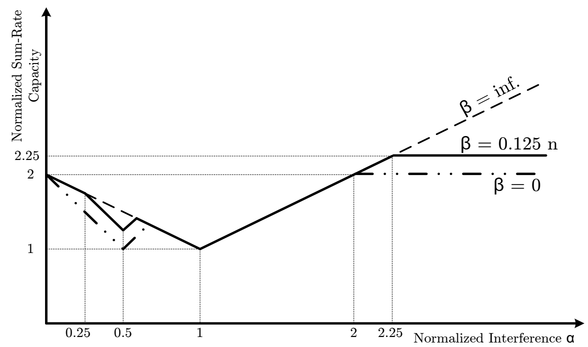


Figure 5.5: Normalized sum-rate capacity for $\beta = 0, \beta = 0.125$ and $\beta = \infty$.

5.6 Gaussian Interference Channel

In this section, we consider the Gaussian IC with rate-limited feedback, described in Section 4.2. We first derive an outer-bound on the capacity region of this network. We then develop an achievability strategy based on the techniques discussed in the previous sections and then show that for symmetric channel gains it achieves a sum-rate within a constant gap to the optimality.

5.6.1 Outer-bound

Theorem 5.4 *The capacity region of the Gaussian IC with rate-limited feedback is included in the closure of the set $\bar{C}$ of (R_1, R_2) satisfying inequalities (5.26a)–(5.26k) over* $0 \leq \rho \leq 1$.

Proof: By symmetry, it suffices to prove the bounds of (5.26a), (5.26b), (5.26c), (5.26g), (5.26h) and (5.26j). The bounds of (5.26a), (5.26b) and (5.26c) are nothing but cutset bounds. The bound of (5.26h) corresponds to the case of infinite feedback rate and was derived in [50]. Hence, proving the bounds of (5.26g) and

$$
\begin{aligned}
R_1 &\leq \log\left(1 + \mathsf{SNR}_1 + \mathsf{INR}_{21} + 2\rho\sqrt{\mathsf{SNR}_1 \cdot \mathsf{INR}_{21}}\right) && \text{(5.26a)}\\
R_1 &\leq \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_1}{1+(1-\rho^2)\mathsf{INR}_{12}}\right) + \log\left(1 + (1-\rho^2)\mathsf{INR}_{12}\right) && \text{(5.26b)}\\
R_1 &\leq \log\left(1 + (1-\rho^2)\mathsf{SNR}_1\right) + C_{\mathsf{FB2}} && \text{(5.26c)}\\
R_2 &\leq \log\left(1 + \mathsf{SNR}_2 + \mathsf{INR}_{12} + 2\rho\sqrt{\mathsf{SNR}_2 \cdot \mathsf{INR}_{12}}\right) && \text{(5.26d)}\\
R_2 &\leq \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_2}{1+(1-\rho^2)\mathsf{INR}_{21}}\right) + \log\left(1 + (1-\rho^2)\mathsf{INR}_{21}\right) && \text{(5.26e)}\\
R_2 &\leq \log\left(1 + (1-\rho^2)\mathsf{SNR}_2\right) + C_{\mathsf{FB1}} && \text{(5.26f)}\\
R_1 + R_2 &\leq \log\left(1 + (1-\rho^2)\mathsf{INR}_{21} + \frac{(1-\rho^2)\mathsf{SNR}_1}{1+(1-\rho^2)\mathsf{INR}_{12}}\right)\\
&\quad + \log\left(1 + (1-\rho^2)\mathsf{INR}_{12} + \frac{(1-\rho^2)\mathsf{SNR}_2}{1+(1-\rho^2)\mathsf{INR}_{21}}\right)\\
&\quad + C_{\mathsf{FB1}} + C_{\mathsf{FB2}} && \text{(5.26g)}\\
R_1 + R_2 &\leq \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_1}{1+(1-\rho^2)\mathsf{INR}_{12}}\right)\\
&\quad + \log\left(1 + \mathsf{SNR}_2 + \mathsf{INR}_{12} + 2\rho\sqrt{\mathsf{SNR}_2 \cdot \mathsf{INR}_{12}}\right) && \text{(5.26h)}\\
R_1 + R_2 &\leq \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_2}{1+(1-\rho^2)\mathsf{INR}_{21}}\right)\\
&\quad + \log\left(1 + \mathsf{SNR}_1 + \mathsf{INR}_{21} + 2\rho\sqrt{\mathsf{SNR}_1 \cdot \mathsf{INR}_{21}}\right) && \text{(5.26i)}\\
2R_1 + R_2 &\leq \log\left(1 + \mathsf{SNR}_1 + \mathsf{INR}_{21} + 2\rho\sqrt{\mathsf{SNR}_1 \cdot \mathsf{INR}_{21}}\right)\\
&\quad + \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_1}{1+(1-\rho^2)\mathsf{INR}_{12}}\right) && \text{(5.26j)}\\
&\quad + \log\left(1 + (1-\rho^2)\mathsf{INR}_{12} + \frac{(1-\rho^2)\mathsf{SNR}_2}{1+(1-\rho^2)\mathsf{INR}_{21}}\right) + C_{\mathsf{FB1}} + C_{\mathsf{FB2}}\\
R_1 + 2R_2 &\leq \log\left(1 + \mathsf{SNR}_2 + \mathsf{INR}_{12} + 2\rho\sqrt{\mathsf{SNR}_2 \cdot \mathsf{INR}_{12}}\right)\\
&\quad + \log\left(1 + \frac{(1-\rho^2)\mathsf{SNR}_2}{1+(1-\rho^2)\mathsf{INR}_{21}}\right) && \text{(5.26k)}\\
&\quad + \log\left(1 + (1-\rho^2)\mathsf{INR}_{21} + \frac{(1-\rho^2)\mathsf{SNR}_1}{1+(1-\rho^2)\mathsf{INR}_{12}}\right) + C_{\mathsf{FB1}} + C_{\mathsf{FB2}}
\end{aligned}
$$

(5.26j) is the main focus of this proof. We will present the proof of (5.26g) here, and defer the proof for remaining bounds to Appendix D.3.

Proof of (5.26g): The proof idea mostly follows the deterministic case proof of 5.21g. The only difference in the Gaussian case is that we define a noisy version of $h_{12}X_1^N$ corresponding to V_1^N in the deterministic case: $S_1^N := h_{12}X_1^N + Z_2^N$. Similarly we define $S_2^N := h_{21}X_2^N + Z_1^N$ to mimic V_2^N. With this, we can now get:

$$N(R_1 + R_2 - \epsilon_N) \stackrel{(a)}{\leq} I\left(W_1; Y_1^N\right) + I\left(W_2; Y_2^N\right)$$
$$\stackrel{(b)}{=} h\left(Y_1^N\right) + h\left(Y_2^N\right) - h\left(S_1^N|W_2\right) - h\left(S_2^N|W_1\right)$$
$$= I\left(S_1^N; W_2\right) + I\left(S_2^N; W_1\right) - h\left(S_1^N|Y_1^N\right) - h\left(S_2^N|Y_2^N\right)$$
$$+ \underbrace{h\left(Y_1^N|S_1^N\right) + h\left(Y_2^N|S_2^N\right)}_{T}$$
$$\stackrel{(c)}{\leq} T + I\left(S_1^N, \tilde{Y}_1^N, W_1; W_2\right) + I\left(S_2^N, \tilde{Y}_2^N, W_2; W_1\right)$$
$$- h\left(S_1^N|Y_1^N, W_1, \tilde{Y}_1^N\right) - h\left(S_2^N|Y_2^N, W_2, \tilde{Y}_2^N\right)$$
$$\stackrel{(d)}{=} T + I\left(\tilde{Y}_1^N; W_2|W_1\right) + I\left(S_1^N; W_2|W_1, \tilde{Y}_1^N\right)$$
$$+ I\left(\tilde{Y}_2^N; W_1|W_2\right) + I\left(S_2^N; W_1|W_2, \tilde{Y}_2^N\right)$$
$$- h\left(Z_1^N|S_1^N, W_2, \tilde{Y}_2^N\right) - h\left(Z_2^N|S_2^N, W_1, \tilde{Y}_1^N\right)$$
$$\stackrel{(e)}{=} \underbrace{T - h\left(Z_1^N\right) - h\left(Z_2^N\right)}_{T'} + I\left(\tilde{Y}_1^N; W_2|W_1\right)$$
$$+ I\left(\tilde{Y}_2^N; W_1|W_2\right) + I\left(Z_2^N; S_2^N|W_1, \tilde{Y}_1^N\right)$$
$$+ I\left(Z_1^N; S_1^N|W_2, \tilde{Y}_2^N\right) - h\left(Z_1^N|W_1, W_2, \tilde{Y}_2^N\right)$$
$$+ h\left(Z_1^N\right) - h\left(Z_2^N|W_1, W_2, \tilde{Y}_1^N\right) + h\left(Z_2^N\right)$$
$$\stackrel{(f)}{=} T' + I\left(\tilde{Y}_1^N; W_2|W_1\right) + I\left(\tilde{Y}_2^N; W_1|W_2\right)$$

$$+ I\left(Z_2^N; S_2^N|W_1, \tilde{Y}_1^N\right) + I\left(Z_1^N; S_1^N|W_2, \tilde{Y}_2^N\right)$$
$$+ I\left(Z_1^N; \tilde{Y}_2^N|W_1, W_2\right) + I\left(Z_2^N; \tilde{Y}_1^N|W_1, W_2\right)$$
$$= T' + I\left(\tilde{Y}_1^N; W_2, Z_2^N|W_1\right) + I\left(\tilde{Y}_2^N; W_1, Z_1^N|W_2\right)$$
$$+ I\left(Z_2^N; S_2^N|W_1, \tilde{Y}_1^N\right) + I\left(Z_1^N; S_1^N|W_2, \tilde{Y}_2^N\right)$$

$$
\begin{aligned}
&= T' + I\left(\tilde{Y}_1^N; W_2|W_1, Z_2^N\right) + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I\left(\tilde{Y}_1^N, S_2^N; Z_2^N|W_1\right) + I\left(\tilde{Y}_2^N, S_1^N; Z_1^N|W_2\right) \\
&\overset{(g)}{\leq} T' + I\left(\tilde{Y}_1^N; W_2|W_1, Z_2^N\right) + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I\left(Z_2^N; \tilde{Y}_1^N, W_2, \tilde{Y}_2^N, Z_1^N|W_1\right) \\
&+ I\left(Z_1^N; \tilde{Y}_2^N, W_1, \tilde{Y}_1^N, Z_2^N|W_2\right) \\
&\overset{(h)}{=} T' + I\left(\tilde{Y}_1^N; W_2|W_1, Z_2^N\right) + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I\left(\tilde{Y}_2^N; Z_2^N|W_1, W_2, Z_1^N\right) + I\left(\tilde{Y}_1^N; Z_1^N|W_1, W_2, Z_2^N\right) \\
&+ I\left(\tilde{Y}_1^N; Z_2^N|W_1, W_2, Z_1^N, \tilde{Y}_2^N\right) \\
&+ I\left(\tilde{Y}_2^N; Z_1^N|W_1, W_2, Z_2^N, \tilde{Y}_1^N\right) \\
&\overset{(i)}{=} T' + I\left(\tilde{Y}_1^N; W_2|W_1, Z_2^N\right) + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I\left(\tilde{Y}_2^N; Z_2^N|W_1, W_2, Z_1^N\right) + I\left(\tilde{Y}_1^N; Z_1^N|W_1, W_2, Z_2^N\right) \\
&= h\left(Y_1^N|S_1^N\right) - h\left(Z_1^N\right) + h\left(Y_2^N|S_2^N\right) - h\left(Z_2^N\right) \\
&+ I\left(\tilde{Y}_1^N; W_2, Z_1^N|W_1, Z_2^N\right) + I\left(\tilde{Y}_2^N; W_1, Z_2^N|W_2, Z_1^N\right) \\
&\leq \sum_{i=1}^{N} \left[h\left(Y_{1i}|S_{1i}\right) - h\left(Z_{1i}\right)\right] + \sum_{i=1}^{N} \left[h\left(Y_{2i}|S_{2i}\right) - h\left(Z_{2i}\right)\right] \\
&+ H\left(\tilde{Y}_1^N|W_1, Z_2^N\right) + H\left(\tilde{Y}_2^N|W_2, Z_1^N\right) \\
&\leq \sum_{i=1}^{N} \left[h\left(Y_{1i}|S_{1i}\right) - h\left(Z_{1i}\right)\right] + \sum_{i=1}^{N} \left[h\left(Y_{2i}|S_{2i}\right) - h\left(Z_{2i}\right)\right] \\
&+ \sum_{i=1}^{N} H\left(\tilde{Y}_{1i}|X_{1i}\right) + \sum_{i=1}^{N} H\left(\tilde{Y}_{2i}|X_{2i}\right),
\end{aligned}
\tag{5.27}
$$

where (a) follows from Fano's inequality; (b) follows from the fact that $h(Y_1^N|W_1) = h(S_2^N|W_1)$ and $h(Y_2^N|W_2) = h(S_1^N|W_2)$ (see Claim 5.3 below); (c) follows from the non-negativity of mutual information and the fact that conditioning reduces entropy; (d) follows from the fact that X_k^N is a function of $(W_k, \tilde{Y}_k^{N-1})$, and the fact that W_1 and W_2 are independent; (e) follows from the fact that X_k^N is a function of $(W_k, \tilde{Y}_k^{N-1})$; (f) follows from the fact that Z_k^N is independent of W_1 and

W_2; (g) holds since S_k^N is a function of $(W_k, \tilde{Y}_k^{N-1}, Z_{3-k}^N)$; ($h$) holds since W_1, W_2, Z_1^N, and Z_2^N are mutually independent; (i) holds since

$$\begin{aligned}
& I\left(\tilde{Y}_1^N; Z_2^N | W_1, W_2, Z_1^N, \tilde{Y}_2^N\right) \\
& \quad = \sum_{i=1}^{N} I\left(\tilde{Y}_{1i}; Z_2^N | W_1, W_2, Z_1^N, \tilde{Y}_2^N, \tilde{Y}_1^{i-1}\right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{N} I\left(\tilde{Y}_{1i}; Z_2^N | W_1, W_2, Z_1^N, \tilde{Y}_2^N, \tilde{Y}_1^{i-1}, X_2^N, X_1^i\right) \\
&= \sum_{i=1}^{N} I\left(\tilde{Y}_{1i}; Z_2^N | W_1, W_2, Z_1^N, \tilde{Y}_2^N, \tilde{Y}_1^{i-1}, X_2^N, X_1^i, Y_1^i\right) \\
&= \sum_{i=1}^{N} I\left(\tilde{Y}_{1i}; Z_2^N | W_1, W_2, Z_1^N, \tilde{Y}_2^N, \tilde{Y}_1^{i}, X_2^N, X_1^i, Y_1^i\right) \\
&= 0. \qquad (5.28)
\end{aligned}$$

Claim 5.3 $h(S_1^N | W_2) = h(Y_2^N | W_2)$.

Proof:

$$\begin{aligned}
h(Y_2^N | W_2) &= \sum h(Y_{2i} | Y_2^{i-1}, W_2) \\
&\overset{(a)}{=} \sum h(S_{1i} | Y_2^{i-1}, W_2) \\
&\overset{(b)}{=} \sum h(S_{1i} | Y_2^{i-1}, W_2, X_2^i, S_1^{i-1}) \\
&\overset{(c)}{=} \sum h(S_{1i} | W_2, S_1^{i-1}) = h(S_1^N | W_2),
\end{aligned}$$

where (a) follows from the fact that Y_{2i} is a function of (X_{2i}, S_{1i}) and X_{2i} is a function of (W_2, Y_2^{i-1}); (b) follows from the fact that X_2^i is a function of (W_2, Y_2^{i-1}) and S_1^{i-1} is a function of (Y_2^{i-1}, X_2^{i-1}); (c) follows from Claim 5.4 (see below). ■

Claim 5.4 *For all $i \geq 1$, X_1^i is a function of (W_1, S_2^{i-1}) and X_2^i is a function of (W_2, S_1^{i-1}).*

Proof: By symmetry, it is enough to prove only one. Notice that X_2^i is a function of (W_2, Y_2^{i-1}) and Y_2^{i-1} is a function of (X_2^{i-1}, S_1^{i-1}). Hence, X_2^i is a function of $(W_2, X_2^{i-1}, S_1^{i-1})$. Iterating the same argument, we conclude that X_2^i is a function of (W_2, X_{21}, S_1^{i-1}). Since X_{21} depends only on W_2, we complete the proof. ■

From the above, we get

$$\begin{aligned} R_1 + R_2 \leq\ & h(Y_1|S_1) - h(Z_1) + h(Y_2|S_2) - h(Z_2) \\ & + C_{\mathsf{FB1}} + C_{\mathsf{FB2}}. \end{aligned} \tag{5.29}$$

where we have used the fact that $H(\tilde{Y}_{ki}|X_{ki}) \leq C_{\mathsf{FBk}}$ and conditioning reduces entropy.

Finally note that for $\rho = \mathbb{E}[X_1 X_2^*]$, we have[2]

$$h(Y_1|S_1) \leq \log 2\pi e \left(1 + (1-\rho^2)\mathsf{INR}_{21} + \frac{(1-\rho^2)\mathsf{SNR}_1}{1 + (1-\rho^2)\mathsf{INR}_{12}}\right). \tag{5.30}$$

Using (5.30), we get the desired upper bound in (5.26g). ■

If we consider the symmetric channel gains, *i.e.*,

$$\begin{aligned} |h_{11}| &= |h_{22}| = |h_d|, \\ |h_{12}| &= |h_{21}| = |h_c|, \end{aligned} \tag{5.31}$$

and

$$\begin{aligned} \mathsf{SNR}_1 &= \mathsf{SNR}_2 = \mathsf{SNR} = |h_d|^2, \\ \mathsf{INR}_{12} &= \mathsf{INR}_{21} = \mathsf{INR} = |h_c|^2, \end{aligned} \tag{5.32}$$

we get the following outer-bound result.

Corollary 5.5 *The sum-rate capacity of the symmetric Gaussian IC with rate-limited feedback is included by the set $\bar{C}_{\mathsf{sym}}$ of $R_1 + R_2$ satisfying*

$$R_1 + R_2 \quad \leq \quad 2\log(1 + \mathsf{SNR}) + C_{\mathsf{FB1}} + C_{\mathsf{FB2}} \tag{5.33a}$$

[2] ρ captures the power gain that can be achieved by making the transmit signals correlated.

$$
\begin{aligned}
R_1 + R_2 &\leq \log\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) \qquad (5.33b)\\
&+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right)\\
R_1 + R_2 &\leq 2\log\left(1 + \mathsf{INR} + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) \qquad (5.33c)\\
&+ C_{\mathsf{FB1}} + C_{\mathsf{FB2}}.
\end{aligned}
$$

Proof: The proof is straight forward and is a direct consequence of the bounds in (5.26c), (5.26f), (5.26g), and (5.26h). For instance, (5.33a) is derived by combining (5.26c) and (5.26f) for $\rho = 0$. Note that $\rho = 0$ maximizes (5.26c) and (5.26f). ■

5.6.2 Acievability Strategy

We first provide a brief outline of the achievability. Our achievable scheme is based on block Markov encoding with backward decoding where the scheme is implemented over B blocks. In each block (with the exception of the last two), new messages are transmitted. At the end of a block, each receiver creates a feedback signal and sends it back to its corresponding transmitter. This will provide each transmitter with part of the other user's information that caused interference. Each transmitter encodes this interfering message and transmit it to its receiver during a different block. Through this part of the transmitted signal, receivers will be able to complete the decoding of the previously received messages. During the last two blocks, no new messages will be transmitted and each transmitter provides its receiver with the interfering message coming from the other transmitter. Later, we let B go to infinity to get our desired result.

As we have seen in Section 5.3, each receiver may need to decode the superposition of the two codewords (corresponding to the other user's cooperative common message and part of its own private message). In order to accomplish this in the Gaussian case, we employ lattice codes.

Lattice Coding Preliminaries

We briefly go over some preliminaries on lattice coding and summerize the results that will be used later. A lattice is a discrete additive subgroup of $\mathbb{R}^n$. The fundamental volume $V_f(\Lambda)$ of a lattice Λ is the reciprocal of the number of lattice points per unit volume.

Given integer p, denote the set of integers modulo p by $\mathbb{Z}_p$. Let $\mathbb{Z}^n \to \mathbb{Z}_p^n$: $v \mapsto \bar{v}$ be the componentwise modulo p operation over integer vectors. Also, let C be a linear (n, k) code over $\mathbb{Z}_p$. The lattice Λ_C defined as

$$\Lambda_C = \{v \in \mathbb{Z}^n : \bar{v} \in C\}, \tag{5.34}$$

is generated with respect to the linear code C (see [34] for details). In [34], it has been shown that there exists good lattice codes for point-to-point communication channels, *i.e.*, codes that achieve a rate close to the capacity of the channel with arbitrary small decoding error probability. We summarize the result here.

Consider a point-to-point communication scenario over an additive noise channel

$$Y = X + Z, \tag{5.35}$$

where X is the transmitted signal with power constraint P, Y is the received signal and Z is the additive noise process with zero mean and variance σ^2.

A set $\mathcal{B}$ of linear codes over $\mathbb{Z}_p$ is called balanced if every nonzero element of $\mathbb{Z}_p^n$ is contained in the same number of codes in $\mathcal{B}$. Define $\mathcal{L}_{\mathcal{B}}$ as

$$\mathcal{L}_{\mathcal{B}} = \{\Lambda_C : C \in \mathcal{B}\}. \tag{5.36}$$

Lemma 5.1 ([34]) *Consider a point-to-point additive noise channel described in (5.35). Let $\mathcal{B}$ be a balanced set of linear (n,k) codes over $\mathbb{Z}_p$. Averaged over all lattices from the set $\mathcal{L}_{\mathcal{B}}$ defined in (5.36), each scaled by $\gamma > 0$ and with a fundamental volume V, we have that for any $\delta > 0$, the average probability of decoding error is bounded by*

$$\bar{P}_e < (1+\delta)\frac{n\frac{1}{2}\log\left(2\pi e\sigma^2\right)}{V}, \tag{5.37}$$

for sufficiently large p and small γ such that $\gamma^n p^{n-k} = V$.

See [34] for the proof. The next lemma describes the existence of a good lattice code for a point-to-point AWGN channel.

Lemma 5.2 ([34]) *Consider a point-to-point additive noise channel described in (5.35) such that the transmitter satisfies a power constraint of P. Then, we can choose a lattice Λ generated using construction A, a shift s[3] and a shaping region S[4] such that the codebook $(\Lambda + s) \cap S$ achieves a rate R with arbitrarily small probability of error if*

$$R \leq \frac{1}{2}\log\left(\frac{P}{\sigma^2}\right). \tag{5.38}$$

In other words, Lemma 5.2 describes the existence of a lattice code with sufficient codewords. See [34] for the proof. For a more comprehensive review of lattice codes see [18, 34, 42].

[3]Shift s is a vector in $\mathbb{R}^n$ and it is required in order to prove of existence of good lattice codes, see [34] for more details.

[4]We need to consider the intersection of a lattice with some shaping region $S \subset \mathbb{R}^n$ to satisfy the power constraint.

Remark 5.6 *In this paper, we consider complex AWGN channels. Similar to Lemma 5.2, one can show that using lattice codes, a rate of* $\log\left(\frac{P}{\sigma^2}\right)$ *is achievable in the complex channel setting.*

Acievability Strategy for $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$

We describe our strategy for the extreme case where $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$ (interchanging user IDs, one can get similar results for $C_{\mathsf{FB2}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB1}} = 0$). Our strategy for any other feedback configuration will be based on a combination of the strategies for these extreme cases.

Codebook Generation and Encoding: The communication strategy consists of B blocks, each of length N channel uses. In block b, $b = 1, 2, \ldots, B-2$, transmitter 1 has four messages $W_1^{(1,b)}, W_1^{(2,b)}, W_1^{(3,b)}$ and $W_1^{(4,b)}$, where $W_1^{(i,b)} \in \{1, 2, \ldots, 2^{NR_1^{(i)}}\}$. Out of these four messages, $W_1^{(1,b)}, W_1^{(2,b)}$ and $W_1^{(4,b)}$ are new messages and in particular $W_1^{(1,b)}$ and $W_1^{(2,b)}$ form the private message of transmitter 1 while $W_1^{(4,b)}$ is the non-cooperative message (as it will be clarified shortly for the feedback strategy, the reason for splitting the private message of transmitter 1 into two parts is that in order to be able to use lattice codes, we would like the codeword corresponding to the cooperative common message of transmitter 2 to be received at the same power level as part of the codeword corresponding to the parivate message of transmitter 1). We will describe $W_1^{(3,b)}$ when we explain the feedback strategy. On the other hand, transmitter 2 has three new independent messages $W_2^{(1,b)}, W_2^{(2,b)}$ and $W_2^{(3,b)}$, the private, the cooperative common, and the non-cooperative common message of transmitter 2 respectively.

At transmitter k, message $W_k^{(i,b)}$ is mapped to a Gaussian codeword $X_k^{(i,b)}$

picked from a codebook of size $2^{NR_k^{(i)}}$ and any element of this codebook is drawn i.i.d. from $\mathcal{CN}(0, P_k^{(i)})$, $(k, i) \in \{(1,1),(1,3),(1,4),(2,1),(2,3)\}$. For notational simplicity, we have removed the superscript N.

Message $W_k^{(2,b)}$ is mapped to $X_k^{(2,b)}$ encoded by lattice $\Lambda_k^{(2,b)}$ with shift $s_k^{(2,b)}$ and spherical shaping region $S_k^{(2,b)}$. This gives a codebook of size $2^{NR_k^{(2)}}$ with power constraint of $P_k^{(2)}$, $k = 1, 2$. Denote this codebook by $(\Lambda_k^{(2,b)} + s_k^{(2,b)}) \cap S_k^{(2,b)}$.

Transmitter k will superimpose all of its transmitted signals to create $X_k^{(b)}$, its transmitted signal during block b, *i.e.*, $X_1^{(b)} = X_1^{(1,b)} + X_1^{(2,b)} + X_1^{(3,b)} + X_1^{(4,b)}$ and $X_2^{(b)} = X_2^{(1,b)} + X_2^{(2,b)} + X_2^{(3,b)}$.

The power assignments should be such that they are non-negative and satisfy the power constraint at each transmitter:

$$
\begin{aligned}
P_1 &= P_1^{(1)} + P_1^{(2)} + P_1^{(3)} + P_1^{(4)} \leq 1, \\
P_2 &= P_2^{(1)} + P_2^{(2)} + P_2^{(3)} \leq 1.
\end{aligned}
\tag{5.39}
$$

Feedback Strategy: Our feedback strategy is inspired by the motivating example in Section 5.3. Remember that in this example, receiver 2 had to feed back the superposition of the two codewords (corresponding to transmitter 1's cooperative common message and part of its private message). To realize this in the Gaussian case, we incorporate lattice coding with appropriate power assignment as part of our strategy.

We set $\mathsf{SNR}P_1^{(2)} = \mathsf{INR}P_2^{(2)}$, so that $X_1^{(2,b)}$ and $X_2^{(2,b)}$ arrive at the same power level at receiver 1 and therefore $h_d X_1^{(2,b)} + h_c X_2^{(2,b)}$ is a lattice point. We refer to this lattice index as $I_{\Lambda_{1,2}}^{(b)}$. Receiver 1 then feeds $\left(I_{\Lambda_{1,2}}^{(b)} \textbf{ mod } 2^{NC_{\mathsf{FB}}}\right)$ back to transmitter 1.

Given $\left(I_{\Lambda_{1,2}}^{(b)} \textbf{ mod } 2^{NC_{\mathsf{FB}}}\right)$, transmitter 1 removes $h_d X_1^{(2,b)}$ and decodes the message index of $W_2^{(2,b)}$. This can be done as long as the total number of lattice points for either of the two aligned messages is less than $2^{NC_{\mathsf{FB}}}$, *i.e.*, $R_1^{(2,b)}, R_2^{(2,b)} \leq C_{\mathsf{FB}}$. Since the feedback transmission itself lasts a block, we set $W_1^{(3,b+2)} = W_2^{(2,b)}$.

Decoding: For notational simplicity, we ignore the block index and from our description it is clear whether the two signals belong to the same block or different ones. We also use the following shorthand notation:

$$P_k^{(1:j)} = P_k^{(1)} + P_k^{(2)} + \ldots + P_k^{(j)} \quad k = 1, 2. \tag{5.40}$$

Our achievable scheme employs different decoding orders depending on the channel gains. In other words, based on the channel gains the number of required messages to achieve the desired sum-rate might vary. In fact based on the channel gains, it might be sufficient to consider fewer messages than suggested above. In such cases, we assume the unnecessary messages to be deterministic (*i.e.*, the corresponding rate to be zero). In particular, we have three different cases.

Case (a) $\log(\mathsf{INR}) \leq \frac{1}{2}\log(\mathsf{SNR})$**:**

In this case, we set $R_1^{(4)} = R_2^{(3)} = 0$. In other words, $W_1^{(4)}$ and $W_2^{(3)}$ are deterministic messages. We then get

$$Y_1 = h_d\left(X_1^{(1)} + X_1^{(2)} + X_1^{(3)}\right) + h_c\left(X_2^{(1)} + X_2^{(2)}\right) + Z_1. \tag{5.41}$$

At the end of each block, receiver 1 first decodes $X_1^{(3)}$ by treating all other codewords as noise. $X_1^{(3)}$ can be decoded with small error probability if

$$R_1^{(3)} \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(3)}}{1 + \mathsf{INR}P_2 + \mathsf{SNR}P_1^{(1:2)}}\right). \tag{5.42}$$

It then removes $h_d X_1^{(3)}$ from the received signal and decodes $X_1^{(1)}$ by treating other codewords as noise. $X_1^{(1)}$ is decodable at receiver 1 with arbitrary small error probability if

$$R_1^{(1)} \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1 + \mathsf{INR}P_2 + \mathsf{SNR}P_1^{(2)}}\right). \tag{5.43}$$

After removing $h_d X_1^{(1)}$, receiver 1 has access to $h_d X_1^{(2)} + h_c X_2^{(2)} + h_c X_2^{(1)} + Z_1$. Since we have set $\mathsf{SNR}P_1^{(2)} = \mathsf{INR}P_2^{(2)}$, $h_d X_1^{(2)} + h_c X_2^{(2)}$ is a lattice point with some index $I_{\Lambda_{1,2}}^{(b)}$. Receiver 1 decodes $I_{\Lambda_{1,2}}^{(b)}$ by treating other codewords as noise, and sends back $\left(I_{\Lambda_{1,2}}^{(b)} \textbf{ mod } 2^{NC_{\mathsf{FB}}}\right)$ to transmitter 1. From Lemma 5.2, decoding with arbitrary small error probability is feasible if

$$\begin{aligned} R_1^{(2)} &\leq \left[\log\left(\frac{\mathsf{SNR}P_1^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)\right]^+, \\ R_2^{(2)} &\leq \left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)\right]^+, \end{aligned} \tag{5.44}$$

Here $[\cdot]^+ = \max\{\cdot, 0\}$.

The decoding at receiver 2 proceeds as follows. At the end of each block, receiver 2 removes $h_c X_1^{(3)}$ from its received signal. Note that $X_1^{(3)}$ is in fact a function of $W_2^{(2,b-2)}$ and thus it is known to receiver 2 (assuming successful decoding in the previous blocks). Therefore, after removing $h_c X_1^{(3)}$, we get

$$Y_2 = h_d\left(X_2^{(1)} + X_2^{(2)}\right) + h_c\left(X_1^{(1)} + X_1^{(2)}\right) + Z_2. \tag{5.45}$$

Receiver 2 now decodes $X_2^{(2)}$ and $X_2^{(1)}$ by treating other codewords as noise. This can be done with arbitrary small error probability if

$$\begin{aligned} R_2^{(2)} &\leq \left[\log\left(\frac{\mathsf{SNR}P_2^{(2)}}{1 + \mathsf{SNR}P_2^{(1)} + \mathsf{INR}P_1^{(1:2)}}\right)\right]^+, \\ R_2^{(1)} &\leq \log\left(1 + \frac{\mathsf{SNR}P_2^{(1)}}{1 + \mathsf{INR}P_1^{(1:2)}}\right). \end{aligned} \tag{5.46}$$

The decoding strategy presented above describes a set of constraints on the rates, which is summarized as follows:

$$\begin{cases} R_1^{(1)} & \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1+\mathsf{INR}P_2+\mathsf{SNR}P_1^{(2)}}\right) \\ R_1^{(2)} & \leq \min\left\{\log\left(\frac{\mathsf{SNR}P_1^{(2)}}{1+\mathsf{INR}P_2^{(1)}}\right)^+, C_{\mathsf{FB}}\right\} \\ R_1^{(3)} & \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(3)}}{1+\mathsf{INR}P_2+\mathsf{SNR}P_1^{(1:2)}}\right) \\ R_2^{(1)} & \leq \log\left(1 + \frac{\mathsf{SNR}P_2^{(1)}}{1+\mathsf{INR}P_1^{(1:2)}}\right) \\ R_2^{(2)} & \leq \min\left\{\left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1+\mathsf{INR}P_2^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \end{cases} \tag{5.47}$$

Therefore, we can achieve a sum-rate

$$R_{\mathsf{SUM}}^{(a)} = R_1^{(1)} + R_1^{(2)} + R_2^{(1)} + R_2^{(2)},$$

arbitrary close to[5]

$$\begin{aligned} R_{\mathsf{SUM}}^{(a)} = \log&\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1 + \mathsf{INR}P_2 + \mathsf{SNR}P_1^{(2)}}\right) \\ &+ \min\left\{\left[\log\left(\frac{\mathsf{SNR}P_1^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \\ &+ \log\left(1 + \frac{\mathsf{SNR}P_2^{(1)}}{1 + \mathsf{INR}P_1^{(1:2)}}\right) \\ &+ \min\left\{\left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\}. \end{aligned} \tag{5.48}$$

Case (b) $\frac{1}{2}\log(\mathsf{SNR}) \leq \log(\mathsf{INR}) \leq \frac{2}{3}\log(\mathsf{SNR})$**:** In this case, we have

$$\begin{aligned} Y_1 = h_d&\left(X_1^{(1)} + X_1^{(2)} + X_1^{(3)} + X_1^{(4)}\right) \\ &+ h_c\left(X_2^{(1)} + X_2^{(2)} + X_2^{(3)}\right) + Z_1. \end{aligned} \tag{5.49}$$

[5]Note that $X_1^{(3,b)}$ is a function of the cooperative common message of transmitter 2, *i.e.*, $W_2^{(2,b-2)}$, hence, it does not contain any new information and it is not considered in the sum-rate.

At the end of each block, receiver 1 first decodes $X_1^{(4)}$ by treating all other codewords as noise, and removes $h_d X_1^{(4)}$ from the received signal. This can be decoded with small error probability if

$$R_1^{(4)} \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(4)}}{1 + \mathsf{INR}P_2 + \mathsf{SNR}P_1^{(1:3)}}\right). \tag{5.50}$$

Next, it decodes $X_1^{(3)}$ by treating other codewords as noise and removes $h_d X_1^{(3)}$ from the received signal. This can be decoded with arbitrary small error probability if

$$R_1^{(3)} \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(3)}}{1 + \mathsf{INR}P_2 + \mathsf{SNR}P_1^{(1:2)}}\right). \tag{5.51}$$

We proceed by decoding the non-cooperative common message of transmitter 2, *i.e.*, $X_2^{(3)}$ by treating other codewords as noise. This can be decoded with arbitrary small error probability if

$$R_2^{(3)} \leq \log\left(1 + \frac{\mathsf{INR}P_2^{(3)}}{1 + \mathsf{INR}P_2^{(1:2)} + \mathsf{SNR}P_1^{(1:2)}}\right). \tag{5.52}$$

It then removes $h_c X_2^{(3)}$ from the received signal, having now access to $h_d X_1^{(2)} + h_c X_2^{(2)} + h_d X_1^{(1)} + h_c X_2^{(1)} + Z_1$. We decode the lattice index of $h_d X_1^{(2)} + h_c X_2^{(2)}$, *i.e.*, $I_{\Lambda_{1,2}}^{(b)}$, by treating other codewords as noise. It then sends back $\left(I_{\Lambda_{1,2}}^{(b)} \ \mathbf{mod}\ 2^{NC_{\mathsf{FB}}}\right)$ to transmitter 1. From Lemma 5.2, decoding with arbitrary small error probability is feasible if

$$\begin{aligned} R_1^{(2)} &\leq \left[\log\left(\frac{\mathsf{SNR}P_1^{(2)}}{1 + \mathsf{INR}P_2^{(1)} + \mathsf{SNR}P_1^{(1)}}\right)\right]^+, \\ R_2^{(2)} &\leq \left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)} + \mathsf{SNR}P_1^{(1)}}\right)\right]^+. \end{aligned} \tag{5.53}$$

After decoding and removing $h_d X_1^{(2)} + h_c X_2^{(2)}$, receiver 1 decodes $X_1^{(1)}$. This can be done with arbitrary small error probability if

$$R_1^{(1)} \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1 + \mathsf{INR}P_2^{(1)}}\right). \tag{5.54}$$

Similar to the previous case, receiver 2 removes $X_1^{(3)}$ from its received signal. The decoding at receiver 2 proceeds as follows. Receiver 2 decodes $X_2^{(3)}$ by treating other codewords as noise and removes $h_d X_2^{(3)}$ from the received signal. Next, $X_2^{(2)}$, the non-cooperative common message of transmitter 1, will be decoded while treating other codewords as noise. Receiver 2 removes $h_d X_2^{(2)}$ from the received signal and then, decodes $X_1^{(4)}$ by treating other codewords as noise. After removing $h_c X_1^{(4)}$, we now decode the private message of transmitter 2, *i.e.*, $X_2^{(1)}$. This can be done with arbitrary small error probability if

$$\begin{aligned} R_2^{(3)} &\leq \log\left(1 + \frac{\mathsf{SNR} P_2^{(3)}}{1 + \mathsf{SNR} P_2^{(1:2)} + \mathsf{INR}(P_1 - P_1^{(3)})}\right), \\ R_2^{(2)} &\leq \left[\log\left(\frac{\mathsf{SNR} P_2^{(2)}}{1 + \mathsf{SNR} P_2^{(1)} + \mathsf{INR}(P_1 - P_1^{(3)})}\right)\right]^+, \\ R_1^{(4)} &\leq \log\left(1 + \frac{\mathsf{INR} P_1^{(4)}}{1 + \mathsf{INR} P_1^{(1:2)} + \mathsf{SNR} P_2^{(1)}}\right), \\ R_2^{(1)} &\leq \log\left(1 + \frac{\mathsf{SNR} P_2^{(1)}}{1 + \mathsf{INR} P_1^{(1:2)}}\right). \end{aligned} \tag{5.55}$$

The decoding strategy presented above describes a set of constraints on the rates, which is summarized as follows:

$$\begin{cases} R_1^{(1)} & \leq \log\left(1 + \frac{\mathsf{SNR} P_1^{(1)}}{1 + \mathsf{INR} P_2^{(1)}}\right) \\ R_1^{(2)} & \leq \min\left\{\left[\log\left(\frac{\mathsf{SNR} P_1^{(2)}}{1 + \mathsf{INR} P_2^{(1)} + \mathsf{SNR} P_1^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \\ R_1^{(3)} & \leq \log\left(1 + \frac{\mathsf{SNR} P_1^{(3)}}{1 + \mathsf{INR} P_2 + \mathsf{SNR} P_1^{(1:2)}}\right) \\ R_1^{(4)} & \leq \log\left(1 + \frac{\mathsf{INR} P_1^{(4)}}{1 + \mathsf{INR} P_1^{(1:2)} + \mathsf{SNR} P_2^{(1)}}\right) \\ R_2^{(1)} & \leq \log\left(1 + \frac{\mathsf{SNR} P_2^{(1)}}{1 + \mathsf{INR} P_1^{(1:2)}}\right) \\ R_2^{(2)} & \leq \min\left\{\left[\log\left(\frac{\mathsf{INR} P_2^{(2)}}{1 + \mathsf{INR} P_2^{(1)} + \mathsf{SNR} P_1^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \\ R_2^{(3)} & \leq \log\left(1 + \frac{\mathsf{INR} P_2^{(3)}}{1 + \mathsf{INR} P_2^{(1:2)} + \mathsf{SNR} P_1^{(1:2)}}\right). \end{cases} \tag{5.56}$$

Therefore, we can achieve a sum-rate

$$R_{\mathsf{SUM}}^{(b)} = R_1^{(1)} + R_1^{(2)} + R_1^{(4)} + R_2^{(1)} + R_2^{(2)} + R_2^{(3)},$$

arbitrary close to[6]

$$R_{\mathsf{SUM}}^{(b)} = \log\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1 + \mathsf{INR}P_2^{(1)}}\right) \tag{5.57}$$

$$\begin{aligned}
&+ \min\left\{\left[\log\left(\frac{\mathsf{SNR}P_1^{(2)}}{1 + \mathsf{INR}P_2^{(1)} + \mathsf{SNR}P_1^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \\
&+ \log\left(1 + \frac{\mathsf{INR}P_1^{(4)}}{1 + \mathsf{INR}P_1^{(1:2)} + \mathsf{SNR}P_2^{(1)}}\right) \\
&+ \log\left(1 + \frac{\mathsf{SNR}P_2^{(1)}}{1 + \mathsf{INR}P_1^{(1:2)}}\right) \\
&+ \min\left\{\left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)} + \mathsf{SNR}P_1^{(1)}}\right)\right]^+, C_{\mathsf{FB}}\right\} \\
&+ \log\left(1 + \frac{\mathsf{INR}P_2^{(3)}}{1 + \mathsf{INR}P_2^{(1:2)} + \mathsf{SNR}P_1^{(1:2)}}\right).
\end{aligned}$$

Case (c) $2\log(\mathsf{SNR}) \leq \log(\mathsf{INR})$**:**

In this case, there is no need to decode the superposition of the two messages. So set $R_1^{(1)}, R_1^{(2)}$ and $R_2^{(1)}$ equal to zero. We then get

$$Y_1 = h_d\left(X_1^{(3)} + X_1^{(4)}\right) + h_c\left(X_2^{(2)} + X_2^{(3)}\right) + Z_1, \tag{5.58}$$

$$Y_2 = h_d\left(X_2^{(2)} + X_2^{(3)}\right) + h_c\left(X_1^{(3)} + X_1^{(4)}\right) + Z_2. \tag{5.59}$$

As for the feedback strategy, receiver 1 decodes $X_2^{(2)}$ by treating other codewords as noise, and sends the lattice index of $W_2^{(2)}$ back to transmitter 1 during the following block. Transmitter 1 later encodes this message as $X_1^{(3)}$ and transmits it. It is worth mentioning that in this case, it is in fact receiver 2 who wants to exploit the feedback link of user 1 to get part of its message. In other words, we have two paths for information flow from transmitter 2 to receiver 2;

[6]Note that $X_1^{(3,b)}$ is a function of the cooperative common message of transmitter 2, *i.e.*, $W_2^{(2,b-2)}$, hence it is not considered in the sum-rate.

one through the direct link between them and the other one through receiver 1, feedback link and transmitter 1. The decoding works very similar to what we described above and we get the following set of constraints to guarantee small error probability at the decoders.

$$\begin{cases} R_1^{(3)} & \leq \log\left(1 + \frac{\mathsf{INR}P_1^{(3)}}{1+\mathsf{SNR}P_2}\right) \\ R_1^{(4)} & \leq \log\left(1 + \frac{\mathsf{SNR}P_1^{(4)}}{1+\mathsf{SNR}P_1^{(3)}}\right) \\ R_2^{(2)} & \leq \min\left\{\left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1+\mathsf{SNR}P_1}\right)\right]^+, C_{\mathsf{FB}}\right\} \\ R_2^{(3)} & \leq \log\left(1 + \frac{\mathsf{SNR}P_2^{(3)}}{1+\mathsf{SNR}P_2^{(2)}}\right) \end{cases} \tag{5.60}$$

As before $X_1^{(3,b)}$ is a function of $W_2^{(2,b-2)}$. Therefore, we can achieve a sum-rate

$$R_{\mathsf{SUM}}^{(c)} = R_1^{(4)} + R_2^{(2)} + R_2^{(3)},$$

arbitrary close to

$$\begin{aligned} R_{\mathsf{SUM}}^{(c)} = \log&\left(1 + \frac{\mathsf{SNR}P_1^{(4)}}{1 + \mathsf{SNR}P_1^{(3)}}\right) \\ &+ \min\left\{\left[\log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{SNR}P_1}\right)\right]^+, C_{\mathsf{FB}}\right\} \\ &+ \log\left(1 + \frac{\mathsf{SNR}P_2^{(3)}}{1 + \mathsf{SNR}P_2^{(2)}}\right). \end{aligned} \tag{5.61}$$

Case (d) $\frac{2}{3}\log(\mathsf{SNR}) \leq \log(\mathsf{INR}) \leq 2\log(\mathsf{SNR})$**:**

As we will show in Appendix D.4, in this regime feedback can at most increase the sum-rate capacity by 4 bits/sec/Hz. Hence, we ignore the feedback and use the non-feedback transmission strategy in [19] (*i.e.*, having only one private and one common message at each transmitter and jointly decoding at receivers).

General Feedback Assignment

We now describe our achievable scheme for general feedback capacity assignment based on a combination of the achievability schemes for the extreme cases. Let $C_{\mathsf{FB1}} = \lambda C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = (1-\lambda)C_{\mathsf{FB}}$, such that $0 \leq \lambda \leq 1$. We call the achievable sum-rate of the extreme case $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$ by $R_{\mathsf{SUM}}^{C_{\mathsf{FB2}}=0}$, and similarly, we refer to the achievable sum-rate of the other extreme case by $R_{\mathsf{SUM}}^{C_{\mathsf{FB1}}=0}$. We split

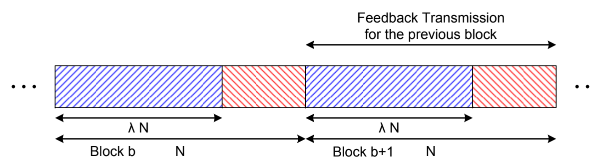


Figure 5.6: Achievability strategy for $C_{\mathsf{FB1}} = \lambda C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = (1-\lambda)C_{\mathsf{FB}}$.

any block b, $b = 1, 2, \ldots, B-2$, of length N into two sub-blocks: b_1 of length λN and b_2 of length $(1-\lambda)N$. See Figure 5.6 for a depiction. During block b_1, we implement the transmission strategy of the extreme case $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$, with a block length of λN; and during block b_2, the achievability scheme of the extreme case $C_{\mathsf{FB1}} = 0$ and $C_{\mathsf{FB2}} = C_{\mathsf{FB}}$, with a block length of $(1-\lambda)N$.

At the end of each sub-block, receivers decode the messages as described before and create the feedback messages. During block $b+1$ the feedback messages of sub-blocks b_1 and b_2 will be sent back to corresponding transmitters, as shown in Figure 5.6. Note that we use C_{FB1} during the entire length of block $b+1$, hence the effective feedback rate of user 1 (total feedback use divided by number of transmission time slots), would be

$$C_{\mathsf{FB1}}^{\mathsf{eff}} = \frac{NC_{\mathsf{FB1}}}{\lambda N} = \frac{\lambda N C_{\mathsf{FB}}}{\lambda N} = C_{\mathsf{FB}}. \tag{5.62}$$

Hence, we can implement the achievability strategy corresponding to the extreme case $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$. Similar argument is valid for the other extreme case. With this achievability scheme, as N goes to infinity, we achieve a sum-rate of $\lambda R_{\mathsf{SUM}}^{C_{\mathsf{FB2}}=0} + (1-\lambda) R_{\mathsf{SUM}}^{C_{\mathsf{FB1}}=0}$.

Power Splitting

We have yet to specify the values of the powers associated with the codewords at the transmitters (*i.e.*, $P_k^{(i)}$: $k \in \{1,2\}$, i $\in \{1,2,3,4\}$). In general, one can solve an optimization problem to find the optimal choice of power level assignments that maximizes the achievable sum-rate. We have performed numerical analysis for this optimization problem. Figure 5.7 shows the gap between our proposed achievable scheme and the outer-bounds in Corollary 5.5 at (a) $\mathsf{SNR} = 20\mathsf{dB}$, (b) $\mathsf{SNR} = 40\mathsf{dB}$, and (c) $\mathsf{SNR} = 60\mathsf{dB}$, for $C_{FB} = 10$ bits. In fact through our numerical analysis, we can see that the gap is at most 4, 5, and 5.5 bits/sec/Hz for the given values of SNR, respectively. Note that sharp points in Figure 5.7 are due to the change of achievability scheme for different values of INR as described before.

In Appendix D.4, we present an explicit choice of power assignments such that the gap between the achievability scheme and the outer-bounds does not scale with SNR. As a result, we get the following Theorem.

Theorem 5.6 *The sum-rate capacity of the Gaussian IC with rate-limited feedback is within at most* 14.8 *bits/sec/Hz of the maximum* $R_1 + R_2$ *satisfying*

$$0 \leq R_1 + R_2 \leq 2\log(1+\mathsf{SNR}) + C_{\mathsf{FB1}} + C_{\mathsf{FB2}} \tag{5.63a}$$

$$0 \leq R_1 + R_2 \leq \log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) \tag{5.63b}$$

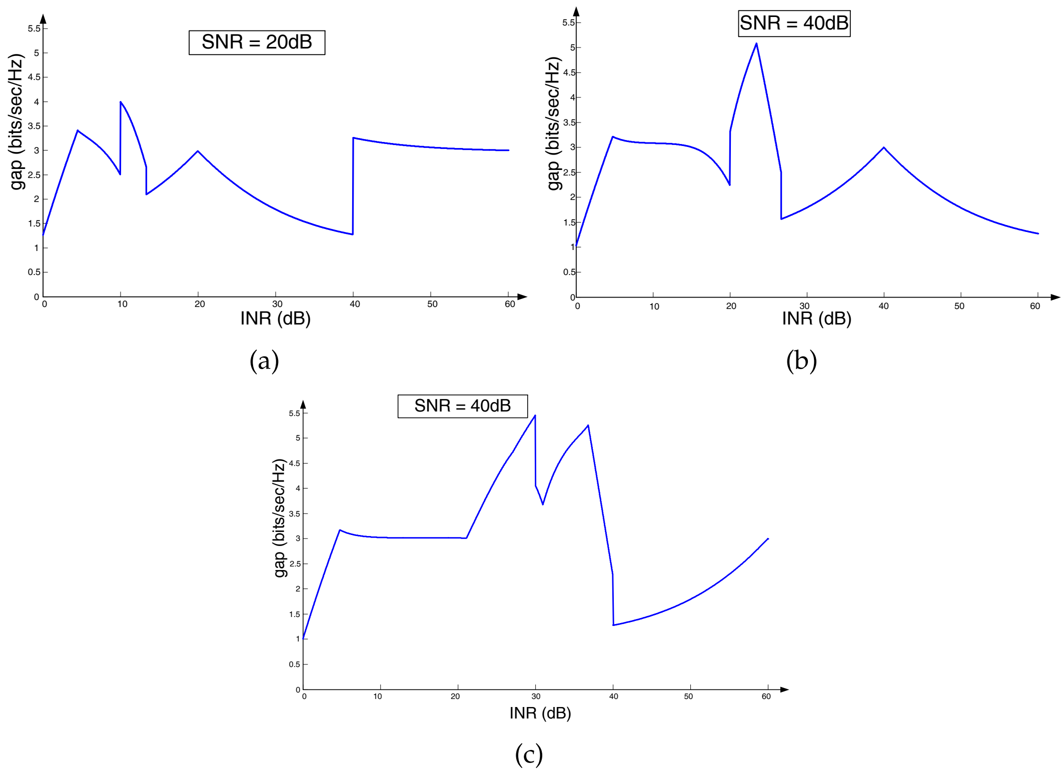


Figure 5.7: Numerical analysis: gap between achievable scheme and the outer-bounds in Corollary 5.5 at (a) SNR = 20dB, (b) SNR = 40dB, and (c) SNR = 60dB for $C_{FB} = 10$ bits.

$$\begin{aligned} & + \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right) \\ 0 \leq R_1 + R_2 & \leq 2\log\left(1 + \mathsf{INR} + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) \\ & + C_{\mathsf{FB1}} + C_{\mathsf{FB2}}. \end{aligned} \tag{5.63c}$$

Remark 5.7 *Note that the given choice of power assignment in Appendix D.4 is not necessarily optimal, and our analysis is pessimistic in the sense that we consider the worst case scenario, and we calculate the gap for the worst case.*

As a corollary, we characterize the symmetric capacity of the two-user Gaussian IC with rate-limited feedback, as defined below, to within a constant number of bits.

Definition 5.1 *The symmetric capacity is defined by*

$$C_{\mathsf{sym}} = \sup\{R : (R, R) \in C\}, \tag{5.64}$$

where C *is the capacity region.*

Corollary 5.7 *For the symmetric Gaussian IC with equal feedback link capacities,* i.e., $C_{\mathsf{FB1}} = C_{\mathsf{FB2}}$, *the presented achievability strategy achieves to within at most* 7.4 *bits/sec/Hz/user to the symmetric capacity* C_{sym} *defined in (5.64), for all channel gains.*

Proof: Theorem 5.6 says that we can achieve to within at most 14.8 bits/sec/Hz of the outerbounds in Corollary 5.5 for any feedback assignment. Therefore, in symmetric IC with equal feedback link capacities $C_{\mathsf{FB1}} = C_{\mathsf{FB2}} = \frac{1}{2}C_{\mathsf{FB1}}$, the gap between the achievability and the symmetric capacity is at most 7.4 bits/sec/Hz/user. ■

5.7 Concluding Remarks

We have addressed the two-user interference channel with rate-limited feedback under three different models: the El Gamal-Costa deterministic model [16], the linear deterministic model [7], and the Gaussian model. We developed new achievable schemes and new outer-bounds for all of the three models. We showed the optimality of our scheme under the linear deterministic model. Under the Gaussian model, we established new outer-bounds on the capacity region with rate-limited feedback, and we proposed a transmission strategy employing lattice codes and the ideas developed in the first two models. Furthermore, we proved that the gap between the achievable sum-rate of the proposed scheme and the outer-bound is bounded by a constant number of bits, independent of the channel gains.

One of the future directions would be to extend this result to the capacity region of the asymmetric two-user Gaussian interference channel with rate-limited feedback. The same achievability scheme can be applied there, however, the gap analysis will be cumbersome. Therefore, one interesting direction is to find out new techniques to bound the gap between the achievable region and the outer-bounds on the capacity region of the asymmetric two-user Gaussian interference channel with rate-limited feedback.

APPENDIX A

CHAPTER 1 OF APPENDIX

A.1 Achievability Proof of Theorem 2.1 [Instantaneous-CSIT]

In this appendix, we provide the achievability proof of Theorem 2.1. Below, we have stated the capacity region of the two-user BFIC with Instantaneous-CSIT (and no OFB).

$$C^{\mathrm{ICSIT}} = \begin{cases} 0 \leq R_i \leq p, & i = 1, 2, \\ R_1 + R_2 \leq 1 - q^2 + pq. \end{cases} \tag{A.1}$$

Remark A.1 *For* $0 \leq p \leq 0.5$*, the capacity region is given by*

$$C^{\mathrm{ICSIT}} = \left\{ R_1, R_2 \in \mathbb{R}^+ \ s.t. \ R_i \leq p, \ i = 1, 2 \right\}. \tag{A.2}$$

while for $0.5 < p \leq 1$*, the outer-bound on* $R_1 + R_2$ *is also active, see Figure A.1.*

With Instantaneous-CSIT, each transmitter knows what channel realization occurs at the time of transmission. Transmitters can take advantage of such knowledge and by pairing different realizations, the optimal rate region as given in Theorem 2.1 can be achieved. We will first describe the achievability strategy for $0 \leq p \leq 0.5$, since it is easier to follow. We then complete the proof by describing the achievability strategy for $0.5 < p \leq 1$.

A.1.1 Achievabiliy Strategy for $0 \leq p \leq 0.5$

Note that the result for $p = 0$ is trivial, so we assume $0 < p \leq 0.5$. Below, we describe the possible pairing opportunities that are useful in this regime and

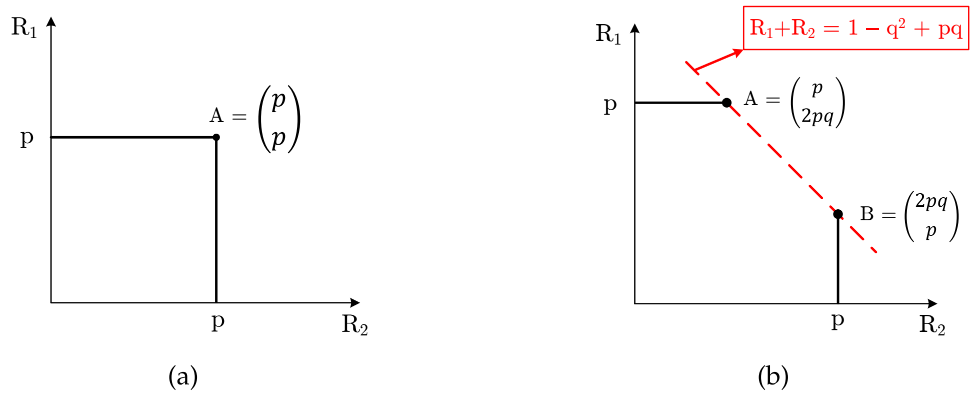


Figure A.1: *Capacity region of the two-user BFIC with Instantaneous-CSIT and for (a)* $0 \le p \le 0.5$*, and (b)* $0.5 < p \le 1$.

then, we describe the achievability scheme. The possible pairing opportunities are as follows.

- Type A [Cases 1 and 15]: In Case 15, only the corss links are equal to 1, therefore, by pairing bits in Case 1 with bits in Case 15, we can cancel out interference in Case 1, see Figure A.2. In other words by pairing the two cases, we can communicate 2 bits interference free.

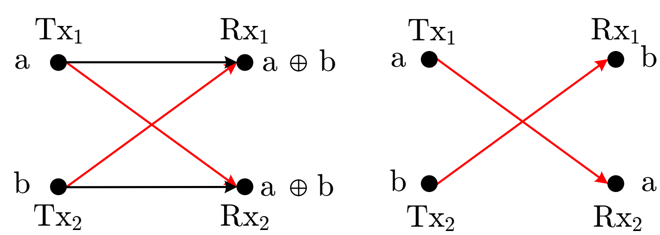


Figure A.2: Pairing opportunity Type A: By pairing Cases 1 and 15, we can communicate two bits interference-free. For instance, receiver one has access to bits $a \oplus b$ and b and as a result, it can decode its desired bit.

- Type B [Cases 2 and 14]: We can pair up Cases 2 and 14 to cancel out interference in Case 2 as depicted in Figure A.3.

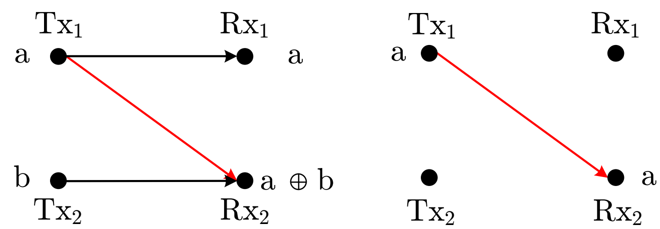


Figure A.3: Pairing opportunity Type B: By pairing Cases 2 and 14, we can communicate two bits interference-free. For instance, receiver two has access to bits $a \oplus b$ and a and as a result, it can decode its desired bit.

- Type C [Cases 3 and 13]: Similar to Type B with swapping user IDs.

We are now ready to provide the achievability scheme for the Instantaneous-CSIT model and for $0 \leq p \leq 0.5$. We first provide an overview of our scheme.

Overview

Our achievability strategy is carried on over $b+1$ communication blocks, each block with n time instants. We describe the achievability strategy for rate tuple

$$(R_1, R_2) = (p, p) . \tag{A.3}$$

Transmitters communicate fresh data bits in the first b blocks and the final block is to help the receivers decode their corresponding bits. At the end, using our scheme, we achieve rate tuple $\frac{b}{b+1}(p, p)$ as $n \to \infty$. Finally, letting $b \to \infty$, we achieve the desired rate tuple. In our scheme the messages transmitted in block j, $j = 1, 2, \ldots, b$, will be decoded at the end of block $j+1$.

Achievability strategy

Let W_i^j be the message of transmitter i in block j, $i = 1, 2$, $j = 1, 2, \ldots, b$. Moreover, let $\mathsf{W}_1^j = a_1^j, a_2^j, \ldots, a_m^j$, and $\mathsf{W}_2^j = b_1^j, b_2^j, \ldots, b_m^j$, where a_i^j's and b_i^j's are picked uniformly and independently from $\{0, 1\}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, b$, and some positive integer m. We set

$$n = m/p + \left(2/p^4\right) m^{2/3}. \tag{A.4}$$

Achievability strategy for block 1: In the first communication block, at each time instant t, Tx_i sends a new data bit (from its initial m bits) if $G_{ii}[t] = 1$, $i = 1, 2$. In other words, Tx_1 sends a new data bit either of the following channel realizations occurs (see Table 2.2): Cases $1, 2, 3, 4, 5, 6, 7$, and 8; while Tx_2 sends a new data bit if either of the following channel realizations occurs: Cases $1, 2, 3, 4, 9, 10, 11$, and 12.

If not specified, the transmitters remain silent. Tx_1 transfers its transmitted bits in Cases 1 and 2 to queues $Q_{1,\mathsf{C}1}^1$ and $Q_{1,\mathsf{C}2}^1$ respectively; and Tx_2 transfers its transmitted bits in Cases 1 and 3 to queues $Q_{2,\mathsf{C}1}^1$ and $Q_{2,\mathsf{C}3}^1$ respectively.

If at the end of block 1, there exists a bit at either of the transmitters that has not yet been transmitted, we consider it as error type-I and halt the transmission.

Remark A.2 *Note that the transmitted bits in Cases* $4, 5, 6, 7, 8, 9, 10, 11$, *and* 12 *are available at their corresponding receivers without any interference. In other words, they are communicated successfully and no retransmission is required.*

Assuming that the transmission is not halted, let random variable N_{i,C_ℓ}^1 denote the number of bits in Q_{i,C_ℓ}^1, $(i, \ell) = (1, 1), (1, 2), (2, 1), (2, 3)$. Since transition of a bit to this queue is distributed as independent Bernoulli RV, upon completion of block 1, we have

$$\mathbb{E}[N^1_{i,\mathsf{C}_\ell}] = \frac{\Pr\left(\text{Case } \ell\right)}{1 - \sum_{i=9,10,\ldots,16} \Pr\left(\text{Case i}\right)} m = \frac{1}{p} \Pr\left(\text{Case } \ell\right) m. \tag{A.5}$$

If the event $\left[N^1_{i,\mathsf{C}_\ell} \geq \mathbb{E}[N^1_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to Q^1_{i,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^1_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Furthermore, using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Achievability strategy for block j, $j = 2, 3, \ldots, b$: In communication block j, $j = 2, 3, \ldots, b$, at each time instant t, Tx_i sends a new data bit (from its initial m bits) if $G_{ii}[t] = 1$, $i = 1, 2$. Transmitter one transfers its transmitted bit in Cases 1 and 2 to queues $Q^j_{1,\mathsf{C}1}$ and $Q^j_{1,\mathsf{C}2}$ respectively; and Tx_2 transfers its transmitted bit in Cases 1 and 3 to queues $Q^j_{2,\mathsf{C}1}$ and $Q^j_{2,\mathsf{C}3}$ respectively. Note that so far the transmission scheme is similar to the first communication block.

Now if at a given time instant Case 15 occurs, Tx_i sends a bit from queue $Q^{j-1}_{i,\mathsf{C}1}$ and removes it from the this queue. If at time instant t Case 15 occurs and $Q^{j-1}_{i,\mathsf{C}1}$ is empty, then Tx_i remains silent. This way, similar to pairing Type A described previously, the transmitted bits in Case 1 of the previous block can be decoded at the corresponding receiver.

Furthermore, if at a given time instant Case 14 (13) occurs, Tx_1 (Tx_2) sends a bit from queue $Q^{j-1}_{1,\mathsf{C}2}$ ($Q^{j-1}_{2,\mathsf{C}3}$) and removes it from the this queue. This is motivated by pairing Type B (C) described previously.

If at the end of block j, there exists a bit at either of the transmitters that has not yet been transmitted, or any of the queues Q^{j-1}_{1,C_1}, Q^{j-1}_{1,C_2}, Q^{j-1}_{2,C_1}, or Q^{j-1}_{2,C_3} is not empty, we consider this event as error type-I and we halt the transmission.

Assuming that the transmission is not halted, let random variable N^j_{i,C_ℓ} denote the number of bits in Q^j_{i,C_ℓ}, $(i,\ell) = (1,1),(1,2),(2,1),(2,3)$. Since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of block j, we have

$$\mathbb{E}[N^j_{i,\mathsf{C}_\ell}] = \frac{\Pr(\text{Case } \ell)}{1 - \sum_{i=9,10,\ldots,16} \Pr(\text{Case i})} m = \frac{1}{p} \Pr(\text{Case } \ell)\, m. \tag{A.6}$$

If the event $\left[N^j_{i,\mathsf{C}_\ell} \geq \mathbb{E}[N^j_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to Q^j_{i,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^j_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Achievability strategy for block $b+1$: In the final communication block, transmitters do not communicate any new data bit.

If at time instant t Case 15 occurs, Tx_i sends a bit from queue $Q^b_{i,\mathsf{C1}}$ and removes it from the this queue. If at time instant t Case 15 occurs and $Q^b_{i,\mathsf{C1}}$ is empty, then Tx_i remains silent. If at time instant t Case 14 (13) occurs, Tx_1 (Tx_2) sends a bit from queue $Q^b_{1,\mathsf{C2}}$ ($Q^b_{2,\mathsf{C3}}$) and removes it from the this queue.

If at the end of block $b+1$, any of the states Q^b_{1,C_1}, Q^b_{1,C_2}, Q^b_{2,C_1}, or Q^b_{2,C_3} is not empty, we consider this event as error type-I and we halt the transmission.

Note that if the transmission is not halted, any bit is either available at its intended receiver interference-free, or the interfering bit is provided to the receiver in the following block. The probability that the transmission strategy halts at the end of each block can be bounded by the summation of error probabilities of types I and II. Using Chernoff-Hoeffding bound, we can show that the probability that the transmission strategy halts at any point approaches zero as $m \to \infty$.

Now, since each block has $n = m/p + \left(2/p^4\right) m^{2/3}$ time instants and the probability that the transmission strategy halts at any point approaches zero as $m \to \infty$, we achieve a rate tuple

$$\frac{b}{b+1}(p, p), \tag{A.7}$$

as $m \to \infty$. Finally letting $b \to \infty$, we achieve the desired rate tuple.

A.1.2 Achievabiliy Strategy for $0.5 < p \leq 1$

For $p = 1$, the capacity region is the same with no, delayed, or instantaneous CSIT. So in this section, we assume $0.5 < p < 1$. By symmetry, it suffices to describe the strategy for point $A = (p, 2pq)$. In this regime, we will take advantage of another pairing opportunity as described below.

• Type D [Cases 2, 3, and 12]: This type of pairing is different from what we have described so far. In all previous types, we paired up cases that had zero capacity to cancel out interference in other cases. However, here all three cases have capacity 1. By pairing all three cases together, we can communicate 4 bits as depicted in Figure A.4.

Remark A.3 *This coding opportunity can be applicable to DoF analysis of wireless networks with linear schemes in the context of $2 \times 2 \times 2$ layered networks (Section III.A of [30]).*

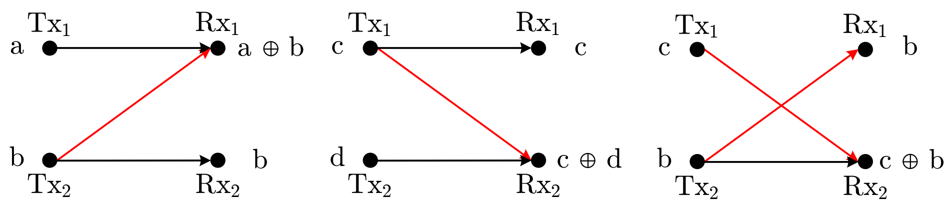


Figure A.4: Pairing opportunity Type D: Cases 2, 3, and 12. Tx_1 uses c to recover b and then it decodes a, similar argument holds for Tx_2. All three cases have capacity 1, and by pairing them, we can communicate 4 bits.

Overview

The achievability is again carried on over $b+1$ communication blocks, each block with n time instants. We describe the achievability strategy for rate tuple

$$(R_1, R_2) = (p, 2pq), \tag{A.8}$$

see Figure A.1(b).

Transmitters communicate fresh data bits in the first b blocks and the final block is to help receivers decode their corresponding bits. At the end, using our scheme, we achieve rate tuple $\frac{b}{b+1}(p, 2pq)$ as $n \to \infty$. Finally, letting $b \to \infty$, we achieve the desired corner point. In our scheme, the transmitted bits in block j, $j = 1, 2, \ldots, b$, will be decoded by the end of block $j+1$.

Achievability strategy

Let W_i^j be the message of transmitter i in block j. We assume $\mathrm{W}_1^j = a_1^j, a_2^j, \ldots, a_m^j$, and $\mathrm{W}_2^j = b_1^j, b_2^j, \ldots, b_{m_2}^j$ for $j = 1, 2, \ldots, b$, where a_i^j's and b_i^j's are picked uniformly and independently from $\{0, 1\}$, for some positive value of m and $m_2 = 2qm$

(note that $2q < 1$). We set

$$n = m/p + \left(2/q^4\right) m^{2/3}. \tag{A.9}$$

Achievability strategy for block 1: In the first communication block, at each time instant t, transmitter one sends a new data bit if $G_{11}[t] = 1$ except Case 1. In other words, Tx_1 sends a new data bit if either of the following channel realizations occurs (see Table 2.2): Cases $2, 3, 4, 5, 6, 7,$ and 8. Transmitter two sends a new data bit if $G_{22}[t] = 1$ except Cases 1 and 12. In other words, Tx_2 sends a new data bit if either of the following channel realizations occurs: Cases $2, 3, 4, 9, 10,$ and 11.

If at time instant t where $t \leq \frac{q^2}{p^2}n$, Case 1 occurs, then each transmitter sends out a new data bit. Then, Tx_i transfers its transmitted bit in Case 1 to queue $Q^1_{i,\mathsf{C1}}$ for $t \leq \frac{q^2}{p^2}n$. If $t > \frac{q^2}{p^2}n$ and Case 1 occurs, then Tx_1 sends out a new data bit while Tx_2 remains silent, see Figure A.5. Note that these bits are delivered to Rx_1 interference-free.

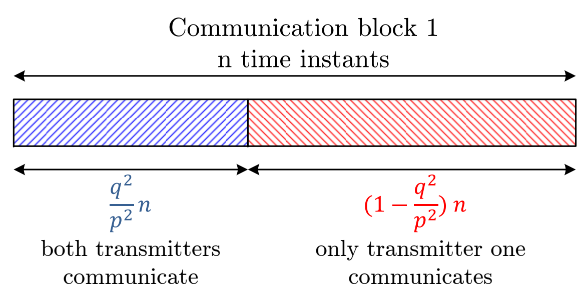


Figure A.5: If Case 1 occurs during communication block 1, then if $t \leq \frac{q^2}{p^2}n$, each transmitter sends out a new data bit. However, if $t > \frac{q^2}{p^2}n$, then Tx_1 sends out a new data bit while Tx_2 remains silent.

If $t \leq \frac{q^2}{p^2}n$, and Case 12 occurs, then Tx_2 sends out a new data bit while Tx_1 remains silent. Note that these bits are delivered to Rx_2 interference-free.

If not specified, the transmitters remain silent. Note that Tx_1 sends a bit if $G_{11}[t] = 1$ (*i.e.* with probability p). On the other hand, Tx_2 sends a bit with probability

$$\sum_{j=2,3,4,9,10,11} \Pr\left(\text{Case j}\right) + \frac{q^2}{p^2} \sum_{j=1,12} \Pr\left(\text{Case j}\right) = 2pq. \tag{A.10}$$

Transmitter one transfers its transmitted bit in Case 2 to queue $Q^1_{1,\mathsf{C2}}$; and Tx_2 transfers its transmitted bit in Case 3 to queue $Q^1_{2,\mathsf{C3}}$. If at the end of block 1, there exists a bit at either of the transmitters that has not yet been transmitted, we consider it as error type-I and halt the transmission.

Remark A.4 *Note that the transmitted bits in Cases* 4, 5, 6, 7, 8, 9, 10, *and* 11 *are available at their corresponding receivers without any interference.*

Assuming that the transmission is not halted, let random variable N^1_{i,C_ℓ} denote the number of bits in Q^1_{i,C_ℓ}, $(i,\ell) = (1,1), (1,2), (2,1), (2,3)$. Since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of block 1, we have

$$\begin{aligned}
\mathbb{E}[N^1_{1,\mathsf{C}_1}] &= \frac{\left(q^2/p^2\right)\Pr\left(\text{Case 1}\right)}{1-\sum_{j=9,10,\ldots,16}\Pr\left(\text{Case j}\right)} m = pq^2 m,\\
\mathbb{E}[N^1_{1,\mathsf{C}_2}] &= \frac{\Pr\left(\text{Case 2}\right)}{1-\sum_{j=9,10,\ldots,16}\Pr\left(\text{Case j}\right)} m = p^2 q m,\\
\mathbb{E}[N^1_{2,\mathsf{C}_1}] &= \frac{\left(q^2/p^2\right)\Pr\left(\text{Case 1}\right)}{\sum_{j=2,3,4,9,10,11}\Pr\left(\text{Case j}\right)+\frac{q^2}{p^2}\sum_{j=1,12}\Pr\left(\text{Case j}\right)} 2qm = pq^2 m,\\
\mathbb{E}[N^1_{2,\mathsf{C}_3}] &= \frac{\Pr\left(\text{Case 3}\right)}{\sum_{j=2,3,4,9,10,11}\Pr\left(\text{Case j}\right)+\frac{q^2}{p^2}\sum_{j=1,12}\Pr\left(\text{Case j}\right)} 2qm = p^2 q m.
\end{aligned} \tag{A.11}$$

If the event $\left[N^1_{i,\mathsf{C}_\ell} \geq \mathbb{E}[N^1_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to

Q^1_{i,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^1_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Achievability strategy for block j, $j = 2, 3, \ldots, b$: In communication block j, $j = 2, 3, \ldots, b$, at each time instant t, transmitter one sends a new data bit if $G_{11}[t] = 1$ except Case 1, while transmitter two sends a new data bit if $G_{22}[t] = 1$ except Cases 1 and 12.

If $t \leq \frac{q^2}{p^2} n$ and Case 1 occurs, then each transmitter sends out a new data bit. Then Tx_i transfers its transmitted bit in Case 1 to queue $Q^j_{i,\mathsf{C}1}$ for $t \leq \frac{q^2}{p^2} n$. If $t > \frac{q^2}{p^2} n$ and Case 1 occurs, then Tx_1 sends out a new data bit while Tx_2 remains silent. Note that these bits are delivered to Rx_1 interference-free.

If $t \leq \frac{q^2}{p^2} n$ and Case 12 occurs, then Tx_2 sends out a new data bit while Tx_1 remains silent. We will exploit channel realization 12 for $t > \frac{q^2}{p^2} n$, to perform pairing Type D.

Transmitter one transfers its transmitted bit in Case 2 to queue $Q^j_{1,\mathsf{C}2}$; and transmitter two transfers its transmitted bit in Case 3 to queue $Q^j_{2,\mathsf{C}3}$. Note that so far the transmission scheme is similar to the first communication block.

Now, if at time instant t Case 15 occurs, Tx_i sends a bit from queue $Q^{j-1}_{i,\mathsf{C}1}$ and removes it from the this queue. If at time instant t Case 15 occurs and $Q^{j-1}_{i,\mathsf{C}1}$ is empty, then Tx_i remains silent. This way, similar to pairing Type A described previously, the transmitted bits in Case 1 of the previous block can be decoded at the corresponding receiver.

Furthermore, if at time instant t Case 14 (13) occurs, Tx_1 (Tx_2) sends a bit from queue $Q^{j-1}_{1,\mathsf{C}2}$ ($Q^{j-1}_{2,\mathsf{C}3}$) and removes it from the this queue. This is motivated by pairing Type B (C) described previously.

Finally, if $t > \frac{q^2}{p^2}n$ and Case 12 occurs, Tx_1 sends a bit from queue $Q^{j-1}_{1,\mathsf{C}2}$ and Tx_2 sends a bit from queue $Q^{j-1}_{2,\mathsf{C}3}$. Each transmitter removes the transmitted bit from the corresponding queue. This is motivated by pairing Type D described above.

If at the end of block j, there exists a bit at either of the transmitters that has not yet been transmitted, or any of the states $Q^{j-1}_{1,\mathsf{C}1}$, $Q^{j-1}_{1,\mathsf{C}2}$, $Q^{j-1}_{2,\mathsf{C}1}$, or $Q^{j-1}_{2,\mathsf{C}3}$ is not empty, we consider it as error type-I and halt the transmission.

Assuming that the transmission is not halted, let random variable N^j_{i,C_ℓ} denote the number of bits in Q^j_{i,C_ℓ}, $(i,\ell) = (1,1),(1,2),(2,1),(2,3)$. Since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of block j, we have

$$\mathbb{E}[N^j_{1,\mathsf{C}_1}] = \frac{\left(q^2/p^2\right)\Pr\left(\text{Case 1}\right)}{1-\sum_{j=9,10,\ldots,16}\Pr\left(\text{Case j}\right)}m = pq^2m,$$
$$\mathbb{E}[N^j_{1,\mathsf{C}_2}] = \frac{\Pr\left(\text{Case 2}\right)}{1-\sum_{j=9,10,\ldots,16}\Pr\left(\text{Case j}\right)}m = p^2qm,$$
$$\mathbb{E}[N^j_{2,\mathsf{C}_1}] = \frac{\left(q^2/p^2\right)\Pr\left(\text{Case 1}\right)}{\sum_{j=2,3,4,9,10,11}\Pr\left(\text{Case j}\right)+\frac{q^2}{p^2}\sum_{j=1,12}\Pr\left(\text{Case j}\right)}2qm = pq^2m,$$
$$\mathbb{E}[N^j_{2,\mathsf{C}_3}] = \frac{\Pr\left(\text{Case 3}\right)}{\sum_{j=2,3,4,9,10,11}\Pr\left(\text{Case j}\right)+\frac{q^2}{p^2}\sum_{j=1,12}\Pr\left(\text{Case j}\right)}2qm = p^2qm. \tag{A.12}$$

If the event $\left[N^j_{i,\mathsf{C}_\ell} \geq \mathbb{E}[N^j_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}\right]$ occurs, we consider it as error type-II and we halt the transmission. At the end of block 1, we add 0's (if necessary) to Q^j_{i,C_ℓ} so that the total number of bits is equal to $\mathbb{E}[N^j_{i,\mathsf{C}_\ell}] + m^{\frac{2}{3}}$. Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Achievability strategy for block $b+1$: In the final communication block, transmitters do not communicate any new data bit.

If at time instant t Case 15 occurs, Tx_i sends a bit from queue $Q^b_{i,\mathsf{C}1}$ and removes it from the this queue. If at time instant t Case 15 occurs and $Q^b_{i,\mathsf{C}1}$ is empty, then Tx_i remains silent. If at time instant t Case 14 (13) or 12 occurs, Tx_1 (Tx_2) sends a bit from queue $Q^b_{1,\mathsf{C}2}$ ($Q^b_{2,\mathsf{C}3}$) and removes it from the this queue.

If at the end of block j any of the states $Q^b_{1,\mathsf{C}1}$, $Q^b_{1,\mathsf{C}2}$, $Q^b_{2,\mathsf{C}1}$, or $Q^b_{2,\mathsf{C}3}$ is not empty, we consider it as error type-I and halt the transmission.

Note that if the transmission is not halted, any bit is either available at its intended receiver interference-free, or the interfering bits is provided to the receiver in the following block. The probability that the transmission strategy halts at the end of each block can be bounded by the summation of error probabilities of types I and II. Using Chernoff-Hoeffding bound, we can show that the probability that the transmission strategy halts at any point approaches zero for $m \rightarrow \infty$.

Now, since each block has $n = m/p + \left(2/q^4\right) m^{2/3}$ time instants and the probability that the transmission strategy halts at any point approaches zero for $m \rightarrow \infty$, we achieve a rate tuple

$$\frac{b}{b+1}\left(p, 2pq\right), \tag{A.13}$$

as $m \rightarrow \infty$. Finally letting $b \rightarrow \infty$, we achieve the desired rate tuple.

A.2 Converse Proof of Theorem 2.1 [Instantaneous-CSIT]

The derivation of the outer-bound on individual rates is simple, however for the completeness of the results, we include the proof here. This outer-bound can be used for other theorems as needed. To derive the outer-bound on R_1, we have

$$\begin{aligned}
nR_1 &= H(W_1) \overset{(a)}{=} H(W_1|G^n) \\
&\overset{(b)}{=} H(W_1|X_2^n, G^n) \\
&\overset{\text{(Fano)}}{\leq} I(W_1; Y_1^n|X_2^n, G^n) + n\epsilon_n \\
&\overset{\text{(data proc.)}}{\leq} I(X_1^n; Y_1^n|X_2^n, G^n) + n\epsilon_n \\
&= H(Y_1^n|X_2^n, G^n) - H(Y_1^n|X_1^n, X_2^n, G^n) + n\epsilon_n \\
&= H(G_{11}^n X_1^n|X_2^n, G^n) + n\epsilon_n \\
&\leq pn + n\epsilon_n,
\end{aligned} \tag{A.14}$$

where $\epsilon_n \to 0$ as $n \to \infty$; (a) holds since message W_1 is independent of G^n; and (b) holds since given G^n, W_1 is independent of X_2^n, see (2.44). Similarly, we have

$$nR_2 \leq pn + n\epsilon_n. \tag{A.15}$$

dividing both sides by n and let $n \to \infty$, we have

$$\begin{cases} R_1 \leq p \\ R_2 \leq p \end{cases} \tag{A.16}$$

The outer-bound on $R_1 + R_2$ follows from the proof of Theorem 2.4 in Section 2.10.

A.3 Achievability Proof of Theorem 2.2: Sum-rate for $0 \leq p < 0.5$

In this appendix, we provide the achievability proof of Theorem 2.2 with Delayed-CSIT and for $0 \leq p < 0.5$. We provide the an achievability strategy for rate tuple

$$R_1 = R_2 = \min\left\{p, \frac{(1-q^2)}{1+(1-q^2)^{-1}p}\right\}. \tag{A.17}$$

Let the messages of transmitters one and two be denoted by $\mathrm{W}_1 = a_1, a_2, \ldots, a_m$, and $\mathrm{W}_2 = b_1, b_2, \ldots, b_m$, respectively, where a_i's and b_i's are picked uniformly and independently from $\{0,1\}$, $i = 1, \ldots, m$. We show that it is possible to communicate these bits in

$$n = \max\left\{m/p, \left(1-q^2\right)^{-1} m + \left(1-q^2\right)^{-2} pm\right\} + O\left(m^{2/3}\right) \tag{A.18}$$

time instants with vanishing error probability (as $m \rightarrow \infty$). Therefore achieving the rates given in (A.17) as $m \rightarrow \infty$.

Phase 1 [uncategorized transmission]: At the beginning of the communication block, we assume that the bits at Tx_i are in queue $Q_{i\rightarrow i}$, $i = 1, 2$. At each time instant, Tx_i sends out a bit from $Q_{i\rightarrow i}$ and this bit will either stay in the initial queue or a transition to a new queue will take place. Table A.1 summarizes the transitions for each channel realization. The arguments are very similar to our discussion in Section 2.5, and the only difference is the way we handle Cases 7, 8, 11, and 12. We provide some details about these cases.

For Cases 7 () and 8 (), in Section 2.5, we updated the status of the transmitted bit of Tx_2 to $Q_{2\rightarrow\{1,2\}}$. However, this scheme is suboptimal for $0 \leq p < 0.5$, and instead we update the status of the transmitted bit of Tx_2 to an intermediate queue $Q_{2,INT}$. Then in Phase 2, we retransmit these bits and upgrade their status once more. Similar story holds for Cases 11 and 12. The main reason for doing this is as follows. As we discussed in Section 2.4, there are many opportunities to combine bits in order to improve the achievable rates. However, we could never combine the bits that were transmitted in Cases 7, 8, 11, or 12 with other bits. This was not an issue for $0.5 \leq p \leq 1$, however for $0 \leq p < 0.5$, we need to find a way to combine these bits with other bits in future time instants. To do so, the only way is to keep them in an intermediate queue and retransmit them again in Phase 2.

Phase 1 goes on for

$$\left(1 - q^2\right)^{-1} m + m^{\frac{2}{3}} \tag{A.19}$$

time instants and if at the end of this phase either of the queues $Q_{i \rightarrow i}$ is not empty, we declare error type-I and halt the transmission.

Assuming that the transmission is not halted, upon completion of Phase 1, we have

$$\begin{aligned}
\mathbb{E}[N_{1,\mathsf{C}_1}] &= \frac{\Pr\left(\text{Case 1}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p^4 m, \\
\mathbb{E}[N_{1 \rightarrow 2|1}] &= \frac{\Pr\left(\text{Case 2}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p^3 q m, \\
\mathbb{E}[N_{1 \rightarrow 1|2}] &= \frac{\sum_{j=14,15} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p q^2 m, \\
\mathbb{E}[N_{1,INT}] &= \frac{\sum_{j=11,12} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p^2 q m,
\end{aligned} \tag{A.20}$$

similarly, we have

$$\mathbb{E}[N_{2,\mathsf{C}_1}] = \frac{\Pr\left(\text{Case 1}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} p^4 m,$$

$$\mathbb{E}[N_{2\rightarrow 1|2}] = \frac{\Pr\left(\text{Case 3}\right)}{1-\sum_{i=9,10,13,16}\Pr\left(\text{Case i}\right)}m = (1-q^2)^{-1}p^3qm,$$
$$\mathbb{E}[N_{2\rightarrow 2|1}] = \frac{\sum_{j=13,15}\Pr\left(\text{Case j}\right)}{1-\sum_{i=9,10,13,16}\Pr\left(\text{Case i}\right)}m = (1-q^2)^{-1}pq^2m,$$
$$\mathbb{E}[N_{2,INT}] = \frac{\sum_{j=7,8}\Pr\left(\text{Case j}\right)}{1-\sum_{i=9,10,13,16}\Pr\left(\text{Case i}\right)}m = (1-q^2)^{-1}p^2qm, \quad \text{(A.21)}$$

If the event $\left[N \geq \mathbb{E}[N] + m^{\frac{2}{3}}\right]$ for $N = N_{i,\mathsf{C}_1}, N_{i\rightarrow i|\bar{i}}, N_{i\rightarrow \bar{i}|i}, N_{i,INT}$, $i = 1,2$, occurs, we consider it as error type-II and we halt the transmission strategy. At the end of Phase 1, we add 0's (if necessary) in order to make queues Q_{i,C_1}, $Q_{i\rightarrow j|\bar{j}}$, and $Q_{i,INT}$ of size equal to $\mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}}$, $\mathbb{E}[N_{i\rightarrow j|\bar{j}}] + m^{\frac{2}{3}}$, and $\mathbb{E}[N_{i,INT}] + m^{\frac{2}{3}}$ respectively, $i = 1,2$, and $j = i, \bar{i}$. Furthermore, using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Phase 2 [upgrading status of interfering bits in Q_{1,C_j}]: In this phase, we focus on the bits in $Q_{1,INT}$ and $Q_{2,INT}$. At each time instant, Tx_i picks a bit from $Q_{i,INT}$ and sends it. This bit will either stay in $Q_{i,INT}$ or a transition to a new queue will take place. Table A.2 describes what happens to the status of the bits if either of the 16 possible cases occurs. Due to symmetry, we only describe the transitions for bits in $Q_{1,INT}$. Consider bit "a" in $Q_{1,INT}$.

- Cases $1,2,3,4$, and 5: The transitions for these cases are consistent with our previous discussions.
- Cases $9, 10, 11, 12, 13$, and 16: In these cases, it is easy to see that no change occurs in the status of bit a.
- Case 6: In this case, bit a is delivered to both receivers and hence, no further transmission is required. Therefore, it joins $Q_{1,F}$.

- Case 7: Here with slight abuse of, Q_{1,C_1} represents the bits of Tx_1 that are received at both receivers with interference but not necessarily in Case 1, $i = 1, 2$. For instance if at a given time, Tx_1 sends a bit from $Q_{1,INT}$ and Case 7 occurs, then this bit joins Q_{1,C_1} since now both receivers have received this bit with interference.
- Case 8: In this case, bit a is available at Rx_2 but it is interfered at Rx_1 by bit b. However, in Case 8 no change occurs for the bits in $Q_{2,INT}$. Therefore, since bit b will be retransmitted until it is provided to Rx_1, no retransmission is required for bit a and it joins $Q_{1,F}$.
- Cases 14 and 15: If either of these cases occur, bit a becomes available at Rx_2 and is needed at Rx_1. Thus, we update the status of such bits to $Q_{1\rightarrow 1|2}$.

Phase 2 goes on for

$$\begin{aligned}&\left(1-\sum_{i=9,10,11,12,13,16} \Pr\left(\text{Case } i\right)\right)^{-1}\left(1-q^2\right)^{-1}p^2qm + 2m^{\frac{2}{3}}\\ &= \left(1-\left[p^2q+q^2\right]\right)^{-1}\left(1-q^2\right)^{-1}p^2qm + 2m^{\frac{2}{3}} \qquad \text{(A.22)}\end{aligned}$$

time instants and if at the end of this phase either of the states $Q_{i,INT}$ is not empty, we declare error type-I and halt the transmission.

Assuming that the transmission is not halted, upon completion of Phase 2, the states $Q_{1,INT}$ and $Q_{2,INT}$ are empty and we have

$$\begin{aligned}\mathbb{E}[N_{i,\mathsf{C}_1}] &= (1-q^2)^{-1}p^4m + \frac{\sum_{j=1,7}\Pr\left(\text{Case j}\right)}{1-\sum_{i=9,10,11,12,13,16}\Pr\left(\text{Case i}\right)}(p^2qm+m^{2/3})\\ &= (1-q^2)^{-1}p^4m + \left(1-\left[p^2q+q^2\right]\right)^{-1}\left(p^4+p^2q^2\right)\left(p^2qm+m^{2/3}\right), \qquad \text{(A.23)}\end{aligned}$$

similarly, we have

$$\mathbb{E}[N_{i\rightarrow\bar{i}|i}] = (1-q^2)^{-1}p^3qm + \frac{\sum_{j=2,4,5}\Pr\left(\text{Case j}\right)}{1-\sum_{i=9,10,11,12,13,16}\Pr\left(\text{Case i}\right)}(p^2qm+m^{2/3})$$

$$= (1-q^2)^{-1}p^3qm + \left(1-\left[p^2q+q^2\right]\right)^{-1}\left(p^2q+p^2q^2\right)\left(p^2qm+m^{2/3}\right), \quad \text{(A.24)}$$

and

$$\mathbb{E}[N_{i\rightarrow i|\bar{i}}] = (1-q^2)^{-1}pq^2m + \frac{\sum_{j=14,15}\Pr\left(\text{Case j}\right)}{1-\sum_{i=9,10,11,12,13,16}\Pr\left(\text{Case i}\right)}(p^2qm+m^{2/3})$$
$$= (1-q^2)^{-1}pq^2m + \left(1-\left[p^2q+q^2\right]\right)^{-1}pq^2\left(p^2qm+m^{2/3}\right). \quad \text{(A.25)}$$

If the event $\left[N \geq \mathbb{E}[N] + m^{\frac{2}{3}}\right]$ for $N = N_{i,\mathsf{C}_1}, N_{i\rightarrow i|\bar{i}}, N_{i\rightarrow \bar{i}|i}$, $i = 1,2$, occurs, we consider it as error type-II and we halt the transmission strategy. At the end of Phase 2, we add 0's (if necessary) in order to make queues Q_{i,C_1} and $Q_{i\rightarrow j|\bar{j}}$ of size equal to $\mathbb{E}[N_{i,\mathsf{C}_1}] + m^{\frac{2}{3}}$, and $\mathbb{E}[N_{i\rightarrow j|\bar{j}}] + m^{\frac{2}{3}}$ respectively, $i = 1,2$, and $j = i, \bar{i}$. Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Phase 3 [encoding and retransmission]: In this phase, Tx_i communicates the bits in $Q_{i\rightarrow i|\bar{i}}$ to Rx_i, $i = 1,2$. However, it is possible to create XOR of these bits with the bits in $Q_{i\rightarrow \bar{i}|i}$ and the bits in Q_{i,C_1} to create bits of common interest. To do so, we first encode the bits in these states using the results of [17], and then we create the XOR of the encoded bits.

In other words, given $\epsilon, \delta > 0$, Tx_i encodes all the bits in $Q_{i\rightarrow i|\bar{i}}$ at rate $p-\delta$ using random coding scheme of [17]. Similarly, Tx_i encodes $q\left(\mathbb{E}[N_{i\rightarrow i|\bar{i}}] + m^{\frac{2}{3}}\right)$ bits from $Q_{i\rightarrow \bar{i}|i}$ and Q_{i,C_1} at rate $pq-\delta$ (if there are less bits in $Q_{i\rightarrow \bar{i}|i}$ and Q_{i,C_1}, then encode all of the bits in these queues). More precisely, first Tx_i encodes bits from $Q_{i\rightarrow \bar{i}|i}$, and if the number of bits in $Q_{i\rightarrow \bar{i}|i}$ is less than $q\left(\mathbb{E}[N_{i\rightarrow i|\bar{i}}] + m^{\frac{2}{3}}\right)$, then Tx_i uses bits in Q_{i,C_1}. Tx_i will then communicate the XOR of these encoded bits.

Note that since Rx_i already knows the bits in $Q_{i\rightarrow \bar{i}|i}$, it can remove the corresponding part of the received signal. Then since the channel from $\mathsf{Tx}_{\bar{i}}$ to Rx_i can be viewed as a binary erasure channel with success probability of pq, from [17], we know that Rx_i can decode $Q_{\bar{i},\mathsf{C}_1}$ with decoding error probability less than or equal to ϵ. Thus, Rx_i can decode the transmitted bits from $Q_{\bar{i},\mathsf{C}_1}$ and use them to decode the bits in Q_{i,C_1}. Then, Rx_i removes the contribution of the bits in Q_{i,C_1} the received signal. Finally, since the channel from Tx_i to Rx_i can be viewed as a binary erasure channel with success probability of p, from [17], we know that Rx_i can decode $Q_{i\rightarrow i|\bar{i}}$ with decoding error probability less than or equal to ϵ.

If an error occurs in decoding of the encoded bits, we halt the transmission. Assuming that the transmission is not halted, at the end of Phase 3, $Q_{i\rightarrow i|\bar{i}}$ becomes empty and there are

$$\left(\mathbb{E}[N_{i\rightarrow \bar{i}|i}] + \mathbb{E}[N_{i,\mathsf{C}_1}] + 2m^{\frac{2}{3}} - q\left(\mathbb{E}[N_{i\rightarrow i|\bar{i}}] + m^{\frac{2}{3}}\right)\right)^{+} \tag{A.26}$$

bits left in $Q_{i\rightarrow \bar{i}|i}$ and Q_{i,C_1}.

If $Q_{i\rightarrow \bar{i}|i}$ and Q_{i,C_1} are also empty, the transmission strategy ends here. Otherwise, we merge the remaining bits in $Q_{i\rightarrow \bar{i}|i}$ (if any) with the bits in Q_{i,C_1} as **Type-III** (see Section 2.4) and put the XOR of them in $Q_{i\rightarrow\{1,2\}}$, $i = 1, 2$. Finally, we need to describe what happens to the remaining bits in Q_{i,C_1}. As mentioned before, a bit in Q_{i,C_1} can be viewed as a bit of common interest by itself. For the remaining bits in Q_{1,C_1}, we put the first half in $Q_{1\rightarrow\{1,2\}}$ (suppose m is picked such that the remaining number of bits is even). Note that if these bits are delivered to Rx_2, then Rx_2 can decode the first half of the bits in Q_{2,C_1}. Therefore, the first half of the bits in Q_{2,C_1} join $Q_{2,F}$.

Phase 4 [communicating bits of common interest]: During Phase 4, we deliver the bits in $Q_{1\rightarrow\{1,2\}}$ and $Q_{2\rightarrow\{1,2\}}$ using the transmission strategy for the two-source multicast problem. More precisely, the bits in $Q_{i\rightarrow\{1,2\}}$ will be considered as the message of transmitter Tx_i and they will be encoded as in the achievability scheme of Lemma 2.2, $i = 1, 2$. Fix $\epsilon, \delta > 0$, from Lemma 2.2 we know that the rate tuple

$$(R_1, R_2) = \frac{1}{2}\left((1-q^2) - \delta, (1-q^2) - \delta\right)$$

is achievable with decoding error probability less than or equal to ϵ. Thus, using Lemma 2.2, we can communicate the remaining bits at rate $(1 - q^2) - \delta$ with decoding error probability less than or equal to ϵ. If an error occurs in decoding of the encoded bits, we halt the transmission.

Using Chernoff-Hoeffding bound and the results of [17], we can show that the probability that the transmission strategy halts at any point approaches zero for $\epsilon, \delta \to 0$ and $m \to \infty$. Moreover, it is easy to verify that for $0 \leq p \leq \left(3 - \sqrt{5}\right)/2$, at the end of Phase 3, $Q_{i \to \bar{i}|i}$ and Q_{i,C_1} are empty and the transmission strategy ends there. However, for $\left(3 - \sqrt{5}\right)/2 < p < 0.5$, the transmission strategy continues to Phase 4. Therefore, we can show that if no error occurs, the transmission strategy end in

$$n = \max\left\{m/p, \left(1-q^2\right)^{-1} m + \left(1-q^2\right)^{-2} pm\right\} + O\left(m^{2/3}\right) \qquad \text{(A.27)}$$

time instants. Therefore achieving the rates given in (A.17).

A.4 Achievability Proof of Theorem 2.2: Corner Point C

In this appendix, we describe the achievability strategy for corner point C depicted in Figure 2.17(b), *i.e.*

$$(R_1, R_2) = \left(pq(1+q), p\right). \qquad \text{(A.28)}$$

Let the messages of transmitters one and two be denoted by $\mathsf{W}_1 = a_1, a_2, \ldots, a_{m_1}$, and $\mathsf{W}_2 = b_1, b_2, \ldots, b_m$, respectively, where data bits a_i's and b_i's are picked uniformly and independently from $\{0, 1\}$, and $m_1 = q(1+q)m$ (suppose the parameters are such that $m, m_1 \in \mathbb{Z}$). Note that for $\left(3 - \sqrt{5}\right)/2 < p \leq 1$, we have $q(1+q) < 1$. We show that it is possible to communicate these bits in

$$n = \frac{1}{p}m + O\left(m^{2/3}\right) \tag{A.29}$$

time instants with vanishing error probability (as $m \to \infty$). Therefore, achieving corner point C as $m \to \infty$. Our transmission strategy consists of five phases as described before.

Phase 1 [uncategorized transmission]: This phase is similar to Phase 1 of the achievability strategy for the optimal sum-rate point A. The main difference is due to the fact that the transmitters start with unequal number of bits. At the beginning of the communication block, we assume that the bits $a_1, a_2, \ldots, a_{m_1}$ at Tx_1 and the bits $b_1, b_2, \ldots, b_{m_1}$ at Tx_2 are in queues $Q_{1\to1}$ and $Q_{2\to2}$ respectively.

Remark A.5 *Note that Tx_2 has m initial bits, however, only m_1 of them are in $Q_{2\to2}$ at the beginning of the communication block.*

At each time instant t, Tx_i sends out a bit from $Q_{i\to i}$, and this bit will either stay in the initial queues or a transition will take place. Based on the channel realizations, a total of 16 possible configurations may occur at any time instant. Table A.3 summarizes the transition from the initial queue for each channel realization.

In comparison to the achievability strategy of the sum-rate point A, we have new queues for the bits:

(a) $Q_{i,OP}$ denotes the bits that have caused interference at the unintended receiver and this interference has to get resolved.

(b) $Q_{1,INT}$ denotes an intermediate queue of the bits at Tx_1 that were transmitted when channel realizations 11 or 12 occurred.

(c) $Q_{i,INT}$ denotes an intermediate queue of the bits at Tx_2 that were transmitted when channel realizations 7 or 8 occurred.

Phase 1 goes on for

$$(1/p - 1)\, m + m^{\frac{2}{3}} \tag{A.30}$$

time instants and if at the end of this phase, either of queues $Q_{i \to i}$ is not empty, we declare error type-I and halt the transmission.

Assuming that the transmission is not halted, let random variable N_{1,C_1}, $N_{i,OP}$, $N_{i \to i|\bar{i}}$, and $N_{i,INT}$ denote the number of bits in Q_{1,C_1}, $Q_{i,OP}$, $Q_{i \to i|\bar{i}}$, and $Q_{i,INT}$ respectively $i = 1, 2$. The transmission strategy will be halted and an error (that we refer to as error type-II) will occur, if any of the following events happens.

$$\begin{aligned}
&N_{1,\mathsf{C}_1} > \mathbb{E}[N_{1,\mathsf{C}_1}] + m^{\frac{2}{3}} \stackrel{\triangle}{=} n_{1,\mathsf{C}_1};\\
&N_{i,OP} > \mathbb{E}[N_{i,OP}] + m^{\frac{2}{3}} \stackrel{\triangle}{=} n_{i,OP}, \quad i = 1, 2;\\
&N_{i \to i|\bar{i}} > \mathbb{E}[N_{i \to i|\bar{i}}] + m^{\frac{2}{3}} \stackrel{\triangle}{=} n_{i \to i|\bar{i}}, \quad i = 1, 2;\\
&N_{i,INT} > \mathbb{E}[N_{i,INT}] + m^{\frac{2}{3}} \stackrel{\triangle}{=} n_{i,INT}, \quad i = 1, 2.
\end{aligned} \tag{A.31}$$

From basic probability, and we have

$$\begin{aligned}
\mathbb{E}[N_{1,\mathsf{C}_1}] &= \frac{\Pr\left(\text{Case 1}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m_1 = (1 - q^2)^{-1} p^4 m_1 = p^3 q m,\\
\mathbb{E}[N_{1,OP}] &= \frac{\Pr\left(\text{Case 2}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m_1 = (1 - q^2)^{-1} p^3 q m_1 = p^2 q^2 m,
\end{aligned}$$

$$\mathbb{E}[N_{2,OP}] = \frac{\Pr(\text{Case 3})}{1-\sum_{i=5,6,14,16}\Pr(\text{Case i})} m_1 = (1-q^2)^{-1} p^3 q m_1 = p^2q^2 m,$$
$$\mathbb{E}[N_{i\to i|\bar{i}}] = \frac{\sum_{j=14,15}\Pr(\text{Case j})}{1-\sum_{i=9,10,13,16}\Pr(\text{Case i})} m_1 = (1-q^2)^{-1}\left(pq^3+p^2q^2\right) m_1 = q^3 m,$$
$$\mathbb{E}[N_{i,INT}] = \frac{\sum_{j=11,12}\Pr(\text{Case j})}{1-\sum_{i=9,10,13,16}\Pr(\text{Case i})} m_1 = (1-q^2)^{-1}\left(p^3q+p^2q^2\right) m_1 = pq^2 m. \tag{A.32}$$

Furthermore, using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

At the end of Phase 1, we add 0's (if necessary) in order to make queues Q_{1,C_1}, $Q_{i,OP}$, $Q_{i\to i|\bar{i}}$, and $Q_{i,INT}$ of size equal to n_{1,C_1}, $n_{i,OP}$, $n_{i\to i|\bar{i}}$, and $n_{i,INT}$ respectively as defined in (A.31), $i = 1,2$. For the rest of this appendix, we assume that Phase 1 is completed and no error has occurred.

Phase 2 [updating status of the bits in $Q_{i,INT}$]: In this phase, we focus on the bits in $Q_{i,INT}$, $i = 1,2$. The ultimate goal is to deliver the bits in $Q_{i,INT}$ to both receivers. At each time instant, Tx_i picks a bit from $Q_{i,INT}$ and sends it. This bit will either stay in $Q_{i,INT}$ or a transition to a new queue will take place. Table A.4 describes what happens to the status of the bits if either of the 16 cases occurs.

Here, we describe what happens to the status of a bit in $Q_{1,INT}$ if either of the 16 channel realizations occur. The description for a bit in $Q_{2,INT}$ is very similar and is summarized in Table A.4. Consider a bit "a" in $Q_{1,INT}$. At each time instant, 16 possible cases may occur:

- Cases 9,10,11,12,13, and 16: In these cases, it is easy to see that no change occurs in the status of bit a.
- Case 6: In this case, bit a is delivered to both receivers and hence, no fur-

ther transmission is required. Therefore, it joins $Q_{1,F}$.

- Case 8: In this case, bit a is available at Rx_2 but it is interfered at Rx_1 by bit b. However, in Case 8 no change occurs for the bits in $Q_{2,INT}$. Therefore, since bit b will be retransmitted until it is provided to Rx_1, no retransmission is required for bit a and it joins $Q_{1,F}$.
- Cases 14 and 15: If either of these cases occur, bit a becomes available at Rx_2 and is needed at Rx_1. Thus, we update the status of such bits to $Q_{1\rightarrow 1|2}$.
- Cases 1,2,3,4,5, and 7: If either of these cases occur, we upgrade the status of bit a to the opportunistic state $Q_{1,OP}$, meaning that from now on bit a has to be provided to either Rx_2 or both receivers such that it causes no further interference. For instance, if Case 2 occurs, providing bit a to both receivers is suffiecient to decode the simultaneously transmitted bits.

Phase 2 goes on for

$$\left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq^2 m + 2m^{\frac{2}{3}} \tag{A.33}$$

time instants, and if at the end of this phase either of the states $Q_{i,INT}$ is not empty, we declare error type-I and halt the transmission.

Assuming that the transmission is not halted, since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of Phase 2, we have

$$\begin{aligned}
\mathbb{E}[N_{1,\mathsf{C}_1}] &= p^3 qm + \frac{\sum_{j=1,3} \Pr(\text{Case j})}{1 - \sum_{i=9,10,11,12,13,16} \Pr(\text{Case i})}(pq^2 m + m^{2/3}) \\
&= p^3 qm + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} p^3 (pq^2 m + m^{2/3}), \\
\mathbb{E}[N_{1,OP}] &= p^2 q^2 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq(pq^2 m + m^{2/3}), \\
\mathbb{E}[N_{2,OP}] &= p^2 q^2 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq(1+p^2)(pq^2 m + m^{2/3}),
\end{aligned}$$

$$\mathbb{E}[N_{i\to i|\bar{i}}] = q^3 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq^2 \left(pq^2 + m^{2/3}\right), \qquad i = 1,2. \quad \text{(A.34)}$$

The transmission strategy will halt and an error (that we refer to as error type-II) will occur, if any of the following events happens.

$$\begin{aligned} &N_{1,\mathsf{C}_1} > \mathbb{E}[N_{1,\mathsf{C}_1}] + m^{\frac{2}{3}} \triangleq n_{1,\mathsf{C}_1}; \\ &N_{i,OP} > \mathbb{E}[N_{i,OP}] + m^{\frac{2}{3}} \triangleq n_{i,OP}, \quad i = 1,2; \\ &N_{i\to i|\bar{i}} > \mathbb{E}[N_{i\to i|\bar{i}}] + m^{\frac{2}{3}} \triangleq n_{i\to i|\bar{i}}, \quad i = 1,2. \end{aligned} \quad \text{(A.35)}$$

Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Again, at the end of Phase 2, we add 0's (if necessary) in order to make queues Q_{1,C_1}, $Q_{i,OP}$, and $Q_{i\to i|\bar{i}}$ of size equal to n_{1,C_1}, $n_{i,OP}$, and $n_{i\to i|\bar{i}}$ respectively as defined in (A.35), $i = 1,2$. For the rest of this appendix, we assume that Phase 2 is completed and no error has occurred.

Note that Tx_2 initially had m fresh data bits but during Phase 1 it only communicated m_1 of them. The rest of those bits will be transmitted during Phase 3 as described below.

Phase 3 [uncategorized transmission vs interference management]: During Phase 3, Tx_1 (the secondary user) communicates $\frac{q}{1+q}(p - q^2)m$ bits from states Q_{1,C_1} and $Q_{1,OP}$ at a rate such that both receivers can decode them at the end of Phase 3, regardless of the transmitted signal of Tx_2. In fact, at Rx_1, we have

$$\Pr\left[G_{11}[t] = 1, G_{21}[t] = 0\right] = pq, \quad \text{(A.36)}$$

and at Rx_2, we have

$$\Pr\left[G_{22}[t] = 0, G_{12}[t] = 1\right] = pq. \quad \text{(A.37)}$$

Hence, using the results of [17], we know that given any $\epsilon, \delta > 0$, Tx_1 can use a random code of rate $pq - \delta$ to encode $\frac{q}{1+q}(p - q^2)m$ bits from states Q_{1,C_1} and $Q_{1,OP}$, and transmits them such that both receivers can decode the transmitted message with error probability less than or equal to ϵ for sufficiently large block length (Tx_1 picks bits from Q_{1,C_1} and if this state becomes empty it starts picking from the bits in $Q_{1,OP}$). Since Rx_2 can decode the transmitted signal of Tx_1 in this phase, we can assume that the encoded bits of Tx_1, do not create any new interference during Phase 3.

We now describe what Tx_2 does during Phase 3. At the beginning of Phase 3, we assume that the bits $b_{m_1+1}, b_{m_1+2}, \ldots, b_m$ at Tx_2 are in state $Q_{2\rightarrow 2}$. At each time instant, Tx_2 picks a bit from $Q_{2\rightarrow 2}$ and sends it. This bit will either stay in $Q_{2\rightarrow 2}$ or a transition occurs as described below.

- Cases $1, 2, 3, 4, 9, 10, 11,$ and 12: In these cases the direct link from Tx_2 to Rx_2 is on. Therefore, since at the end of block (assuming large enough block length), we can decode and remove the transmitted signal of Tx_1, the transmitted bit of Tx_2 leaves $Q_{2\rightarrow 2}$ and joins $Q_{2,F}$.
- Cases $7, 8, 13,$ and 15: In these cases (assuming the transmitted signal of Tx_1 can be removed), the transmitted bit of Tx_2 becomes available at Rx_1 while it is required at Rx_2. Thus, the transmitted bit of Tx_2 leaves $Q_{2\rightarrow 2}$ and joins $Q_{2\rightarrow 2|1}$.
- Cases $5, 6, 14,$ and 16: In these cases, no change hanppens in the status of the transmitted bit from Tx_2.

Phase 3 goes on for

$$\frac{(p - q^2)}{(1 - q^2)}m + m^{\frac{2}{3}} \tag{A.38}$$

time instants and if at the end of this phase there is a bit left in $Q_{2\to 2}$ or an error occurs in decoding the transmitted signal of Tx_1, we declare error type-I and halt the transmission. Note that during Phase 3, the number of bits in $Q_{1\to 1|2}$ and $Q_{2,OP}$ remain unchanged.

Assuming that the transmission is not halted, since transition of a bit to this state is distributed as independent Bernoulli RV, upon completion of Phase 2, we have

$$\mathbb{E}[N_{1,\mathsf{C}_1}] = \left[p^3qm + \left(1 - \left[p^3q + 2pq^2 + q^4\right]\right)^{-1} p^3(pq^2m + m^{2/3}) - \frac{q}{1+q}(p-q^2)m\right]^+,$$
$$\begin{aligned}\mathbb{E}[N_{1,OP}] = {} & p^2q^2m + \left(1 - \left[p^3q + 2pq^2 + q^4\right]\right)^{-1} pq(pq^2m + m^{2/3}) \\ & - \left[\frac{q}{1+q}(p-q^2)m - p^3qm - \left(1 - \left[p^3q + 2pq^2 + q^4\right]\right)^{-1} p^3(pq^2m + m^{2/3})\right]^+.\end{aligned} \tag{A.39}$$

For $\left(3 - \sqrt{5}\right)/2 \leq p \leq 1$, $\mathbb{E}[N_{1,OP}]$ is non-negative. The transmission strategy will halt and an error (that we refer to as error type-II) will occur, if any of the following events happens.

(a) $N_{1,\mathsf{C}_1} > \mathbb{E}[N_{1,\mathsf{C}_1}] + m^{\frac{2}{3}}$;

(b) $N_{1,OP} > \mathbb{E}[N_{1,OP}] + m^{\frac{2}{3}}$;

(c) $N_{2\to 2|1} > \mathbb{E}[N_{2\to 2|1}] + m^{\frac{2}{3}}$.

Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II decreases exponentially with m.

Phase 4 [delivering interference-free bits and interference management]: In Phase 4, Tx_1 will communicate all the bits in $Q_{1\to 1|2}$. However, it is possible to create XOR of these bits with bits in $Q_{1,OP}$ in order to create bits of common interest. To do so, we first encode the bits in these states using the results of [17], and then we create the XOR of the encoded bits. On the other hand, Tx_2 will do the same to part of the bits in $Q_{2\rightarrow 2|1}$ and $Q_{2,OP}$.

More precisely, for any $\epsilon, \delta > 0$, Tx_1 encodes all the bits in $Q_{1\rightarrow 1|2}$ at rate $p - \delta$ using random coding scheme of [17]. Similarly, Tx_1 encodes

$$q^4 + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} p^2 q^5 \tag{A.40}$$

bits from $Q_{1,OP}$ at rate $pq - \delta$. Then Tx_1 will communicate the XOR of these encoded bits.

During Phase 3, Tx_2 encodes same number of the bits as in $Q_{1\rightarrow 1|2}$ from $Q_{2\rightarrow 2|1}$ at rate $p - \delta$ and

$$q^4 + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} p^2 q^5$$

bits from $Q_{2,OP}$ at rate $pq - \delta$ using random coding scheme of [17]. Then Tx_2 will communicate the XOR of these encoded bits.

Since Rx_2 already has access to the bits in $Q_{1\rightarrow 1|2}$ and $Q_{2,OP}$, it can remove their contribution from the received signals. Then for sufficiently large block length, Rx_2 can decode the transmitted bits from $Q_{1,OP}$ with decoding error probability less than or equal to ϵ. After decoding and removing this part, Rx_2 can decode the encoded bits from $Q_{2\rightarrow 2|1}$ with decoding error probability less than or equal to ϵ.

On the other hand, since Rx_1 already has access to the bits in $Q_{2\rightarrow 2|1}$ and $Q_{1,OP}$, it can remove their contribution from the received signals. Then for sufficiently large block length, Rx_1 can decode the transmitted bits from $Q_{2,OP}$ with decoding error probability less than or equal to ϵ. Finally, after decoding and removing this part, Rx_1 can decode the encoded bits from $Q_{1\to 1|2}$ with decoding error probability less than or equal to ϵ.

Phase 4 goes on for

$$\left[q^3 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq^2 \left(pq^2 + m^{2/3}\right)\right] / (p - \delta) \tag{A.41}$$

time instants. If an error occurs in decoding any of the encoded signals in Phase 4, we consider it as error and we halt the transmission strategy.

At the end of Phase 4, $Q_{1\to 1|2}$ becomes empty. Define

$$\beta = (pq - \delta)\left[q^3 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq^2 \left(pq^2 + m^{2/3}\right)\right] / (p - \delta), \tag{A.42}$$

note that as $\epsilon, \delta \to 0$, β agrees with the expression given in (A.40).

Upon completion of Phase 4, we have

$$\begin{aligned}
\mathbb{E}[N_{1,OP}] &= p^2 q^2 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq(pq^2 m + m^{2/3}) \\
&\quad - \left[\frac{q}{1+q}(p - q^2)m - p^3 qm - \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} p^3(pq^2 m + m^{2/3})\right]^+ \\
&\quad - \frac{q}{1+q}(p - q^2)m - \beta, \\
\mathbb{E}[N_{2,OP}] &= p^2 q^2 m + \left(1 - \left[p^3 q + 2pq^2 + q^4\right]\right)^{-1} pq(1 + p^2)(pq^2 m + m^{2/3}) - \beta, \\
\mathbb{E}[N_{2\to 2|1}] &= \frac{pq}{(1 - q^2)}\left(p - q^2\right)m.
\end{aligned} \tag{A.43}$$

The transmission strategy will halt and an error of type-II will occur, if any of the following events happens.

(a) $N_{i,OP} > \mathbb{E}[N_{i,OP}] + m^{\frac{2}{3}}$, $i = 1, 2$;

(b) $N_{2\to 2|1} > \mathbb{E}[N_{2\to 2|1}] + m^{\frac{2}{3}}$.

Using Chernoff-Hoeffding bound, we can show that the probability of errors of type II decreases exponentially with m. Furthermore, the probability that an error occurs in decoding any of the encoded signals in Phase 4 can be made arbitrary small as $m \rightarrow \infty$.

Phase 5 [delivering interference-free bits and interference management]: In Phase 5, the transmitters will communicate the remaining bits in $Q_{1,OP}$, Q_{1,C_1}, $Q_{2\rightarrow 2|1}$, and $Q_{2,OP}$. Tx_1 will communicate all the bits in $Q_{1,OP}$ and Q_{1,C_1} such that for sufficiently large block length, both receivers can decode them with arbitrary small error. On the other hand, Tx_2 will communicate the bits in $Q_{2\rightarrow 2|1}$ and $Q_{2,OP}$ similar to Phase 4 with one main difference. Since both receivers can completely remove the contribution of Tx_1 at the end of the block, Tx_2 can send the bits in $Q_{2,OP}$ at a higher rate of p as opposed to pq during Phase 4.

More precisely, for any $\epsilon, \delta > 0$, Tx_1 using random coding scheme of [17], encodes all the bits in $Q_{1,OP}$ and Q_{1,C_1} at rate $pq - \delta$ and communicates them. On the other hand, Tx_2 using random coding, encodes all the bits in $Q_{2\rightarrow 2|1}$ and all bits in $Q_{2,OP}$ at rate $p - \delta$. Then, Tx_2 communicates the XOR of its encoded bits.

Since Rx_1 already has access to the bits in $Q_{2\rightarrow 2|1}$ and $Q_{1,OP}$, it can remove the corresponding parts of the transmitted signals. Then for sufficiently large block length, Rx_1 can decode the transmitted bits from Q_{1,C_1} and $Q_{2,OP}$ with decoding error probability less than or equal to ϵ.

Finally, since Rx_2 already has access to the bits in $Q_{2,OP}$, it can remove the corresponding part of the transmitted signal. Then for sufficiently large block length, Rx_2 can decode the transmitted bits from $Q_{1,OP}$ and Q_{1,C_1} with decoding error probability less than or equal to ϵ. After decoding and removing this part,

Rx_2 can decode the encoded bits from $Q_{2\rightarrow 2|1}$ with decoding error probability less than or equal to ϵ.

Phase 5 goes on for

$$\left[\frac{pq}{(1-q^2)}\left(p-q^2\right)m+2m^{2/3}\right]/\left(p-\delta\right) \tag{A.44}$$

time instants. If an error occurs in decoding any of the encoded signals in Phase 5, we consider it as error and halt the transmission strategy. It is straight forward to verify that at the end of Phase 5, if the transmission is not halted, all states are empty and all bits are successfully delivered.

The probability that the transmission strategy halts at any point can be bounded by the summation of error probabilities of types I-II and the probability that an error occurs in decoding the encoded bits. Using Chernoff-Hoeffding bound and the results of [17], we can show that the probability that the transmission strategy halts at any point approaches zero for $\epsilon, \delta \rightarrow 0$ and $m \rightarrow \infty$. Moreover, the total transmission requires

$$\frac{1}{p}m+6m^{2/3} \tag{A.45}$$

time instants. Thus, Tx_1 achieves a rate of $pq(1+q)$ while Tx_2 achieves a rate of p.

This completes the achievability proof of Theorem 2.2.

A.5 Proof of Lemma 2.2

In this appendix, we provide the proof of Lemma 2.2. We first derive the outer-bound and then we describe the achievability. The outer-bound on R_i is the same as in Section A.2.

Suppose there are encoders and decoders at the transmitters and receivers respectively, such that each receiver can decode both messages with arbitrary small decoding error probability as the block length goes to infinity. We have

$$\begin{aligned}
&n(R_1 + R_2 - \epsilon_n) \\
&\overset{(a)}{\leq} I(W_1, W_2; Y_1^n | G^n) \\
&= H(Y_1^n | G^n) - H(Y_1^n | W_1, W_2, G^n) \\
&\overset{(b)}{=} H(Y_1^n | G^n) \\
&\overset{(c)}{\leq} \sum_{t=1}^{n} H(Y_1[t] | G^n) \\
&\leq \left(1 - q^2\right) n,
\end{aligned} \tag{A.46}$$

where $\epsilon_n \to 0$ as $n \to \infty$; and (a) follows from the fact that the messages and G^n are mutually independent, Fano's inequality, and the fact that Rx_1 should be able to decode both messages; (b) holds since the received signal Y_1^n is a deterministic function of W_1, W_2, and G^n; and (c) follows from the fact that conditioning reduces entropy. Dividing both sides by n and let $n \to \infty$, we get

$$R_1 + R_2 \leq 1 - q^2. \tag{A.47}$$

Below, we provide the achievability proof of Lemma 2.2. Let $\mathrm{W}_i \in \{1, 2, \ldots, 2^{nR_i}\}$ denote the message of user i.

In [17], it has been shown that for a binary erasure channel with success probability p, and for any $\epsilon, \delta > 0$, as long as the communication rate is less than or equal to $p - \delta$, we can have decoding error probability less than or equal to ϵ.

Codebook generation is as follows. Transmitter i creates 2^{nR_i} ($R_1 = p - \delta$ and $R_2 = pq - \delta$) independent codewords where each entry of the codewords is an i.i.d. Bernoulli 0.5 RV. For message index j, transmitter i will send the j^{th} codeword. Note that we can view the channel from Tx_2 to Rx_1 as a binary erasure channel with success probability pq (whenever $G_{11}[t] = 0$ and $G_{21}[t] = 1$, we get a clean observation of $X_2[t]$). Therefore, since $R_2 = pq - \delta$, Rx_1 can decode W_2 with arbitrary small decoding error probability as $n \to \infty$ and remove X_2^n from its received signal. After removing X_2^n, we can view the channel from Tx_1 to Rx_1 as a binary erasure channel with success probability p (whenever $G_{11}[t] = 1$, we get a clean observation of $X_1[t]$). Therefore, since $R_1 = p - \delta$, Rx_1 can decode W_1 with arbitrary small decoding error probability as $n \to \infty$. Similar argument holds for Rx_2. This completes the achievability proof of corner point

$$(R_1, R_2) = (p, pq)\,. \tag{A.48}$$

Similarly, we can achieve corner point

$$(R_1, R_2) = (pq, p)\,. \tag{A.49}$$

Therefore with time sharing, we can achieve the entire region as described in Lemma 2.2.

A.6 Achievability Proof of Theorem 2.3: Corner Point $\left(1 - q^2, 0\right)$

By symmetry, it suffices to describe the achievability strategy for corner ponit

$$(R_1, R_2) = \left(1 - q^2, 0\right), \tag{A.50}$$

when transmitters have delayed knowledge of channel state information and noiseless output feedback links are available from the receivers to the transmitters.

Our achievability strategy is carried on over $b+1$ communication blocks each block with n time instants. Transmitter one communicates fresh data bits in the first b blocks and the final block is to help receiver one decode its corresponding bits. Transmitter and receiver two act as a relay to facilitate the communication between transmitter and receiver one. At the end, using our scheme we achieve rate tuple $\frac{b}{b+1}\left(1-q^2,0\right)$ as $n \rightarrow \infty$. Finally, letting $b \rightarrow \infty$, we achieve the desired corner point.

Let W_1^j be the message of Tx_1 in block j, $j = 1,2,\ldots,b$. We assume $\mathrm{W}_1^j = a_1^j, a_2^j, \ldots, a_m^j$. We set

$$n = \left(1-q^2\right)^{-1} m + m^{2/3}. \tag{A.51}$$

Achievability strategy for block 1: At the beginning of the communication block, we assume that the bits at Tx_1 are in queue (or state) $Q^1_{1\rightarrow 1}$. At each time instant t, Tx_1 sends out a bit from $Q^1_{1\rightarrow 1}$, and this bit will leave this queue if at least one of the outgoing links from Tx_1 was equal to 1 at the time of transmission. On the other hand, Tx_2 remains silent during the first communication block. If at the end of the communication block, queue $Q^1_{1\rightarrow 1}$ is not empty, we declare error type-I and halt the transmission.

At the end of first block, using output feedback links, transmitter two has access to the bits of Tx_1 communicated in the first block. More precisely, Tx_2 has access to the bits of Tx_1 communicated in Cases 11, 12, 14, and 15 during the first communication block. Note that the bits communicated in these cases are available at Rx_2 and have to be provided to Rx_1. Transmitter two transfers these bits to $Q^1_{2\rightarrow 1|2}$.

Assuming that the transmission is not halted, let $N^1_{2\rightarrow 1|2}$ denote the number of bits in queue $Q^1_{2\to 1|2}$. The transmission strategy will be halted and an error type-II will occur, if $N^1_{2\to 1|2} > \mathbb{E}[N^1_{2\to 1|2}] + pqm^{\frac{2}{3}}$. From basic probability, we know that

$$\mathbb{E}[N^1_{2\to 1|2}] = \frac{\sum_{j=11,12,14,15} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} pqm. \tag{A.52}$$

Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II and decreases exponentially with m.

Achievability strategy for block j, $j = 2, 3, \ldots, b$: In the communication block j, Tx_2 treats the bits in $Q^{j-1}_{2\to 1|2}$ as its message and it uses a random code of rate $pq - \delta$ to transmit them. Note that the channel from Tx_2 to Rx_1 can be modeled as a point-to-point erasure channel (any time $G_{11}[t] = 1$ or $G_{21}[t] = 0$, we consider an erasure has taken place). Hence from [17], we know that for any $\epsilon, \delta > 0$ and sufficiently large block length, a rate of $pq - \delta$ is achievable from Tx_2 to Rx_1 with decoding error probability less than or equal to ϵ. Note that at rate $pq - \delta$, both receivers will be able to decode and hence remove the transmitted signal of Tx_2 at the end of communication block. If an error occurs in decoding the transmitted signal of Tx_2, we consider it as error and halt the transmission strategy.

On the other hand, the transmission strategy for Tx_1 is the same as block 1 for the first b blocks (all but the last block). At the end of communication block j, using output feedback links, transmitter two has access to the bits of Tx_1 communicated in Cases 11, 12, 14, and 15 during the communication block j. Transmitter two transfers these bits to $Q^j_{2\to 1|2}$. If at the end of the communication block, queue $Q^j_{1\to 1}$ is not empty, we declare error type-I and halt the transmission.

Assuming that the transmission is not halted, let $N^{j}_{2\to 1|2}$ denote the number of bits in queue $Q^{j}_{2\to 1|2}$. The transmission strategy will be halted and an error type-II will occur, if $N^{j}_{2\to 1|2} > \mathbb{E}[N^{j}_{2\to 1|2}] + pqm^{\frac{2}{3}}$. From basic probability, we know that

$$\mathbb{E}[N^{j}_{2\to 1|2}] = \frac{\sum_{j=11,12,14,15} \Pr\left(\text{Case j}\right)}{1 - \sum_{i=9,10,13,16} \Pr\left(\text{Case i}\right)} m = (1-q^2)^{-1} pqm. \tag{A.53}$$

Using Chernoff-Hoeffding bound, we can show that the probability of errors of types I and II and decreases exponentially with m.

Achievability strategy for block $b+1$: Finally in block $b+1$, no new data bit is transmitted and Tx_2 only communicates the bits of Tx_1 communicated in the previous block in Cases 11, 12, 14, and 15 as described before.

We can show that the probability that the transmission strategy halts at any point approaches zero as $m \to \infty$.

Decoding: At the end of block $j+1$, Rx_1 decodes the transmitted message of Tx_2 in block $j+1$ and removes it from the received signal. Together with the bits it has obtained during block j, it can decode message W^{j}_{1}. Using similar idea, Rx_2 uses backward decoding to cancel out interfernce in the previous blocks to decode all messages.

This completes the achievability proof for corner ponit

$$(R_1, R_2) = \left(1-q^2, 0\right).$$

Table A.1: Summary of Phase 1 for the Achievability Scheme of Corner Point B. Bit "a" represents a bit in $Q_{1\to 1}$ while bit "b" represents a bit in $Q_{2\to 2}$.

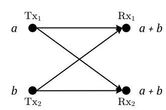
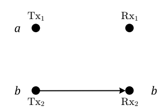
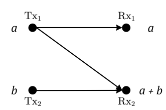
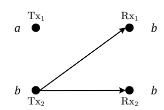
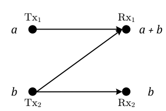
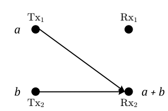
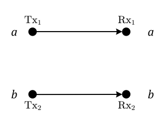
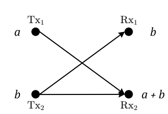
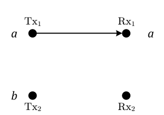
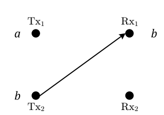
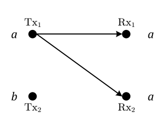
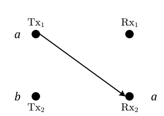
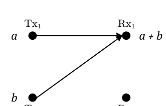
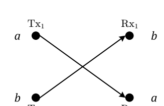
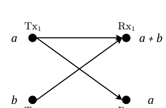
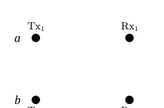
ID	ch. at time n	transition	ID	ch. at time n	transition
1		$\begin{cases} a \to Q_{1,\mathsf{C}_1} \\ b \to Q_{2,\mathsf{C}_1} \end{cases}$	9		$\begin{cases} a \to Q_{1\to 1} \\ b \to Q_{2,F} \end{cases}$
2		$\begin{cases} a \to Q_{1\to 2\|1} \\ b \to Q_{2,F} \end{cases}$	10		$\begin{cases} a \to Q_{1\to 1} \\ b \to Q_{2,F} \end{cases}$
3		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2\to 1\|2} \end{cases}$	11		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
4		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,F} \end{cases}$	12		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
5		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2\to 2} \end{cases}$	13		$\begin{cases} a \to Q_{1\to 1} \\ b \to Q_{2\to 2\|1} \end{cases}$
6		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2\to 2} \end{cases}$	14		$\begin{cases} a \to Q_{1\to 1\|2} \\ b \to Q_{2\to 2} \end{cases}$
7		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	15		$\begin{cases} a \to Q_{1\to 1\|2} \\ b \to Q_{2\to 2\|1} \end{cases}$
8		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	16		$\begin{cases} a \to Q_{1\to 1} \\ b \to Q_{2\to 2} \end{cases}$

Table A.2: Summary of Phase 2 for the Achievability Scheme of Corner Point B. Bit "a" represents a bit in $Q_{1,INT}$ while bit "b" represents a bit in $Q_{2,INT}$.

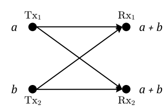
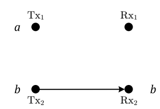
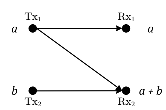
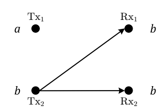
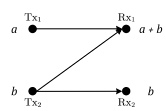
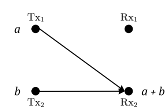
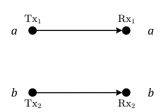
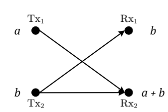
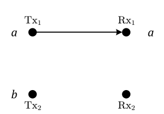
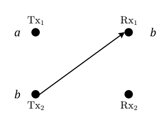
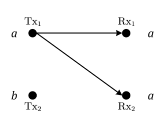
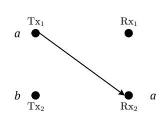
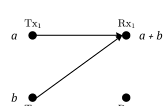
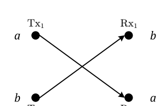
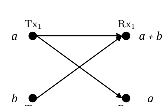
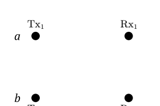
ID	ch. at time n	transition	ID	ch. at time n	transition
1		$a \rightarrow Q_{1,\mathsf{C}_1}$, $b \rightarrow Q_{2,\mathsf{C}_1}$	9		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2\rightarrow 1\|2}$
2		$a \rightarrow Q_{1\rightarrow 2\|1}$, $b \rightarrow Q_{2,F}$	10		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2,F}$
3		$a \rightarrow Q_{1,F}$, $b \rightarrow Q_{2\rightarrow 1\|2}$	11		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2,\mathsf{C}_1}$
4		$a \rightarrow Q_{1\rightarrow 2\|1}$, $b \rightarrow Q_{2\rightarrow 1\|2}$	12		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2,F}$
5		$a \rightarrow Q_{1\rightarrow 2\|1}$, $b \rightarrow Q_{2,INT}$	13		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2\rightarrow 2\|1}$
6		$a \rightarrow Q_{1,F}$, $b \rightarrow Q_{2,INT}$	14		$a \rightarrow Q_{1\rightarrow 1\|2}$, $b \rightarrow Q_{2,INT}$
7		$a \rightarrow Q_{1,\mathsf{C}_1}$, $b \rightarrow Q_{2,INT}$	15		$a \rightarrow Q_{1\rightarrow 1\|2}$, $b \rightarrow Q_{2\rightarrow 2\|1}$
8		$a \rightarrow Q_{1,F}$, $b \rightarrow Q_{2,INT}$	16		$a \rightarrow Q_{1,INT}$, $b \rightarrow Q_{2,INT}$

Table A.3: Summary of Phase 1 for the Achievability Scheme of Corner Point C. Bit "a" represents a bit in $Q_{1\to1}$ while bit "b" represents a bit in $Q_{2\to2}$.

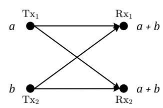
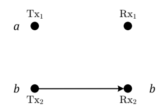
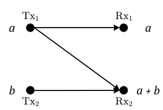
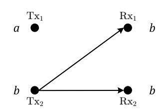
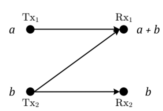
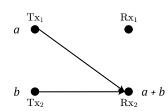
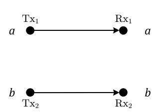
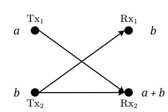
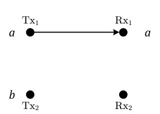
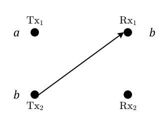
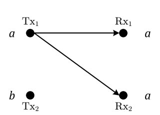
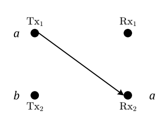
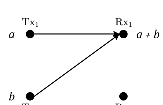
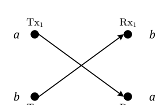
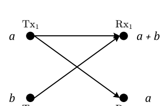
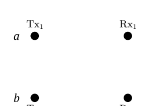
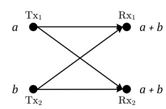
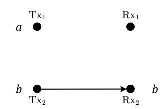
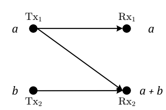
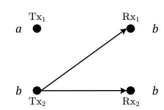
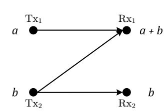
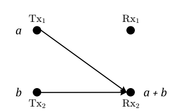
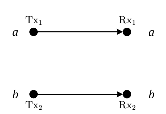
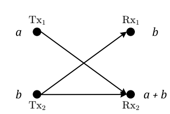
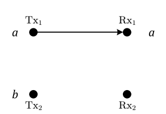
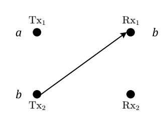
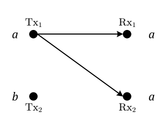
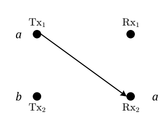
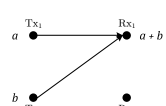
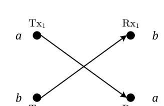
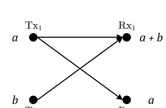
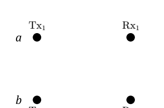
ID	ch. at time n	transition	ID	ch. at time n	transition
1		$\begin{cases} a \to Q_{1,\mathrm{C}_1} \\ b \to Q_{2,F} \end{cases}$	9		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2,F} \end{cases}$
2		$\begin{cases} a \to Q_{1,OP} \\ b \to Q_{2,F} \end{cases}$	10		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2,F} \end{cases}$
3		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,OP} \end{cases}$	11		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
4		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,F} \end{cases}$	12		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
5		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2\to2} \end{cases}$	13		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to2\vert1} \end{cases}$
6		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2\to2} \end{cases}$	14		$\begin{cases} a \to Q_{1\to1\vert2} \\ b \to Q_{2\to2} \end{cases}$
7		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	15		$\begin{cases} a \to Q_{1\to1\vert2} \\ b \to Q_{2\to2\vert1} \end{cases}$
8		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	16		$\begin{cases} a \to Q_{1\to1} \\ b \to Q_{2\to2} \end{cases}$

Table A.4: Summary of Phase 2 for the Achievability Scheme of Corner Point B. Bit "a" represents a bit in $Q_{1,INT}$ while bit "b" represents a bit in $Q_{2,INT}$.

ID	ch. at time n	transition	ID	ch. at time n	transition
1		$\begin{cases} a \to Q_{1,\mathsf{C}_1} \\ b \to Q_{2,F} \end{cases}$	9		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,OP} \end{cases}$
2		$\begin{cases} a \to Q_{1,OP} \\ b \to Q_{2,OP} \end{cases}$	10		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
3		$\begin{cases} a \to Q_{1,\mathsf{C}_1} \\ b \to Q_{2,OP} \end{cases}$	11		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,OP} \end{cases}$
4		$\begin{cases} a \to Q_{1,OP} \\ b \to Q_{2,OP} \end{cases}$	12		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,F} \end{cases}$
5		$\begin{cases} a \to Q_{1,OP} \\ b \to Q_{2,INT} \end{cases}$	13		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2\to 2\vert 1} \end{cases}$
6		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	14		$\begin{cases} a \to Q_{1\to 1\vert 2} \\ b \to Q_{2,INT} \end{cases}$
7		$\begin{cases} a \to Q_{1,OP} \\ b \to Q_{2,INT} \end{cases}$	15		$\begin{cases} a \to Q_{1\to 1\vert 2} \\ b \to Q_{2\to 2\vert 1} \end{cases}$
8		$\begin{cases} a \to Q_{1,F} \\ b \to Q_{2,INT} \end{cases}$	16		$\begin{cases} a \to Q_{1,INT} \\ b \to Q_{2,INT} \end{cases}$

APPENDIX B

CHAPTER 2 OF APPENDIX

B.1 Lattice Quantizer

We will incorporate lattice quantization in our transmission stratgey. Here, we provide the basic concepts and the result we need. For an elaborated discussion see [76, 77].

An n-dimensional lattice Λ is defined by a set of n basis column vectors $g_1, g_2, \ldots, g_n$ in $\mathbb{R}^n$. The lattice Λ is composed of all integral combinations of the basis vectors, *i.e.*

$$\Lambda = \{\ell = \mathcal{G}.i : \ i \in \mathbb{Z}^n\} \tag{B.1}$$

where the $n \times n$ generator matrix $\mathcal{G}$ is given by $\mathcal{G} = [g_1|g_2|\ldots|g_n]$.

A quantizer is defined by a set of code points and a partition which is associated with it. The code points of an n-dimensional *lattice quantizer* form an n-dimensional lattice $\Lambda = \{\ell_i\}$, *i.e.*

$$\ell_i \in \mathbb{R}^n, \qquad \ell_0 = 0, \qquad \ell_i + \ell_j \in \Lambda \quad \forall i, j. \tag{B.2}$$

The partition $\mathcal{P} = \{P_i\}$ associated with the lattice quantizer is a collection of disjoint regions (whose union covers $\mathbb{R}^n$) which satisfy

$$P_i = \ell_i + P_0 = \{x^n : x^n - \ell_i \in P_0\} \tag{B.3}$$

i.e. the ith cell is a shift of the basic cell P_0 by the ith point of the lattice. The lattice quantizer $Q_n = \{\Lambda, \mathcal{P}\} : \mathbb{R}^n \to \mathbb{R}^n$ maps every vector $x^n \in \mathbb{R}^n$ into the lattice point that is associated with the cell containing x^n, *i.e.*

$$Q_n(x^n) = \ell_i, \qquad \text{if } x^n \in P_i. \tag{B.4}$$

The quantization noise is defined as $Z_Q^n = Q_n(x^n + \underline{u}) - \left(x^n + \underline{u}\right)$, where $\underline{u}$ is the dither for dithered quantization.

In [76], authors have shown that the optimal lattice quantizer is white and the autocorrelation of its quantization noise is

$$\mathbf{A}_Z = G_n^{\text{opt}} V^{2/n} \mathbf{I}, \tag{B.5}$$

where V is the volume of the basic cell and $\mathbf{I}$ is the identity matrix. Furthermore, as $n \to \infty$, $G_n^{\text{opt}} \to 1/2\pi e$ (*i.e.* that there exist good lattice quantizers). Due to the dither, the distortion in this procedure is *independent* of the source signal.

B.2 Determining D such that $R_Q(D)/C_{2\times 1} \leq 1$

As mentioned in Section 3.6, we are interested in $P > 2$. Using Jensen's inequality, we have

$$\begin{aligned} R_Q(D) &= \mathbb{E}\left[\log_2\left(1 + \frac{P}{2D}\left(\|\mathbf{g}\|_2^2 + \|\mathbf{h}\|_2^2\right)\right)\right] \\ &\leq \log_2\left(1 + \frac{P}{2D}\mathbb{E}\left[\|\mathbf{g}\|_2^2 + \|\mathbf{h}\|_2^2\right]\right) = \log_2\left(1 + \frac{2P}{D}\right). \end{aligned} \tag{B.6}$$

Moreover, from [54], we have

$$\begin{aligned} C_{2\times 1} &= \int_0^\infty \log_2\left(1 + P\lambda/2\right)\lambda e^{-\lambda} d\lambda \\ &= \int_0^1 \log_2\left(1 + P\lambda/2\right)\lambda e^{-\lambda} d\lambda + \int_1^\infty \log_2\left(1 + P\lambda/2\right)\lambda e^{-\lambda} d\lambda \end{aligned}$$

$$= \sum_{m=1}^{\infty} \int_{2^{-m}}^{2^{1-m}} \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda + \int_1^2 \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda$$
$$+ \sum_{j=1}^{45} \int_{2+.1(j-1)}^{2+.1j} \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda + \int_{6.5}^{\infty} \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda$$

$$\overset{(a)}{\geq} \log_2(1+P/2) \int_0^1 \lambda e^{-\lambda} d\lambda - \underbrace{\sum_{m=1}^{\infty} m \int_{2^{-m}}^{2^{1-m}} \lambda e^{-\lambda} d\lambda}_{<0.4} + \log_2(1+P/2) \int_1^2 \lambda e^{-\lambda} d\lambda$$
$$+ \log_2(1+P/2) \int_2^{6.5} \lambda e^{-\lambda} d\lambda + \underbrace{\sum_{j=1}^{45} \log_2\left[\frac{1+(2+.1(j-1))P/2}{1+P/2}\right] \int_{2+.1(j-1)}^{2+.1j} \lambda e^{-\lambda} d\lambda}_{>0.4}$$
$$+ \int_{6.5}^{\infty} \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda > \log_2(1+P/2) \int_0^{\infty} \lambda e^{-\lambda} d\lambda = \log_2(1+P/2)\,. \quad \text{(B.7)}$$

where (a) holds since

$$\begin{aligned}
&\sum_{m=1}^{\infty} \int_{2^{-m}}^{2^{1-m}} \log_2(1+P\lambda/2)\,\lambda e^{-\lambda} d\lambda \\
&\geq \sum_{m=1}^{\infty} \int_{2^{-m}}^{2^{1-m}} \log_2(1+2^{-m}P/2)\,\lambda e^{-\lambda} d\lambda \qquad \text{(B.8)} \\
&\geq \sum_{m=1}^{\infty} \int_{2^{-m}}^{2^{1-m}} \left[\log_2(1+P/2) - \log_2(2^m)\right] \lambda e^{-\lambda} d\lambda \\
&= \log_2(1+P/2) \int_0^1 \lambda e^{-\lambda} d\lambda - \sum_{m=1}^{\infty} m \int_{2^{-m}}^{2^{1-m}} \lambda e^{-\lambda} d\lambda,
\end{aligned}$$

and $\sum_{m=1}^{\infty} m \int_{2^{-m}}^{2^{1-m}} \lambda e^{-\lambda} d\lambda$ converges since

$$\left\{ m \int_{2^{-m}}^{2^{1-m}} \lambda e^{-\lambda} d\lambda \right\}_{m=1}^{\infty}, \qquad \text{(B.9)}$$

is a Cauchy sequence. Thus, we have $R_Q(4)/C_{2\times 1} < 1$.

APPENDIX C

CHAPTER 3 OF APPENDIX

C.1 Converse Proof for Case (2) of Theorem 4.2

Converse proof for case (2): If a destination is only connected to relay A_2, assign channel gain of 0 to the link from A_2 to such destination. Set all the other channel gains equal to $h \in \mathbb{C}$. We claim that in such network, a destination connected to both relays should be able to decode all messages (note that with our choice of channel gains, there is no message for destinations that are only connected to relay A_2).

Destinations that are connected to both relays receive same signals with different noise terms. Therefore, since each one of them is able to decode its message, then it should be able to decode the rest of the messages intended for destinations that are connected to both relays. They decode and remove such messages from the received signal. The remaining signal is the same (up to noise terms) as that of destinations that are only connected to relay A_1. Therefore, those messages are also decodable at a destination that is connected to both relays.

We assume local view at the sources, therefore to achieve a normalized sum-rate of α, each source should transmit at a rate greater than or equal to $\alpha \log(1 + |h|^2) - \tau$, since from each source's point of view, it is possible that the other S-D pairs have capacity 0. We get

$$d_{\text{out}}(\mathsf{A}_1)(\alpha \log(1 + |h|^2) - \tau) \leq \log(1 + d_{\text{out}}(\mathsf{A}_1) \times |h|^2), \tag{C.1}$$

or equivalently

$$(d_{\text{out}}(\mathsf{A}_1)\alpha - 1)\log(1+|h|^2) \leq \log(d_{\text{out}}(\mathsf{A}_1)) + d_{\text{out}}(\mathsf{A}_1)\tau. \tag{C.2}$$

Since this has to hold for all values of h, and α and τ are independent of h, we get $\alpha \leq \frac{1}{d_{\text{out}}(\mathsf{A}_1)}$.

C.2 Proof of Lemma 4.1

First note that by increasing noise variances and decreasing power constraint, we only decrease the capacity. Thus, we get $C(\kappa\sigma^2, P/\kappa) \leq C(\sigma^2, P)$. To prove the other inequality, we use the results in [9]. The cut-set bound $\bar{C}$ is defined as

$$\bar{C}(\sigma^2, P) = \max_{p(\{X_j\}_{\forall j \in \mathcal{V}})} \min_{\Omega \in \Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}), \tag{C.3}$$

where $\Lambda_D = \{\Omega : \mathsf{S} \in \Omega, \mathsf{D} \in \Omega^c\}$ is the set of all S-D cuts. [1] Also $\bar{C}_{i.i.d}(\sigma^2, P) = \min_{\Omega \in \Lambda_D} \log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_\Omega \mathbf{G}_\Omega^*|$ is the cut-set bound evaluated for i.i.d. $\mathcal{N}(0, P)$ input distributions and $\mathbf{G}_\Omega$ is the transfer matrix associated with the cut Ω, *i.e.* the matrix relating the vector of all the inputs at the nodes in Ω, denoted by $\mathbf{X}_\Omega$, to the vector of all the outputs in Ω^c, denoted by $\mathbf{Y}_{\Omega^c}$, as in $\mathbf{Y}_{\Omega^c} = \mathbf{G}_\Omega \mathbf{X}_\Omega + \mathbf{Z}_{\Omega^c}$ where $\mathbf{Z}_{\Omega^c}$ is the noise vector. In [9], it has been shown that

$$\bar{C}_{i.i.d}(\sigma^2, P) - 15|\mathcal{V}| \leq C(\sigma^2, P) \leq \bar{C}_{i.i.d}(\sigma^2, P) + 2|\mathcal{V}|, \tag{C.4}$$

where $|\mathcal{V}|$ is the total number of nodes in the network. Similarly, we have

$$\begin{aligned} &\bar{C}_{i.i.d}(\kappa\sigma^2, P/\kappa) - 15|\mathcal{V}| \leq C(\kappa\sigma^2, P/\kappa) \\ &C(\kappa\sigma^2, P/\kappa) \leq \bar{C}_{i.i.d}(\kappa\sigma^2, P/\kappa) + 2|\mathcal{V}|. \end{aligned} \tag{C.5}$$

[1] A cut Ω is a subset of $\mathcal{V}$ such that $\mathsf{S} \in \Omega$, $\mathsf{D} \notin \Omega$, and $\Omega^c = \mathcal{V} \setminus \Omega$.

Now, we will show that

$$C(\sigma^2, P) - C(\kappa\sigma^2, P/\kappa) \leq |\mathcal{V}| \left(2 \log \kappa + 17\right). \tag{C.6}$$

For any S-D cut $\Omega \in \Lambda_D$, $\frac{P}{\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*$ is a positive semi-definite matrix. Hence, there exists a unitary matrix $\mathbf{U}$ such that $\mathbf{U}\mathbf{G}_{diag}\mathbf{U}^* = \frac{P}{\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*$ where $\mathbf{G}_{diag}$ is a diagonal matrix. Refer to the non-zero elements in $\mathbf{G}_{diag}$ as g_{ii}'s. We then have

$$\begin{aligned}
&\log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| - \log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| \\
&= \log|\mathbf{I} + \mathbf{U}\mathbf{G}_{diag}\mathbf{U}^*| - \log|\mathbf{I} + \frac{1}{\kappa^2}\mathbf{U}\mathbf{G}_{diag}\mathbf{U}^*| \\
&= \log|\mathbf{U}\mathbf{U}^* + \mathbf{U}\mathbf{G}_{diag}\mathbf{U}^*| - \log|\mathbf{U}\mathbf{U}^* + \frac{1}{\kappa^2}\mathbf{U}\mathbf{G}_{diag}\mathbf{U}^*| \\
&= \log\left(|\mathbf{U}||\mathbf{I} + \mathbf{G}_{diag}||\mathbf{U}^*|\right) - \log\left(|\mathbf{U}||\mathbf{I} + \frac{1}{\kappa^2}\mathbf{G}_{diag}||\mathbf{U}^*|\right) \\
&= \log|\mathbf{I} + \mathbf{G}_{diag}| - \log|\mathbf{I} + \frac{1}{\kappa^2}\mathbf{G}_{diag}| \\
&= \operatorname{tr}\log\left(\mathbf{I} + \mathbf{G}_{diag}\right) - \operatorname{tr}\log\left(\mathbf{I} + \frac{1}{\kappa^2}\mathbf{G}_{diag}\right) \\
&= \sum_i \log\left(1 + g_{ii}\right) - \sum_i \log\left(1 + \frac{1}{\kappa^2}g_{ii}\right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_i \log\left(\frac{1 + g_{ii}}{1 + \frac{1}{\kappa^2}g_{ii}}\right) \\
&\overset{(a)}{\leq} \sum_i \lim_{g_{ii}\to\infty} \log\left(\frac{1 + g_{ii}}{1 + \frac{1}{\kappa^2}g_{ii}}\right) \\
&= \sum_i \log\kappa^2 \leq 2|\mathcal{V}|\log\kappa,
\end{aligned} \tag{C.7}$$

where (a) follows from the fact that $\frac{1+g_{ii}}{1+\frac{1}{\kappa^2}g_{ii}}$ is monotonically increasing in g_{ii}.

Now suppose that $\min_{\Omega\in\Lambda_D}\log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| = \log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_{\Omega'}\mathbf{G}_{\Omega'}^*|$. Hence, from (C.7), we have

$$\begin{aligned}
&\min_{\Omega\in\Lambda_D}\log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| - \min_{\Omega\in\Lambda_D}\log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| \\
&= \min_{\Omega\in\Lambda_D}\log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_\Omega\mathbf{G}_\Omega^*| - \log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_{\Omega'}\mathbf{G}_{\Omega'}^*|
\end{aligned}$$

$$\leq \log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_{\Omega'}\mathbf{G}^*_{\Omega'}| - \log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_{\Omega'}\mathbf{G}^*_{\Omega'}|$$
$$\overset{(a)}{\leq} 2|\mathcal{V}|\log\kappa, \tag{C.8}$$

where (a) follows from (C.7). Hence, from (C.4) and (C.5) we have

$$C(\sigma^2, P) - C(\kappa\sigma^2, P/\kappa) \leq \min_{\Omega\in\Lambda_D} \log|\mathbf{I} + \frac{P}{\sigma^2}\mathbf{G}_{\Omega}\mathbf{G}^*_{\Omega}|$$
$$- \min_{\Omega\in\Lambda_D} \log|\mathbf{I} + \frac{P}{\kappa^2\sigma^2}\mathbf{G}_{\Omega}\mathbf{G}^*_{\Omega}| + 17|\mathcal{V}| \overset{(a)}{\leq} |\mathcal{V}|\left(2\log\kappa + 17\right), \tag{C.9}$$

where (a) follows from (C.8). Therefore, we get

$$C(\sigma^2, P) - \tau \leq C(\kappa\sigma^2, P/\kappa) \leq C(\sigma^2, P), \tag{C.10}$$

where $\tau = |\mathcal{V}|\left(2\log\kappa + 17\right)$ is a constant independent of channel gains.

C.3 Proof of Lemma 4.2

Consider a $3 \times 3 \times 3$ network where there exists a path from S_i to D_j, for some $i \neq j$, without loss of generality suppose $i = 2$ and $j = 1$. Then, one of the graphs in Fig. C.1 is a subgraph of the network connectivity graph $\mathcal{G}$. First, suppose the graph in Fig. C.1(a) is a subgraph of $\mathcal{G}$. Assign channel gain of $h \in \mathbb{C}$ to the links of the subgraph, and channel gain of 0 to the links that are not in the graph of Fig. C.1(a).

With this assignment of the channel gains, it is straightforward to see that

$$H\left(W_i|Y^n_{\mathsf{A}_i}, L_{\mathsf{A}_i}, \mathsf{SI}\right) \leq n\epsilon_n, \qquad i = 1, 2. \tag{C.11}$$

Basically, relay A_1 has all the information that destinations D_1 and D_2 require in order to decode their messages. We conclude that

$$n\left(R_1 + R_2 - \epsilon_n\right) \leq h\left(Y^n_{\mathsf{A}_1}|L_{\mathsf{A}_1}, \mathsf{SI}\right), \tag{C.12}$$

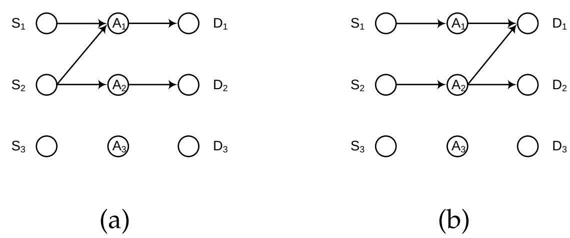


Figure C.1: *If a path exists from S_i to D_j, for some $i \neq j$, then the network connectivity graph has the graph in (a) or (b) as subgraph.*

where $\epsilon_n \to 0$ as $n \to \infty$.

The MAC capacity at relay A_1, gives us

$$2(\alpha \log(1 + |h|^2) - \tau) \leq \log(1 + 2 \times |h|^2), \tag{C.13}$$

which results in

$$(2\alpha - 1) \log(1 + |h|^2) \leq \log(2) + 2\tau. \tag{C.14}$$

Since this has to hold for all values of h, and α and τ are independent of h, we get $\alpha \leq \frac{1}{2}$. The proof for the graph in Fig. C.1(b) is very similar and omitted.

APPENDIX D

CHAPTER 4 OF APPENDIX

D.1 Proof of Theorem 5.2

Proof of (5.21a) (cutset bound): Starting with Fano's inequality, we get

$$N(R_1 - \epsilon_N) \leq I(W_1; Y_1^N) \leq \sum H(Y_{1i}),$$

where the second inequality follows from the fact that conditioning reduces entropy. If (R_1, R_2) is achievable, then $\epsilon_N \to 0$ as $N \to \infty$. Thus we obtain the left term of the bound. Notice that this is a cutset bound, as the bound is obtained assuming that the two transmitters fully collaborate.

To obtain the right term, we consider

$$\begin{aligned}
N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N, Y_2^N, W_2) \\
&\overset{(a)}{=} \sum H(Y_{1i}, Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}, X_2^i) \\
&= \sum H(Y_{1i} | W_2, Y_1^{i-1}, Y_2^i, X_2^i) \\
&+ \sum H(Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}, X_2^i) \\
&\overset{(b)}{=} \sum H(Y_{1i} | W_2, Y_1^{i-1}, Y_{2i}, X_{2i}, U_{1i}) \\
&+ \sum H(Y_{2i} | W_2, Y_1^{i-1}, X_{2i}, U_{1i}) \\
&\overset{(c)}{\leq} \sum H(Y_{1i} | X_{2i}, Y_{2i}, U_{1i}) + \sum H(Y_{2i} | X_{2i}, U_{1i}) \\
&\overset{(d)}{\leq} \sum H(Y_{1i} | V_{1i}, V_{2i}, U_{1i}) + \sum H(Y_{2i} | X_{2i}, U_{1i}),
\end{aligned}$$

where (a) follows from the fact that W_1 is independent from W_2, and X_2^i is a function of (W_2, Y_2^{i-1}); (b) follows from the fact that $U_{1i} := (X_2^{i-1}, \tilde{Y}_2^{i-1})$ and $\tilde{Y}_2^{i-1}$ is a function of Y_2^{i-1}; (c) follows from the fact that conditioning reduces entropy; (d) follows from the fact that (V_{1i} is a function of (X_{2i}, Y_{2i}), Y_{2i} is a function of (X_{2i}, V_{1i})

and V_{2i} is a function of X_{2i}. Thus we get the right term of the bound. Notice that this is a cutset bound, as the bound is obtained assuming that the two receivers fully collaborate.

Proof of (5.21b) (cutset bound): Starting with Fano's inequality, we get:

$$
\begin{aligned}
N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N, \tilde{Y}_2^N, W_2) \\
&\overset{(a)}{=} \sum H(Y_{1i}, \tilde{Y}_{2i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_2^i) \\
&= \sum H(Y_{1i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_2^i) \\
&+ \sum H(\tilde{Y}_{2i} | W_2, Y_1^i, \tilde{Y}_2^{i-1}, X_2^i) \\
&\overset{(b)}{\leq} \sum H(Y_{1i} | X_{2i}, U_{1i}) + \sum H(\tilde{Y}_{2i} | Y_{1i}, X_{2i}) \\
&\overset{(c)}{\leq} \sum H(Y_{1i} | X_{2i}, U_{1i}) + NC_{\mathsf{FB2}},
\end{aligned}
$$

where (a) follows from the fact that W_1 is independent from W_2, and X_2^i is a function of $(W_2, \tilde{Y}_2^{i-1})$; (b) follows from the fact that conditioning reduces entropy; (d) follows from $H(\tilde{Y}_{2i} | Y_{1i}, X_{2i}) \leq C_{\mathsf{FB2}}$. Therefore, we get the desired bound.

Proof of (5.21e): Starting with Fano's inequality, we get

$$
\begin{aligned}
N(R_1 + R_2 - \epsilon_N) &\leq I(W_1; Y_1^N | W_2) + I(W_2; Y_2^N) \\
&= H(Y_1^N | W_2) + I(W_2; Y_2^N) \\
&= H(Y_1^N | W_2) + H(Y_2^N) \\
&- \left\{ H(Y_1^N, Y_2^N | W_2) - H(Y_1^N | Y_2^N, W_2) \right\} \\
&= H(Y_1^N | Y_2^N, W_2) - H(Y_2^N | Y_1^N, W_2) + H(Y_2^N) \\
&\overset{(a)}{=} \sum H(Y_{1i} | W_2, Y_1^{i-1}, Y_2^N, X_2^i, V_{1i}) + H(Y_2^N) \\
&\overset{(b)}{\leq} \sum \left[H(Y_{1i} | V_{2i}, V_{1i}, U_{1i}) + H(Y_{2i}) \right],
\end{aligned}
$$

where (a) follows from the fact that X_2^i is a function of (W_2, Y_2^{i-1}) and V_{1i} is a function of (X_{2i}, Y_{2i}); (b) follows from the fact that U_{1i} is a function of (X_2^{i-1}, Y_2^{i-1}) and conditioning reduces entropy.

Proof of (5.21h):

$$
\begin{aligned}
& N(2R_1 + R_2 - \epsilon_N) \\
& \leq I(W_1; Y_1^N) + I(W_1; Y_1^N|W_2) + I(W_2; Y_2^N) \\
& \overset{(a)}{\leq} [H(Y_1^N) - H(Y_1^N|W_1)] \\
& + I(W_1; Y_1^N, V_1^N|W_2) + [H(Y_2^N) - H(Y_2^N|W_2)] \\
& \overset{(b)}{=} [H(Y_1^N) - H(V_2^N|W_1)] + H(V_1^N|W_2) \\
& + H(Y_1^N|W_2, V_1^N) + [H(Y_2^N) - H(V_1^N|W_2)] \\
& = H(Y_1^N) + H(Y_1^N|W_2, V_1^N) \\
& + H(Y_2^N) - [H(V_2^N) - I(W_1; V_2^N)] \\
& \overset{(c)}{\leq} I(W_1; V_2^N) + H(Y_1^N) + H(Y_1^N|W_2, V_1^N) \\
& + H(Y_2^N, V_2^N) - H(V_2^N) \\
& \overset{(d)}{\leq} I(W_1; V_2^N, W_2, \tilde{Y}_2^N) + H(Y_1^N) \\
& + H(Y_1^N|W_2, V_1^N) + H(Y_2^N|V_2^N) \\
& \overset{(e)}{=} I(W_1; \tilde{Y}_2^N|W_2) + H(Y_1^N) \\
& + H(Y_1^N|W_2, V_1^N, X_2^N, \tilde{Y}_2^N) + H(Y_2^N|V_2^N) \\
& \overset{(f)}{\leq} NC_{\mathsf{FB2}} + \sum H(Y_{1i}) + \sum H(Y_{2i}|V_{2i}) \\
& + \sum H(Y_{1i}|V_{1i}, V_{2i}, U_{1i}),
\end{aligned}
$$

where (a) follows from the fact that adding information increases mutual information; (b) follows from Claim 5.1; (c) follows from providing V_2^N to receiver 2; (d) follows from the fact that adding information increases mutual information; follows from the fact that V_k^N is a function of $(W_k, \tilde{Y}_k^{N-1})$; (e) follows from the fact that X_2^N is a function of (W_2, V_1^{N-1}) (by Claim 5.2) and $\tilde{Y}_2^N$ is a function of (X_2^N, V_1^N); (f) follows from the fact that $U_{1i} := (X_2^{i-1}, \tilde{Y}_2^{i-1})$, $U_{2i} := (X_1^{i-1}, \tilde{Y}_1^{i-1})$, $H(\tilde{Y}_2^N|W_2) \leq NC_{\mathsf{FB2}}$ and conditioning reduces entropy.

To complete the proof, we will show that given $U_i := (U_{1i}, U_{2i})$, X_{1i} and X_{2i} are conditionally independent. Remember that our input distribution is of the form of $p(u_1, u_2)p(x_1|u_1, u_2)p(x_2|u_1, u_2)$.

Claim D.1 *Given $U_i := (U_{1i}, U_{2i}) = (X_2^{i-1}, \tilde{Y}_2^{i-1}, X_1^{i-1}, \tilde{Y}_1^{i-1})$, X_{1i} and X_{2i} are conditionally independent.*

Proof: The proof is based on the dependence-balance-bound technique [28, 74]. For completeness we describe details. We first show that $I(W_1; W_2|U_i) = 0$, implying that W_1 and W_2 are independent given U_i. We will then show that X_{1i} and X_{2i} are conditionally independent given U_i.

Consider

$$
\begin{aligned}
0 \leq I(W_1; W_2|U_i) &\overset{(a)}{=} I(W_1; W_2|U_i) - I(W_1; W_2) \\
&\overset{(b)}{=} -H(W_1) - H(W_2) - H(U_i) + H(W_1, W_2) \\
&\quad + H(W_1, U_i) + H(W_2, U_i) - H(W_1, W_2, U_i) \\
&\overset{(c)}{=} -H(U_i) + H(U_i|W_1) + H(U_i|W_2) \\
&\overset{(d)}{=} \sum_{j=1}^{i-1} \Big[-H(X_{1j}, X_{2j}|X_1^{j-1}, X_2^{j-1}) \\
&\qquad + H(X_{1j}, X_{2j}|W_1, X_1^{j-1}, X_2^{j-1}) \\
&\qquad + H(X_{1j}, X_{2j}|W_2, X_1^{j-1}, X_2^{j-1}) \Big] \\
&\overset{(e)}{=} \sum_{j=1}^{i-1} \Big[-H(X_{1j}, X_{2j}|X_1^{j-1}, X_2^{j-1}) \\
&\qquad + H(X_{2j}|W_1, X_1^{j}, X_2^{j-1}) + H(X_{1j}|W_2, X_1^{j-1}, X_2^{j}) \Big] \\
&= \sum_{j=1}^{i-1} \Big[-H(X_{1j}|X_1^{j-1}, X_2^{j-1}) + H(X_{1j}|W_2, X_1^{j-1}, X_2^{j}) \\
&\qquad - H(X_{2j}|X_1^{j}, X_2^{j-1}) + H(X_{2j}|W_1, X_1^{j}, X_2^{j-1}) \Big] \overset{(f)}{\leq} 0,
\end{aligned}
$$

where (a) follows from $I(W_1; W_2) = 0$; (b) follows from the chain rule; (c) follows from the chain rule and $H(U_i|W_1, W_2) = 0$; (d) follows from the fact that $(\tilde{Y}_{1j}, \tilde{Y}_{2j})$ is a function of $(\tilde{X}_{1j}, \tilde{X}_{2j})$; ($e$) follows from the fact that X_{kj} is a function of $(W_k, X_1^{j-1}, X_2^{j-1})$; ($f$) follows from the fact that conditioning reduces entropy. Therefore, $I(W_1; W_2|U_i) = 0$, which shows the independence of W_1 and W_2 given U_i.

Notice that X_{ki} is a function of $(W_k, X_1^{i-1}, X_2^{i-1})$. Hence, it easily follows that $I(X_{1i}; X_{2i}|U_i) = I(X_{1i}; X_{2i}|X_1^{i-1}, X_2^{i-1}) = 0$. This proves the independence of X_{1i} and X_{2i} given U_i. ■

D.2 Achievability Proof of Theorem 5.3

With the choice of distribution given in (5.23), we have

$$\delta_1 = I(\hat{Y}_1; Y_1|U, U_2, X_1) = 0, \tag{D.1a}$$

$$\delta_2 = I(\hat{Y}_2; Y_2|U, U_1, X_2) = 0, \tag{D.1b}$$

$$I(U, V_2, X_1; Y_1) = \max(n_{11}, n_{21}), \tag{D.1c}$$

$$I(U, V_1, X_2; Y_2) = \max(n_{22}, n_{12}), \tag{D.1d}$$

$$I(X_1; Y_1|U, V_1, V_2) = (n_{11} - n_{12})^+, \tag{D.1e}$$

$$I(X_2; Y_2|U, V_1, V_2) = (n_{22} - n_{21})^+, \tag{D.1f}$$

$$I(U_2; Y_1|U, X_1) = \min(n_{21}, C_{\mathsf{FB1}}), \tag{D.1g}$$

$$I(U_1; Y_2|U, X_2) = \min(n_{12}, C_{\mathsf{FB2}}), \tag{D.1h}$$

$$\begin{aligned} &I(X_1; Y_1|U, U_1, V_2) = (n_{11} - n_{12})^+ \\ &+ \min\{n_{11}, (n_{12} - C_{\mathsf{FB2}})^+\}, \end{aligned} \tag{D.1i}$$

$$I(X_2; Y_2|U, U_2, V_1) = (n_{22} - n_{21})^+$$

$$+\min\left\{n_{22},(n_{21}-C_{\mathsf{FB1}})^+\right\}, \tag{D.1j}$$

$$I(X_1,V_2;Y_1|U,V_1,U_2)= \tag{D.1k}$$

$$(n_{21}-C_{\mathsf{FB1}})^+ + \left[(n_{11}-n_{12})^+ - n_{21}\right]^+$$

$$+\min\left\{(n_{11}-n_{12})^+,\min(n_{21},C_{\mathsf{FB1}})\right\},$$

$$I(X_2,V_1;Y_2|U,V_2,U_1)= \tag{D.1l}$$

$$(n_{12}-C_{\mathsf{FB2}})^+ + \left[(n_{22}-n_{21})^+ - n_{12}\right]^+$$

$$+\min\left\{(n_{22}-n_{21})^+,\min(n_{12},C_{\mathsf{FB2}})\right\},$$

$$I(X_1,V_2;Y_1|U,U_1,U_2)= \tag{D.1m}$$

$$(n_{11}-n_{12})^+ + \min\left\{n_{11},(n_{12}-C_{\mathsf{FB2}})^+\right\}$$

$$+\left[n_{21}-\max\left\{n_{11},\min(n_{21},C_{\mathsf{FB1}})\right\}\right]^+$$

$$+\Big[\min\Big\{n_{21},\left[n_{11}-(n_{12}-C_{\mathsf{FB2}})^+\right]^+\Big\}$$

$$-\max\left\{\min(n_{21},C_{\mathsf{FB1}}),(n_{11}-n_{12})^+\right\}\Big]^+,$$

$$I(X_2,V_1;Y_2|U,U_1,U_2)= \tag{D.1n}$$

$$(n_{22}-n_{21})^+ + \min\left\{n_{22},(n_{21}-C_{\mathsf{FB1}})^+\right\}$$

$$+\left[n_{12}-\max\left\{n_{22},\min(n_{12},C_{\mathsf{FB2}})\right\}\right]^+$$

$$+\Big[\min\Big\{n_{12},\left[n_{22}-(n_{21}-C_{\mathsf{FB1}})^+\right]^+\Big\}$$

$$-\max\left\{\min(n_{12},C_{\mathsf{FB2}}),(n_{22}-n_{21})^+\right\}\Big]^+.$$

Using this computation, one can show that the inequalities of (5.10g) and (5.10h) are implied by (5.10b), (5.10d), (5.10e) and (5.10f); the inequality (5.10k) is implied by (5.10b), (5.10d) and (5.10j); and the inequality (5.10m) is implied by (5.10b), (5.10d) and (5.10l). We omit the tedious calculation. With further computation, we get:

$$R_1 \le \max(n_{11},n_{21}) \tag{D.2a}$$

$$R_1 \le (n_{11}-n_{12})^+ + \min\left\{n_{11},(n_{12}-C_{\mathsf{FB2}})^+\right\}$$

$$
\begin{aligned}
&& + \quad \min(n_{12}, C_{\mathsf{FB2}}) && \text{(D.2b)}\\
R_2 &\leq& \max(n_{22}, n_{12}) && \text{(D.2c)}\\
R_2 &\leq& (n_{22}-n_{21})^+ + \min\{n_{22}, (n_{21}-C_{\mathsf{FB1}})^+\} &&\\
&& + \quad \min(n_{21}, C_{\mathsf{FB1}}) && \text{(D.2d)}\\
R_1+R_2 &\leq& (n_{11}-n_{12})^+ + \max(n_{22}, n_{12}) && \text{(D.2e)}\\
R_1+R_2 &\leq& (n_{22}-n_{21})^+ + \max(n_{11}, n_{21}) && \text{(D.2f)}
\end{aligned}
$$

$$
\begin{aligned}
R_1+R_2 &\leq& \max\{(n_{11}-n_{12})^+, n_{21}\} && \text{(D.2g)}\\
&+& \max\{(n_{22}-n_{21})^+, n_{12}\} &&\\
&+& \min\{(n_{11}-n_{12})^+, n_{21}, C_{\mathsf{FB1}}\} &&\\
&+& \min\{(n_{22}-n_{21})^+, n_{12}, C_{\mathsf{FB2}}\} &&\\
2R_1+R_2 &\leq& (n_{11}-n_{12})^+ + \max(n_{11}, n_{21}) && \text{(D.2h)}\\
&+& \max\{(n_{22}-n_{21})^+, n_{12}\} &&\\
&+& \min\{(n_{22}-n_{21})^+, n_{12}, C_{\mathsf{FB2}}\} &&\\
R_1+2R_2 &\leq& (n_{22}-n_{21})^+ + \max(n_{22}, n_{12}) && \text{(D.2i)}\\
&+& \max\{(n_{11}-n_{12})^+, n_{21}\} &&\\
&+& \min\{(n_{11}-n_{12})^+, n_{21}, C_{\mathsf{FB1}}\}. &&
\end{aligned}
$$

We will show that the inequalities developed above are equivalent to the capacity region in Theorem 5.3. Note that (D.2b) can be written as

$$
R_1 \leq \begin{cases} n_{11} + C_{\mathsf{FB2}}, & n_{11} + C_{\mathsf{FB2}} \leq n_{12}; \\ \max(n_{11}, n_{12}), & \text{otherwise.} \end{cases}
$$

This shows that this inequality is implied by (5.22a) and (5.22b). Similarly, (D.2d) is implied by (5.22c) and (5.22d). Next consider (D.2g),

• if $C_{\mathsf{FB1}} \leq \min\{(n_{11} - n_{12})^+, n_{21}\}$,
$C_{\mathsf{FB2}} \leq \min\{(n_{22} - n_{21})^+, n_{12}\}$:

$$R_1 + R_2 \leq \max\{(n_{11} - n_{12})^+, n_{21}\} + \max\{(n_{22} - n_{21})^+, n_{12}\} + C_{\mathsf{FB1}} + C_{\mathsf{FB2}}, \tag{D.3}$$

• if $C_{\mathsf{FB1}} > \min\{(n_{11} - n_{12})^+, n_{21}\}$,
$C_{\mathsf{FB2}} > \min\{(n_{22} - n_{21})^+, n_{12}\}$:

$$R_1 + R_2 \leq \max(n_{11}, n_{12}) + \max(n_{21}, n_{22}), \tag{D.4}$$

• if $C_{\mathsf{FB1}} \leq \min\{(n_{11} - n_{12})^+, n_{21}\}$,
$C_{\mathsf{FB2}} > \min\{(n_{22} - n_{21})^+, n_{12}\}$:

$$R_1 + R_2 \leq \max\{(n_{11} - n_{12})^+, n_{21}\} + C_{\mathsf{FB1}} + n_{12} + (n_{22} - n_{21})^+, \tag{D.5}$$

• and finally, if $C_{\mathsf{FB1}} > \min\{(n_{11} - n_{12})^+, n_{21}\}$,
$C_{\mathsf{FB2}} \leq \min\{(n_{22} - n_{21})^+, n_{12}\}$:

$$R_1 + R_2 \leq \max\{(n_{22} - n_{21})^+, n_{12}\} + C_{\mathsf{FB2}} + n_{21} + (n_{11} - n_{12})^+. \tag{D.6}$$

Note that the first case is implied by (5.22g); and the second case is implied by (5.22a) and (5.22c). Also notice that the third case is implied by (5.22c) and (5.22i); and the last case is implied by (5.22a) and (5.22h). Lastly, we consider

(D.2h):

$$2R_1 + R_2 \leq \begin{cases} (n_{11} - n_{12})^+ + \max(n_{11}, n_{21}) \\ + \max\{(n_{22} - n_{21})^+, n_{12}\} + C_{\mathsf{FB2}}, \\ \qquad \text{if } C_{\mathsf{FB2}} \leq \min\{(n_{22} - n_{21})^+, n_{12}\}; \\ \max(n_{11}, n_{12}) + \max(n_{11}, n_{21}) \\ +(n_{22} - n_{21})^+, \\ \qquad \text{if } C_{\mathsf{FB2}} > \min\{(n_{22} - n_{21})^+, n_{12}\}. \end{cases}$$

Note that the first case is implied by (5.22h); and the second case is implied by (5.22a) and (5.22f). Similarly, it can be shown that (D.2i) is implied by (5.22i), (5.22c) and (5.22e). Therefore, the inequalities of (D.2a)-(D.2i) are equivalent to those of (5.22a)-(5.22i), thus proving the achievablity of Theorem 5.3.

D.3 Proof of Theorem 5.4

Proof of (5.26a) and (5.26b): Starting with Fano's inequality, we get

$$\begin{aligned} N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N) \\ &\leq \sum [h(Y_{1i}) - h(Y_{1i}|W_1, Y_1^{i-1}, X_{1i})] \\ &= \sum [h(Y_{1i}) - h(Z_{1i})], \end{aligned}$$

where the second inequality follows from the fact that conditioning reduces entropy and X_{1i} is a function of (W_1, Y_1^{i-1}); and the third equality follows from the memoryless property of the channel. If (R_1, R_2) is achievable, then $\epsilon_N \to 0$ as $N \to \infty$. Assume that X_1 and X_2 have covariance ρ, *i.e.*, $\rho = \mathbb{E}[X_1 X_2^*]$. We can then obtain (5.26a).

To obtain (5.26b), consider

$$\begin{aligned}
N(R_1 - \epsilon_N) &\le I(W_1; Y_1^N, Y_2^N | W_2) \\
&= \sum h(Y_{1i}, Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}) - h(Y_1^N, Y_2^N | W_1, W_2) \\
&\overset{(a)}{=} \sum h(Y_{1i}, Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}, X_{2i}) \\
&\quad - \sum [h(Z_{1i}) + h(Z_{2i})] \\
&\overset{(b)}{=} \sum h(Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}, X_{2i}) \\
&\quad + \sum h(Y_{1i} | W_2, Y_2^{i}, X_{2i}, S_{1i}) - \sum [h(Z_{1i}) + h(Z_{2i})] \\
&\overset{(c)}{\le} \sum [h(Y_{2i} | X_{2i}) - h(Z_{2i})] \\
&\quad + \sum [h(Y_{1i} | X_{2i}, S_{1i}) - h(Z_{1i})],
\end{aligned}$$

where (a) follows from the fact that X_2^i is a function of (W_2, Y_2^{i-1}) and $h(Y_1^N, Y_2^N | W_1, W_2) = \sum [h(Z_{1i}) + h(Z_{2i})]$ (see Claim D.2 below); (b) follows from the fact that $S_1^i := h_{12} X_1^i + Z_2^i$ is a function of (Y_2^i, X_2^i); (c) follows from the fact that conditioning reduces entropy. Hence, we get

$$\begin{aligned}
R_1 &\le h(Y_2 | X_2) - h(Z_2) + h(Y_1 | X_2, S_1) - h(Z_1) \\
&\overset{(a)}{\le} \log\left(1 + (1 - |\rho|^2)\mathsf{INR}_{12}\right) \\
&\quad + \log\left(1 + \frac{(1 - |\rho|^2)\mathsf{SNR}_1}{1 + (1 - |\rho|^2)\mathsf{INR}_{12}}\right),
\end{aligned}$$

where (a) follows from the fact that

$$h(Y_2 | X_2) \le \log 2\pi e \left(1 + (1 - |\rho|^2)\mathsf{INR}_{12}\right), \tag{D.7}$$

$$h(Y_1 | X_2, S_1) \le \log 2\pi e \left(1 + \frac{(1 - |\rho|^2)\mathsf{SNR}_1}{1 + (1 - |\rho|^2)\mathsf{INR}_{12}}\right). \tag{D.8}$$

The inequality of (D.8) is obtained as follows. Given (X_2, S_1), the variance of Y_1 is upper-bounded by

$$\mathrm{Var}\,[Y_1 | X_2, S_1] \le K_{Y_1} - K_{Y_1 (X_2, S_1)} K_{(X_2, S_1)}^{-1} K_{Y_1 (X_2, S_1)}^{*},$$

where

$$
\begin{aligned}
K_{Y_1} &= E\left[|Y_1|^2\right] \\
&= 1 + \mathsf{SNR}_1 + \mathsf{INR}_{21} + \rho h_{11}^* h_{21} + \rho^* h_{11} h_{21}^*, \\
K_{Y_1(X_2,S_1)} &= E\left[Y_1[X_2^*, S_1^*]\right] \\
&= \left[\rho h_{11} + h_{21}, h_{12}^* h_{11} + \rho^* h_{21} h_{12}^*\right], \\
K_{(X_2,S_1)} &= E\left[\begin{bmatrix} |X_2|^2 & X_2 S_1^* \\ X_2^* S_1 & |S_1|^2 \end{bmatrix}\right] \\
&= \begin{bmatrix} 1 & \rho^* h_{12}^* \\ \rho h_{12} & 1 + \mathsf{INR}_{12} \end{bmatrix}.
\end{aligned}
\tag{D.9}
$$

By further calculation, we can get (D.8).

Claim D.2 $h(Y_1^N, Y_2^N | W_1, W_2) = h(Y_1^N, S_1^N | W_1, W_2) = \sum[h(Z_{1i}) + h(Z_{2i})]$.

Proof:

$$
\begin{aligned}
h(Y_1^N, Y_2^N | W_1, W_2) &= \sum h(Y_{1i}, Y_{2i} | W_1, W_2, Y_1^{i-1}, Y_2^{i-1}) \\
&\overset{(a)}{=} \sum h(Y_{1i}, Y_{2i} | W_1, W_2, Y_1^{i-1}, Y_2^{i-1}, X_{1i}, X_{2i}) \\
&\overset{(b)}{=} \sum h(Z_{1i}, Z_{2i} | W_1, W_2, Y_1^{i-1}, Y_2^{i-1}, X_{1i}, X_{2i}) \\
&\overset{(c)}{=} \sum \left[h(Z_{1i}) + h(Z_{2i})\right],
\end{aligned}
$$

where (a) follows from the fact that X_{1i} is a function of (W_1, Y_1^{i-1}) and X_{2i} is a function of (W_2, Y_2^{i-1}); (b) follows from the fact that $Y_{1i} = h_{11}X_{1i} + h_{21}X_{2i} + Z_{1i}$ and $S_{1i} := h_{12}X_{1i} + Z_{2i}$; (c) follows from the memoryless property of the channel and the independence assumption of Z_{1i} and Z_{2i}. Similarly, one can show that $h(Y_1^N, S_1^N | W_1, W_2) = \sum[h(Z_{1i}) + h(Z_{2i})]$. ■

Proof of (5.26c): Starting with Fano's inequality, we get

$$
\begin{aligned}
N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N, \tilde{Y}_2^N | W_2) \\
&= \sum h(Y_{1i}, \tilde{Y}_{2i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}) - h(Y_1^N, \tilde{Y}_2^N | W_1, W_2) \\
&\overset{(a)}{\leq} \sum h(Y_{1i}, \tilde{Y}_{2i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_{2i}) - \sum h(Z_{1i}) \\
&= \sum H(\tilde{Y}_{2i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_{2i}) \\
&\quad + \sum h(Y_{1i} | W_2, Y_1^{i-1}, \tilde{Y}_2^{i}, X_{2i}) - \sum h(Z_{1i}) \\
&\overset{(b)}{\leq} NC_{\mathsf{FB2}} + \sum [h(Y_{1i} | X_{2i}) - h(Z_{1i})],
\end{aligned}
$$

where (a) follows from the fact that $h(Y_1^N, \tilde{Y}_2^N | W_1, W_2) \geq \sum h(Z_{1i})$ (see Claim D.3 below) and X_{2i} is a function of $(W_2, \tilde{Y}_2^{i-1})$; (c) follows from the fact that $H(\tilde{Y}_{2i}) \leq C_{\mathsf{FB2}}$ and conditioning reduces entropy. So we get

$$
\begin{aligned}
R_1 &\leq h(Y_1 | X_2) - h(Z_1) + C_{\mathsf{FB2}} \\
&\leq \log\left(1 + (1 - |\rho|^2)\mathsf{SNR}_1\right) + C_{\mathsf{FB2}}.
\end{aligned}
$$

Claim D.3 $h(Y_1^N, \tilde{Y}_2^N | W_1, W_2) \geq \sum h(Z_{1i})$.

Proof:

$$
\begin{aligned}
h(Y_1^N, \tilde{Y}_2^N | W_1, W_2) &= H(\tilde{Y}_2^N | W_1, W_2) \\
&\quad + h(Y_1^N | W_1, W_2, \tilde{Y}_2^N) \\
&\overset{(a)}{\geq} \sum h(Y_{1i} | W_1, W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_{1i}, X_{2i}) \\
&= \sum h(Z_{1i} | W_1, W_2, Y_1^{i-1}, \tilde{Y}_2^{i-1}, X_{1i}, X_{2i}) \\
&\overset{(b)}{=} \sum h(Z_{1i}),
\end{aligned}
$$

where (a) follows from the fact that entropy is nonnegative and X_{1i} is a function of (W_1, Y_1^{i-1}) and X_{2i} is a function of $(W_2, \tilde{Y}_2^{i-1})$; (b) follows from the memoryless property of the channel. ■

Proof of (5.26h):

$$
\begin{aligned}
&N(R_1 + R_2 - \epsilon_N) \le I(W_1; Y_1^N) + I(W_2; Y_2^N) \\
&\overset{(a)}{\le} I(W_1; Y_1^N, S_1^N, W_2) + I(W_2; Y_2^N) \\
&\overset{(b)}{=} h(Y_1^N, S_1^N|W_2) - h(Y_1^N, S_1^N|W_1, W_2) + I(W_2; Y_2^N) \\
&\overset{(c)}{=} h(Y_1^N, S_1^N|W_2) - \sum [h(Z_{1i}) + h(Z_{2i})] + I(W_2; Y_2^N) \\
&\overset{(d)}{=} h(Y_1^N|S_1^N, W_2) - \sum h(Z_{1i}) + h(Y_2^N) - \sum h(Z_{2i}) \\
&\overset{(e)}{=} h(Y_1^N|S_1^N, W_2, X_2^N) - \sum h(Z_{1i}) + h(Y_2^N) - \sum h(Z_{2i}) \\
&\overset{(f)}{\le} \sum_{i=1}^{N} [h(Y_{1i}|S_{1i}, X_{2i}) - h(Z_{1i}) + h(Y_{2i}) - h(Z_{2i})],
\end{aligned}
$$

where (a) follows from the fact that adding information increases mutual information (providing a *genie*); (b) follows from the independence of W_1 and W_2; (c) follows from $h(Y_1^N, S_1^N|W_1, W_2) = \sum [h(Z_{1i}) + h(Z_{2i})]$ (see Claim D.2); (d) follows from $h(S_1^N|W_2) = h(Y_2^N|W_2)$ (see Claim 5.3); (e) follows from the fact that X_2^N is a function of (W_2, S_1^{N-1}) (see Claim 5.4); (f) follows from the fact that conditioning reduces entropy. Hence, we get

$$R_1 + R_2 \le h(Y_1|S_1, X_2) - h(Z_1) + h(Y_2) - h(Z_2).$$

Note that

$$h(Y_2) \le \log 2\pi e \left(1 + \mathsf{SNR}_2 + \mathsf{INR}_{12} + 2|\rho| \sqrt{\mathsf{SNR}_2 \cdot \mathsf{INR}_{12}}\right). \tag{D.10}$$

From (D.8) and (D.10), we get the desired upper bound.

Proof of (5.26j):

$$
\begin{aligned}
&N(2R_1 + R_2 - \epsilon_N) \le I(W_1; Y_1^N) \\
&\quad + I(W_1; Y_1^N|W_2) + I(W_2; Y_2^N) \\
&\quad \overset{(a)}{\le} [h(Y_1^N) - h(Y_1^N|W_1)] + I(W_1; Y_1^N, S_1^N|W_2)
\end{aligned}
$$

$$+ [h(Y_2^N) - h(Y_2^N|W_2)]$$

$$\overset{(b)}{=} [h(Y_1^N) - h(Y_1^N|W_1)] + h(Y_1^N, S_1^N|W_2)$$

$$- \sum [h(Z_{1i}) + h(Z_{2i})] + [h(Y_2^N) - h(Y_2^N|W_2)]$$

$$\overset{(c)}{=} [h(Y_1^N) - h(S_2^N|W_1)] + h(Y_2^N) + h(Y_1^N|W_2, S_1^N)$$

$$- \sum [h(Z_{1i}) + h(Z_{2i})]$$

$$= h(Y_1^N) - \sum [h(Z_{1i}) + h(Z_{2i})] + h(Y_1^N|W_2, S_1^N)$$

$$+ I(W_1; S_2^N) - h(S_2^N) + h(Y_2^N)$$

$$= h(Y_1^N) - \sum [h(Z_{1i}) + h(Z_{2i})] + h(Y_1^N|W_2, S_1^N)$$

$$+ I(W_1; S_2^N) + I(S_2^N; Y_2^N) + h(Y_2^N|S_2^N)$$

$$- I(S_2^N; Y_2^N) - h(S_2^N|Y_2^N)$$

$$\overset{(d)}{\leq} \underbrace{h(Y_1^N) - h(Z_1^N) - h(Z_2^N) + h(Y_1^N|W_2, S_1^N) + h(Y_2^N|S_2^N)}_{T}$$

$$+ I\left(S_2^N, \tilde{Y}_2^N, W_2; W_1\right) - h\left(S_2^N|Y_2^N, W_2, \tilde{Y}_2^N\right)$$

$$= T + I\left(\tilde{Y}_2^N; W_1|W_2\right) + I\left(S_2^N; W_1|W_2, \tilde{Y}_2^N\right)$$

$$- h\left(S_2^N|Y_2^N, W_2, \tilde{Y}_2^N\right)$$

$$= T + I\left(\tilde{Y}_2^N; W_1|W_2\right) + I\left(Z_1^N; W_1|W_2, \tilde{Y}_2^N\right)$$

$$- h\left(Z_1^N|S_1^N, W_2, \tilde{Y}_2^N\right)$$

$$= \underbrace{T - h(Z_1^N)}_{T'} + I\left(\tilde{Y}_2^N; W_1|W_2\right) + h(Z_1^N)$$

$$- h\left(Z_1^N|W_1, W_2, \tilde{Y}_2^N\right) + I(Z_1^N; S_1^N|W_2, \tilde{Y}_2^N)$$

$$\overset{(e)}{=} T' + I(Z_1^N; S_1^N|W_2, \tilde{Y}_2^N) + I\left(\tilde{Y}_2^N; W_1, Z_1^N|W_2\right)$$

$$= T' + I(Z_1^N; S_1^N, \tilde{Y}_2^N|W_2) + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right)$$

$$\overset{(f)}{\leq} T' + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right)$$

$$+ I(Z_1^N; \tilde{Y}_2^N, W_1, \tilde{Y}_1^N, Z_2^N|W_2)$$

$$
\begin{aligned}
&\overset{(g)}{=} T' + I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I(Z_1^N; \tilde{Y}_1^N|W_1, W_2, Z_2^N) \\
&\overset{(h)}{=} h(Y_1^N) - h(Z_1^N) + h(Y_2^N|S_2^N) - h(Z_2^N) \\
&+ h(Y_1^N|W_2, S_1^N, X_2^N) - h(Z_1^N) \\
&+ I\left(\tilde{Y}_2^N; W_1|W_2, Z_1^N\right) \\
&+ I(\tilde{Y}_1^N; Z_1^N|W_1, W_2, Z_2^N) \\
&\overset{(i)}{\leq} NC_{\mathsf{FB1}} + NC_{\mathsf{FB2}} + \sum[h(Y_{1i}) - h(Z_{1i})] \\
&+ \sum[h(Y_{1i}|S_{1i}, X_{2i}) - h(Z_{1i})] \\
&+ \sum[h(Y_{2i}|S_{2i}) - h(Z_{2i})],
\end{aligned} \tag{D.11}
$$

where (a) follows from the non-negativity of mutual information; (b) follows from $h(Y_1^N, S_1^N|W_1, W_2) = \sum[h(Z_{1i}) + h(Z_{2i})]$ (by Claim D.2); (c) follows from Claim 5.3; (d) follows from the non-negativity of mutual information and the fact that conditioning reduces entropy; (e) is true since Z_k^N, W_1, and W_2 are mutually independent; (f) follows from the fact that S_1^N is a function of $(W_1, \tilde{Y}_1^N, Z_2^N)$; ($g$) can be obtained by taking similar steps as in (5.28); (h) follows from Claim 5.4; (i) follows from the fact that $H(\tilde{Y}_k^N) \leq NC_{\mathsf{FBk}}$ and conditioning reduces entropy.

Also note that

$$
h(Y_1|X_2, S_1) \leq \log 2\pi e\left(1 + \frac{(1-|\rho|^2)\mathsf{SNR}_1}{1 + (1-|\rho|^2)\mathsf{INR}_{12}}\right). \tag{D.12}
$$

Therefore, we get the desired bound.

D.4 Gap Analysis of Theorem 5.6

We show that our proposed achievability strategy in Section 5.6.2 results in a sum-rate to within at most 14.8 bits/sec/Hz of the outerbounds in Corollary 5.5. It is sufficient to prove this for the extreme case of feedback capacity assignment, *i.e.*, where $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$ (or symmetrically $C_{\mathsf{FB1}} = 0$ and $C_{\mathsf{FB2}} = C_{\mathsf{FB}}$). The reason is as follows. Consider our achievability strategy for the general feedback strategy described previously, and let $C_{\mathsf{FB1}} = \lambda C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = (1-\lambda)C_{\mathsf{FB}}$, such that $0 \leq \lambda \leq 1$. Under these assumptions, for any value of λ, the outerbounds on sum-rate in (5.33a),(5.33b), and (5.33c) would be the same, call the minimum of them C^*. Assuming that we can achieve to within 14.8 bits/sec/Hz of this outer-bound in the extreme cases, then, with the described achievability scheme for general feedback assignment, we can achieve

$$\begin{aligned} R_{\mathsf{SUM}} &= \lambda R_{\mathsf{SUM}}^{C_{\mathsf{FB2}}=0} + (1-\lambda)\, R_{\mathsf{SUM}}^{C_{\mathsf{FB1}}=0} \\ &\geq \lambda\,(C^* - 14.8) + (1-\lambda)\,(C^* - 14.8) = (C^* - 14.8)\,. \end{aligned} \tag{D.13}$$

We now prove our claim for the extreme cases. By symmetry, we only need to analyze the gap in one case, say $C_{\mathsf{FB1}} = C_{\mathsf{FB}}$ and $C_{\mathsf{FB2}} = 0$. We assume that $\mathsf{INR} \geq 1$, since for the case when $\mathsf{INR} < 1$, by ignoring the feedback and treating interference as noise, we can achieve a sum-rate of

$$2\log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right), \tag{D.14}$$

which is at most within 2.6 bits of outerbound (5.33b) in Corollary 5.5:

$$\begin{aligned} &\log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) + \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right) \\ &- 2\log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) \end{aligned}$$

$$\overset{(\mathsf{INR}\leq 1)}{\leq} \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right)$$
$$- \log\left(1 + \frac{\mathsf{SNR}}{2}\right)$$
$$\overset{(\mathsf{INR}\leq 1)}{\leq} \log(2 + 3\mathsf{SNR}) - \log(1 + \mathsf{SNR}) + 1$$
$$= \log\left(\frac{2 + 3\mathsf{SNR}}{1 + \mathsf{SNR}}\right) + 1$$
$$\leq \log(3) + 1 \leq 2.6. \tag{D.15}$$

We consider five different subcases.

Case (a) $\log(\mathsf{INR}) \leq \frac{1}{2}\log(\mathsf{SNR})$**:**

For this case, we pick the following set of power levels[1]:

$$\begin{cases} P_1^{(1)} = \left(\frac{1}{\mathsf{INR}} - \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}\right)^+ \\ P_1^{(2)} = \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} \\ P_1^{(3)} = \frac{1}{\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} \\ P_2^{(1)} = \frac{1}{\mathsf{INR}} \\ P_2^{(2)} = \frac{1}{2\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} \end{cases} \tag{D.16}$$

Note that the power levels are non-negative, and at transmitter 1, we have

$$P_1 = P_1^{(1)} + P_1^{(2)} + P_1^{(3)}$$
$$= \left(\frac{1}{\mathsf{INR}} - \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}\right)^+$$
$$+ \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} + \frac{1}{\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}$$
$$\overset{(a)}{=} \frac{1}{\mathsf{INR}} - \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}$$
$$+ \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} + \frac{1}{\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}$$
$$= \frac{1}{\mathsf{INR}} + \frac{1}{\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}$$

[1]Remember that starting the beginning of Section 5.6.2, we have assumed that $\mathsf{INR} \geq 1$. Hence, we are not encountering division by zero in power assignments of (D.16).

$$\leq \frac{1}{\mathsf{INR}} + \frac{\mathsf{INR}-1}{\mathsf{INR}} \leq 1, \tag{D.17}$$

where (a) follows from the fact that

$$\frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} \leq \frac{\mathsf{INR}-1}{\mathsf{SNR}} \overset{(\mathsf{INR}^2 \leq \mathsf{SNR})}{\leq} \frac{1}{\mathsf{INR}}. \tag{D.18}$$

At transmitter 2,

$$\begin{aligned}
P_2 &= P_2^{(1)} + P_2^{(2)} \\
&= \frac{1}{\mathsf{INR}} + \frac{1}{2\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\} \\
&\leq \frac{1}{\mathsf{INR}} + \frac{\mathsf{INR}-1}{\mathsf{INR}} \leq 1.
\end{aligned} \tag{D.19}$$

By plugging the given values of power levels into our achievable sum-rate $R_{\mathsf{SUM}}^{(a)}$ defined in (5.48), we get

$$\begin{aligned}
R_{\mathsf{SUM}}^{(a)} = \log&\left(1 + \frac{\mathsf{SNR}P_1^{(1)}}{1 + \mathsf{INR}P_2^{(1:2)} + \mathsf{SNR}P_1^{(2)}}\right) \\
&+ \log\left(\frac{\mathsf{SNR}P_1^{(2)}}{2}\right)^+ + \log\left(1 + \frac{\mathsf{SNR}P_2^{(1)}}{2}\right) \\
&+ \log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)^+ \\
&= \log\left(\frac{2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(1:2)}}{2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(2)}}\right) \\
&+ \log\left(\frac{\mathsf{SNR}P_1^{(2)}}{2}\right)^+ + \log\left(1 + \frac{\mathsf{SNR}}{2\mathsf{INR}}\right) \\
&+ \log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)^+ \\
&\overset{(a)}{\geq} \log\left(\frac{2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(1:2)}}{2\left(1 + \mathsf{SNR}P_1^{(2)}\right)} \times \frac{1 + \mathsf{SNR}P_1^{(2)}}{4}\right) \\
&+ \log\left(1 + \frac{\mathsf{SNR}}{2\mathsf{INR}}\right) + \log\left(\frac{\mathsf{INR}P_2^{(2)}}{1 + \mathsf{INR}P_2^{(1)}}\right)^+ \\
&= \log\left(2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(1:2)}\right) + \log\left(1 + \frac{\mathsf{SNR}}{2\mathsf{INR}}\right)
\end{aligned}$$

$$+ \log\left(\frac{\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}-1\}}{4}\right)^+ - 3$$

$$\begin{aligned}
&\geq \log\left(2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(1:2)}\right) + \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
&+ \min\{C_{\mathsf{FB}}, \log\left(\mathsf{INR}-1\right)^+\} - 6 \\
&\geq \log\left(2 + \mathsf{INR}P_2^{(2)} + \mathsf{SNR}P_1^{(1:2)}\right) + \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
&+ \min\{C_{\mathsf{FB}}, \log\left(1+\mathsf{INR}\right)\} - \log(3) - 6, \qquad \text{(D.20)}
\end{aligned}$$

where (a) follows from the assumption $\mathsf{SNR} \geq \mathsf{INR} \geq 1$, and the last inequality holds since

$$\log\left(\mathsf{INR}-1\right)^+ \geq \log\left(1+\mathsf{INR}\right) - 3 \ \ \forall\, \mathsf{INR} \geq 1. \qquad \text{(D.21)}$$

• If $C_{\mathsf{FB}} \leq \log\left(1+\mathsf{INR}\right)$: Considering the outerbound in (5.33c), in this case we can write

$$\begin{aligned}
R_1 + R_2 &\leq 2\log\left(1 + \mathsf{INR} + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) + C_{\mathsf{FB}} \\
&\leq 2\log\left(1 + \mathsf{INR} + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) + C_{\mathsf{FB}} \\
&= 2\log\left(1 + \frac{\mathsf{INR}^2 + \mathsf{SNR}}{\mathsf{INR}}\right) + C_{\mathsf{FB}} \\
&\overset{(\mathsf{INR}^2 \leq \mathsf{SNR})}{\leq} 2\log\left(1 + \frac{2\mathsf{SNR}}{\mathsf{INR}}\right) + C_{\mathsf{FB}} \\
&\leq 2\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) + C_{\mathsf{FB}} + 2. \qquad \text{(D.22)}
\end{aligned}$$

The gap between the achievable sum-rate of (D.20) and the outerbound in

(D.22), is upper bounded by

$$
\begin{aligned}
& 8+\log(3)+2\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)+C_{\mathsf{FB}} \\
& -\log\left(2+\mathsf{INR}P_2^{(2)}+\mathsf{SNR}P_1^{(1:2)}\right) \\
& -\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)-C_{\mathsf{FB}} \\
& \leq 8+\log(3)+\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)-\log\left(2+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& \leq 8+\log(3).
\end{aligned}
$$

• If $C_{\mathsf{FB}} > \log(1+\mathsf{INR})$: Considering the outerbound in (5.33b), in this case we can write

$$
\begin{aligned}
R_1+R_2 \leq\ & \log\left(1+\frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) \\
& +\log\left(1+\mathsf{SNR}+\mathsf{INR}+2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right) \\
\leq\ & \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& +\log\left(1+\mathsf{SNR}+\mathsf{INR}+2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right).
\end{aligned}
\tag{D.23}
$$

The gap between the achievable sum-rate of (D.20) and the outerbound in

(D.23), is upper bounded by

$$
\begin{aligned}
& 6+\log(3)+\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& +\log\left(1+\mathsf{SNR}+\mathsf{INR}+2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right) \\
& -\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)-\log\left(2+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& -\log(1+\mathsf{INR}) \\
& \leq 6+\log(3)-\log\left(2+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& +\log\left(1+\frac{\mathsf{SNR}+2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}}{\mathsf{INR}}\right) \\
& \leq 6+\log(3)-\log\left(2+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) \\
& +\log\left(1+\frac{3\mathsf{SNR}}{\mathsf{INR}}\right) \\
& \leq 6+2\log(3).
\end{aligned}
$$

Hence, we conclude that the gap between the inner-bound and the outer-bound is at most $8+\log(3) \leq 9.6$ bits/sec/Hz.

Case (b) $\frac{1}{2}\log(\mathsf{SNR}) \leq \log(\mathsf{INR}) \leq \frac{2}{3}\log(\mathsf{SNR})$**:**

For this case, we pick the following set of power levels[2]:

$$
\left\{
\begin{array}{lcl}
P_1^{(1)} & = & \left(\frac{1}{\mathsf{INR}}-\frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}},\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\}\right)^+ \\
P_1^{(2)} & = & \frac{1}{\mathsf{SNR}}\min\{2^{C_{\mathsf{FB}}},\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\} \\
P_1^{(3)} & = & \frac{1}{\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}},\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\} \\
P_1^{(4)} & = & \left(1-P_1^{(1:3)}\right)^+ \\
P_2^{(1)} & = & \frac{1}{\mathsf{INR}} \\
P_2^{(2)} & = & \frac{1}{2\mathsf{INR}}\min\{2^{C_{\mathsf{FB}}},\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\} \\
P_2^{(3)} & = & \left(1-P_2^{(1:2)}\right)^+
\end{array}
\right. \qquad \text{(D.24)}
$$

[2] $\mathsf{INR} \geq 1$.

All the power levels are non-negative. Also, we have

$$
\begin{aligned}
&P_1^{(1)} + P_1^{(2)} + P_1^{(3)} \\
&= \left(\frac{1}{\mathsf{INR}} - \frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\}\right)^+ \\
&+ \frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} + \frac{1}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} \\
&\overset{(a)}{=} \frac{1}{\mathsf{INR}} - \frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} \\
&+ \frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} + \frac{1}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} \\
&= \frac{1}{\mathsf{INR}} + \frac{1}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} \\
&\leq \frac{\mathsf{SNR}^2}{\mathsf{INR}^4} \overset{(\sqrt{\mathsf{SNR}} \leq \mathsf{INR})}{\leq} \frac{\sqrt{\mathsf{SNR}}}{\mathsf{INR}} \leq 1,
\end{aligned}
\tag{D.25}
$$

where (a) follows from the fact that

$$
\frac{1}{\mathsf{SNR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\} \leq \frac{1}{\sqrt{\mathsf{SNR}}} \leq \frac{1}{\mathsf{INR}}. \tag{D.26}
$$

Since $P_1^{(4)} = \left(1 - P_1^{(1:3)}\right)^+$, we conclude that $P_1 \leq 1$. Similarly, we can show that $P_2 \leq 1$.

By plugging the given values of power levels into our achievable sum-rate $R_{\mathsf{SUM}}^{(b)}$ defined in (5.57), we have

$$
\begin{aligned}
R_{\mathsf{SUM}}^{(b)} &= \log\left(\frac{1 + \mathsf{INR}P_2^{(1)} + \mathsf{SNR}P_1^{(1:2)}}{1 + \mathsf{INR}P_2^{(1)}}\right) \\
&+ \log\left(\frac{2\mathsf{INR} + \mathsf{SNR} + \mathsf{INR}^2 - 2\mathsf{INR}}{2\mathsf{INR} + \mathsf{SNR}}\right) \\
&+ \log\left(1 + \frac{\mathsf{SNR}}{2\mathsf{INR}}\right) \\
&+ \log\left(\frac{2\mathsf{INR} + \mathsf{SNR} + \mathsf{INR}^2 - 3/2\mathsf{INR}}{2\mathsf{INR} + \mathsf{SNR}}\right) \\
&+ \min\{C_{\mathsf{FB}}, \log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\right)^+\} - 3
\end{aligned}
$$

$$\geq \log\left(\frac{2+\mathsf{SNR}P_1^{(1:2)}}{2}\right)$$

$$+\log\left(1+\frac{\mathsf{SNR}}{2\mathsf{INR}}\right)+2\log\left(\frac{\mathsf{INR}^2}{3\mathsf{SNR}}\right)$$

$$+\min\{C_{\mathsf{FB}},\log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\right)^+\}-3$$

$$\geq 2\log\left(1+\frac{\mathsf{SNR}}{2\mathsf{INR}}\right)+2\log\left(\frac{\mathsf{INR}^2}{3\mathsf{SNR}}\right)$$

$$+\min\{C_{\mathsf{FB}},\log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\right)^+\}-3$$

$$\geq 2\log\left(\frac{1+\mathsf{SNR}}{2\mathsf{INR}}\right)+2\log\left(\mathsf{INR}^2\right)-2\log\left(3\mathsf{SNR}\right)$$

$$+\min\{C_{\mathsf{FB}},\log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\right)^+\}-3$$

$$=2\log\left(1+\mathsf{SNR}\right)-2\log\left(\mathsf{INR}\right)$$

$$+4\log\left(\mathsf{INR}\right)-2\log\left(3\mathsf{SNR}\right)$$

$$+\min\{C_{\mathsf{FB}},\log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\right)^+\}-3$$

$$\geq 2\log\left(1+\mathsf{INR}\right)$$

$$+\min\{C_{\mathsf{FB}},\log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3}-1\right)^+\}-5-2\log\left(3\right). \tag{D.27}$$

We first simplify the outerbounds in (5.33b) and (5.33c). Considering the outerbound in (5.33c), for this case we can write

$$R_1+R_2 \leq 2\log\left(1+\mathsf{INR}+\frac{\mathsf{SNR}}{1+\mathsf{INR}}\right)+C_{\mathsf{FB}}$$

$$\leq 2\log\left(1+\mathsf{INR}+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)+C_{\mathsf{FB}}$$

$$\overset{(\mathsf{SNR}\leq\mathsf{INR}^2)}{\leq} 2\log\left(1+\mathsf{INR}+\mathsf{INR}\right)+C_{\mathsf{FB}}$$

$$=2\log\left(1+2\mathsf{INR}\right)+C_{\mathsf{FB}}$$

$$\leq 2\log\left(1+\mathsf{INR}\right)+C_{\mathsf{FB}}+2, \tag{D.28}$$

and for the outerbound in (5.33b), we have

$$\begin{aligned}
R_1 + R_2 &\leq \log\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) \\
&+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right) \\
&\overset{(\mathsf{INR}\leq\mathsf{SNR})}{\leq} \log\left(1 + \mathsf{INR} + \mathsf{SNR}\right) - \log\left(1 + \mathsf{INR}\right) \\
&+ \log\left(1 + 4\mathsf{SNR}\right) \\
&\overset{(\mathsf{INR}\leq\mathsf{SNR})}{\leq} 2\log\left(1 + \mathsf{SNR}\right) - \log\left(1 + \mathsf{INR}\right) + 3. \qquad \text{(D.29)}
\end{aligned}$$

We consider two possible scenarios based on C_{FB},

(1) $C_{\mathsf{FB}} \leq \log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\right)^+$:

With this assumption, we have

$$R_{\mathsf{SUM}}^{(b)} \geq 2\log\left(1 + \mathsf{INR}\right) + C_{\mathsf{FB}} - 5 - 2\log\left(3\right), \qquad \text{(D.30)}$$

note that, in this case

$$\begin{aligned}
C_{\mathsf{FB}} &\leq \log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\right)^+ \leq \log\left(1 + \frac{\mathsf{SNR}^2}{\mathsf{INR}^3}\right) \\
&\overset{(\sqrt{\mathsf{SNR}}\leq\mathsf{INR})}{\leq} \log\left(1 + \mathsf{INR}\right), \qquad \text{(D.31)}
\end{aligned}$$

therefore, from (D.28) we get

$$R_1 + R_2 \leq 2\log\left(1 + \mathsf{INR}\right) + C_{\mathsf{FB}} + 2. \qquad \text{(D.32)}$$

Hence, the gap between the achievable sum-rate of (D.30) and the outer-bound in (D.32) is at most $7 + 2\log(3) \leq 10.2$ bits/sec/Hz.

(2) $C_{\mathsf{FB}} \geq \log\left(\frac{\mathsf{SNR}^2}{\mathsf{INR}^3} - 1\right)^+$:

We have

$$R_{\mathsf{SUM}}^{(b)} \geq 2\log\left(1 + \mathsf{INR}\right) + \log\left(1 + \frac{\mathsf{SNR}^2}{\mathsf{INR}^3}\right)^+ - 5 - 3\log\left(3\right)$$

$$\geq 2\log(1+\mathsf{INR}) + 2\log(\mathsf{SNR})$$
$$-3\log(\mathsf{INR}) - 5 - 3\log(3) \tag{D.33}$$
$$\geq 2\log(1+\mathsf{INR}) + 2\log(1+\mathsf{SNR})$$
$$-3\log(1+\mathsf{INR}) - 7 - \log(3)$$
$$= 2\log(1+\mathsf{SNR}) - \log(1+\mathsf{INR}) - 7 - 3\log(3).$$

Hence, the gap between the achievable sum-rate of (D.33) and the outer-bound in (D.29) is at most $10 + 3\log(3) \leq 14.8$ bits/sec/Hz. As a result, the gap between the inner-bound and the minimum of the outer-bounds in (5.33b) and (5.33c) is at most 14.8 bits/sec/Hz.

Case (c) $2\log(\mathsf{SNR}) \leq \log(\mathsf{INR})$**:**

• If $\mathsf{SNR} \leq 1$, pick
$$P_2^{(2)} = P_1^{(3)} = \frac{1}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}\},$$
and set all other power levels equal to zero. By plugging the given values of power levels into our achievable sum-rate $R_{\mathsf{SUM}}^{(c)}$ defined in (5.61), we get
$$R_{\mathsf{SUM}}^{(c)} = \log\left(\frac{\min\{2^{C_{\mathsf{FB}}}, \mathsf{INR}\}}{2}\right)$$
$$= \min\{C_{\mathsf{FB}}, \log(\mathsf{INR})\} - 1$$
$$\geq \min\{C_{\mathsf{FB}}, \log(1+\mathsf{INR})\} - 2. \tag{D.34}$$

Consider the outerbounds in (5.33a) and (5.33b), under the assumptions of case (c) and $\mathsf{SNR} \leq 1$, we have

$$R_1 + R_2 \leq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}},$$
$$\log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right)$$

$$+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right)\}$$
$$\leq \min\{2\log(2) + C_{\mathsf{FB}}, \log(2) + \log(2 + 3\mathsf{INR})\}$$
$$\leq \min\{2 + C_{\mathsf{FB}}, 1 + \log(3) + \log(1 + \mathsf{INR})\}$$
$$\leq \min\{C_{\mathsf{FB}}, \log(1 + \mathsf{INR})\} + 2.6. \tag{D.35}$$

Therefore, with the given choice of power levels the achievable sum-rate of (D.34) is within 2.6 bits/sec/Hz of the minimum of the outerbounds in (5.33a) and (5.33b).

• If $\mathsf{SNR} \geq 1$, pick

Pick the following set of power levels:

$$\begin{cases} P_1^{(1)} = 0 \\ P_1^{(2)} = 0 \\ P_1^{(3)} = \frac{\mathsf{SNR}}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{INR}}{\mathsf{SNR}^2}\} \\ P_1^{(4)} = 1 - P_1^{(3)} \\ P_2^{(1)} = 0 \\ P_2^{(2)} = \frac{1}{\mathsf{INR}} \min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{INR}}{\mathsf{SNR}^2}\} \\ P_2^{(3)} = 1 - P_2^{(2)} \end{cases} \tag{D.36}$$

It is straight forward to check that the power levels are non-negative and they satisfy the power constraint at the transmitters. By plugging the given values of power levels into our achievable sum-rate $R_{\mathsf{SUM}}^{(c)}$ defined in (5.61), we get

$$R_{\mathsf{SUM}}^{(c)} = \log\left(1 + \frac{\mathsf{SNR}(1 - \frac{1}{\mathsf{SNR}})}{2}\right)$$
$$+ \log\left(1 + \frac{\mathsf{SNR}(1 - \frac{1}{\mathsf{SNR}})}{2}\right) + \log\left(\frac{\min\{2^{C_{\mathsf{FB}}}, \frac{\mathsf{INR}}{\mathsf{SNR}^2}\}}{2}\right)$$

$$\geq 2\log(1+\mathsf{SNR}) + \min\{C_{\mathsf{FB}}, \log\left(\frac{\mathsf{INR}}{\mathsf{SNR}^2}\right)\} - 3$$

$$= \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}},$$

$$2\log(1+\mathsf{SNR}) + \log\left(\frac{\mathsf{INR}}{\mathsf{SNR}^2}\right)\} - 3$$

$$\geq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}}, \log(1+\mathsf{INR})\} - 4. \tag{D.37}$$

Consider the outerbounds in (5.33a) and (5.33b), under the assumptions of case (c), we have

$$R_1 + R_2 \leq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}},$$

$$\log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right)$$

$$+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR}\cdot\mathsf{INR}}\right)\}$$

$$\leq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}},$$

$$\log(2) + \log(1+4\mathsf{INR})\}$$

$$\leq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}}, 3 + \log(1+\mathsf{INR})\}$$

$$\leq \min\{2\log(1+\mathsf{SNR}) + C_{\mathsf{FB}}, \log(1+\mathsf{INR})\} + 3. \tag{D.38}$$

Therefore, with the given choice of power levels the achievable sum-rate of (5.61) is within 7 bits/sec/Hz of the minimum of the outerbounds in (5.33a) and (5.33b).

Case (d) $\frac{2}{3}\log(\mathsf{SNR}) \leq \log(\mathsf{INR}) \leq \log(\mathsf{SNR})$**:**

In this case feedback is not needed. The achievability scheme of [19] for Gaussian IC without feedback, results in a sum-rate to within 1 bit/sec/Hz of

$$\log(1+\mathsf{SNR}) + \log\left(1 + \frac{\mathsf{SNR}}{1+\mathsf{INR}}\right). \tag{D.39}$$

For the outerbound in (5.33b), in this case we have

$$\begin{aligned}
R_1 + R_2 \leq \log&\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) \\
&+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR} \cdot \mathsf{INR}}\right) \\
&\leq \log\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) + \log\left(1 + 4\mathsf{SNR}\right) \\
&\leq \log\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) + \log\left(1 + \mathsf{SNR}\right) + 2.
\end{aligned} \tag{D.40}$$

Therefore, the achievable sum-rate of [19] is within 3 bits/sec/Hz of the outerbound in (5.33b).

Case (e) $\log(\mathsf{SNR}) \leq \log(\mathsf{INR}) \leq 2\log(\mathsf{SNR})$**:**

In this case feedback is not needed. The achievability scheme of [19] for Gaussian IC without feedback, results in a sum-rate to within 1 bit/sec/Hz of

$$\log\left(1 + \mathsf{SNR} + \mathsf{INR}\right). \tag{D.41}$$

For the outerbound in (5.33b), in this case we have

$$\begin{aligned}
R_1 + R_2 \leq \log&\left(1 + \frac{\mathsf{SNR}}{1 + \mathsf{INR}}\right) \\
&+ \log\left(1 + \mathsf{SNR} + \mathsf{INR} + 2\sqrt{\mathsf{SNR} \cdot \mathsf{INR}}\right) \\
&\overset{(\mathsf{SNR} \leq \mathsf{INR})}{\leq} \log(2) + \log\left(1 + 4\mathsf{INR}\right) \\
&\leq \log\left(1 + \mathsf{INR}\right) + 3.
\end{aligned} \tag{D.42}$$

Therefore, the achievable sum-rate of [19] is within 4 bits/sec/Hz of the outerbound in (5.33b).

BIBLIOGRAPHY

[1] M. J. Abdoli, A. Ghasemi, and A. K. Khandani. On the degrees of freedom of K-user SISO interference and X channels with delayed CSIT. *Submitted to IEEE Transactions on Information Theory, Available online at: Arxiv preprint arXiv:1109.4314*, 2011.

[2] M.J. Abdoli, A. Ghasemi, and A.K. Khandani. On the degrees of freedom of k-user siso interference and x channels with delayed csit. *arXiv preprint arXiv:1109.4314*, 2011.

[3] V. Aggarwal, A. S. Avestimehr, and A. Sabharwal. Sum-capacity of interference networks with partial information. *in Proceedings of International Symposium on Information Theory (ISIT)*, 2010.

[4] V. Aggarwal, A. S. Avestimehr, and A. Sabharwal. On achieving local view capacity via maximal independent graph scheduling. *IEEE Transactions on Information Theory*, 57(5):2711–2729, May 2011.

[5] V. Aggarwal, Y. Liu, and A. Sabharwal. Message passing in distributed wireless networks. *in Proceedings of International Symposium on Information Theory (ISIT)*, 2009.

[6] V. Aggarwal, Y. Liu, and A. Sabharwal. Sum-capacity of interference channels with a local view: Impact of distributed decisions. *submitted to IEEE Transactions on Information Theory*, 2009.

[7] A. S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. *IEEE Transactions on Information Theory*, 57(4), April 2011.

[8] A.S. Avestimehr, A. Sezgin, and D.N.C. Tse. Capacity of the two-way relay channel within a constant gap. *European Transactions on Telecommunications*, 21(4):363–374, 2010.

[9] S. Avestimehr, S. Diggavi, and D. Tse. A deterministic approach to wireless relay networks. *Proceedings of Allerton Conference on Communication, Control, and Computing*, 2007.

[10] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Stat.*, 30:1229–1241, 1959.

[11] G. Bresler, A. Parekh, and D. Tse. The approximate capacity of the many-to-one and one-to-many Gaussian interference channels. *IEEE Transactions on Information Theory*, 56(9):4566–4592, 2010.

[12] G. Bresler and D. Tse. The two-user gaussian interference channel: a deterministic view. *European Transactions on Telecommunications*, 19(4):333–354, 2008.

[13] Jinyuan Chen and Petros Elia. Toward the performance vs. feedback tradeoff for the two-user MISO broadcast channel. *arXiv preprint arXiv:1306.1751*, 2013.

[14] T. M. Cover and Abbas El-Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25:572–584, September 1979.

[15] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[16] A. El-Gamal and M. H. Costa. The capacity region of a class of deterministic interference channels. *IEEE Transactions on Information Theory*, 28(2):343–346, March 1982.

[17] Peter Elias. The noisy channel coding theorem for erasure channels. *The American Mathematical Monthly*, pages 853–862, 1974.

[18] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Transactions on Information Theory*, 51(10):3401–3416, 2005.

[19] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory*, 54(12):5534–5562, December 2008.

[20] N. T. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channenl can increase with feedback. *IEEE Transactions on Information Theory*, 21(1):100–102, January 1975.

[21] Abbas El Gamal and Young-Han Kim. Lecture notes on network information theory. *arXiv:1001.3404*, January 2010.

[22] AE Gamal. The feedback capacity of degraded broadcast channels (corresp.). *IEEE Transactions on Information Theory*, 24(3):379–381, 1978.

[23] M. Gastpar and G. Kramer. On noisy feedback for interference channels. *In Proceedings Asilomar Conference on Signals, Systems and Computers*, October 2006.

[24] Allen Gersho. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory*, 25(4):373–380, 1979.

[25] A. Ghasemi, A. S. Motahari, and A. K . Khandani. On the degrees of freedom of X channel with delayed CSIT. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 909–912, Saint-Petersburg, Russia, July 2011.

[26] T. Gou, S. A. Jafar, S. W. Jeon, and S. Y. Chung. Aligned interference neutralization and the degrees of freedom of the $2\times2\times2$ interference channel. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 2751–2755, 2011.

[27] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, 27:49–60, January 1981.

[28] Andries P. Hekstra and Frans M. J. Willems. Dependence balance bounds for single-output two-way channels. *IEEE Transactions on Information Theory*, 35:44–53, January 1989.

[29] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[30] Ibrahim Issa, Silas L Fong, and A Salman Avestimehr. Two-hop interference channels: Impact of linear schemes. *arXiv preprint arXiv:1309.0898*, 2013.

[31] J. Jiang, Y. Xin, and H. K. Garg. Discrete memoryless interference channels with feedback. *CISS 41st Annual Conference*, March 2007.

[32] G. Kramer. Feedback strategies for white Gaussian interference networks. *IEEE Transactions on Information Theory*, 48(6):1423–1438, 2002.

[33] G. Kramer. Correction to feedback strategies for white Gaussian interference networks, and a capacity theorem for gaussian interference channels with feedback. *IEEE Transactions on Information Theory*, 50(6):1373–1374, 2004.

[34] H. A. Loeliger. Averaging bounds for lattices and linear codes. *IEEE Transactions on Information Theory*, November 1997.

[35] M. A. Maddah-Ali and D. N. T. Tse. Completely stale transmitter channel state information is still very useful. In *Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2010.

[36] Mohammad Ali Maddah-Ali and David Tse. Completely stale transmitter channel state information is still very useful. *IEEE Transactions on Information Theory*, 58(7):4418–4431, 2012.

[37] H. Maleki, S.A. Jafar, and S. Shamai. Retrospective interference alignment. *Arxiv preprint arXiv:1009.3593*, 2010.

[38] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath. Ergodic interference alignment. *Arxiv.org*, 2011. avilable at arxiv.org/abs/0901.4379.

[39] L. Ozarow. The capacity of the white gaussian multiple access channel with feedback. *IEEE Transactions on Information Theory*, 30(4):623–629, 1984.

[40] L Ozarow and S Leung-Yan-Cheong. An achievable region and outer bound for the gaussian broadcast channel with feedback (corresp.). *IEEE Transactions on Information Theory*, 30(4):667–671, 1984.

[41] J.M. Phillips. Chernoff-hoeffding inequality and applications. *arXiv preprint arXiv:1209.6396*, 2012.

[42] G. Poltyrev. On coding without restrictions for the awgn channel. *IEEE Transactions on Information Theory*, 40(2):409–417, 1994.

[43] Vinod Prabhakaran and Pramod Viswanath. Interference channels with source cooperation. *submitted to IEEE Transaction on Information Theory (online available at arXiv:0905.3109v1)*, May 2009.

[44] A. Raja, V. M. Prabhakaran, and P. Viswanath. The two user gaussian compound interference channel. *IEEE Transactions on Information Theory*, 55(11):5100–5120, 2009.

[45] A. Sahai, V. Aggarwal, M. Yuksel, and A. Sabharwal. On channel output feedback in deterministic interference channels. *CISS 41st Annual Conference*, October 2009.

[46] A. Sezgin, A.S. Avestimehr, M.A. Khajehnejad, and B. Hassibi. Divide-and-conquer: Approaching the capacity of the two-pair bidirectional gaussian relay network. *IEEE Transactions on Information Theory*, 58(4):2434–2454, 2012.

[47] C. E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, September 1956.

[48] I. Shomorony and A. S. Avestimehr. Degrees of freedom of two-hop wireless networks: "everyone gets the entire cake". In *50th Annual Allerton Conference on Communication, Control, and Computing*, 2012.

[49] I. Shomorony and A. S. Avestimehr. Two-unicast wireless networks: Characterizing the degrees-of-freedom. *IEEE Transactions on Information Theory*, 59(1):353–383, 2013.

[50] C. Suh and D.N.C. Tse. Feedback capacity of the gaussian interference channel to within 2 bits. *IEEE Transactions on Information Theory*, 57(5):2667–2685, 2011.

[51] Changho Suh and David Tse. The feedback capacity region of El Gamal-Costa deterministic interference channel. *Proceedings of Allerton Conference on Communication, Control, and computing*, September 2009.

[52] R. Tandon, S. Mohajer, V. Poor, and S. Shamai. Degrees of freedom region of the MIMO interference channel with output feedback and delayed CSIT. *accepted for publication in IEEE Transactions on Information Theory*, 2012.

[53] R. Tandon and S. Ulukus. Dependence balance based outer bounds for Gaussian networks with cooperation and feedback. *submitted to the IEEE Transactions on Information Theory*, December 2008.

[54] Emre Telatar. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595, 1999.

[55] David NC Tse and Roy D Yates. Fading broadcast channels with state information at the receivers. *IEEE Transactions on Information Theory*, 58(6):3453–3471, 2012.

[56] Daniela Tuninetti. On InterFerence Channel with Generalized Feedback (IFC-GF). *IEEE International Symposium on Information Theory*, June 2007.

[57] A. Vahid and A. S. Avestimehr. The two-user deterministic interference channel with rate-limited feedback. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, July 2010.

[58] A. Vahid, M.A. Maddah-Ali, and A.S. Avestimehr. Approximate capacity of the two-user MISO broadcast channel with delayed CSIT. In *51st Annual Allerton Conference on Communication, Control, and Computing*, 2013.

[59] A. Vahid, C. Suh, and A. S. Avestimehr. Interference channels with rate-limited feedback. *IEEE Transactions on Information Theory*, 58(5):2788–2812, 2012.

[60] Alireza Vahid, Vaneet Aggarwal, A Salman Avestimehr, and Ashutosh Sabharwal. On the capacity of multi-hop wireless networks with partial network knowledge. In *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1030–1037. IEEE, 2010.

[61] Alireza Vahid, Vaneet Aggarwal, A Salman Avestimehr, and Ashutosh Sabharwal. Wireless network coding with local network views: Coded layer scheduling. *arXiv preprint arXiv:1106.5742*, 2011.

[62] Alireza Vahid, Mohammad Ali Maddah-Ali, and A Salman Avestimehr. Interference channel with binary fading: Effect of delayed network state information. In *49th Annual Allerton Conference on Communication, Control, and Computing*, pages 894–901, 2011.

[63] Alireza Vahid, Mohammad Ali Maddah-Ali, and A Salman Avestimehr. Communication through collisions: Opportunistic utilization of past receptions. *to appear in IEEE Infocom 2014. arXiv preprint arXiv:1312.0116*, 2013.

[64] Alireza Vahid, Mohammad Ali Maddah-Ali, and A Salman Avestimehr. Binary Fading Interference Channel with No CSIT. *to be presented in International Symposium on Information Theory (ISIT)*, 2014.

[65] Alireza Vahid, Mohammad Ali Maddah-Ali, and Amir Salman Avestimehr. Binary fading interference channel with delayed feedback. In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1882–1886. IEEE, 2012.

[66] Alireza Vahid, Mohammad Ali Maddah-Ali, and Amir Salman Avestimehr. Capacity region of the two-user MISO broadcast channel with delayed CSIT to within constant number of bits. *available on arXiv*, 2013.

[67] Alireza Vahid, Mohammad Ali Maddah-Ali, and Amir Salman Avestimehr. Capacity results for binary fading interference channels with delayed CSIT. *arXiv preprint arXiv:1301.5309*, 2013.

[68] C. S. Vaze and M. K. Varanasi. The degrees of freedom region of the two-user MIMO broadcast channel with delayed CSI. Dec. 2010. arxiv.org/abs/1101.0306.

[69] Chinmay S Vaze and Mahesh K Varanasi. The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT. *IEEE Transactions on Information Theory*, 58(8):5354–5374, 2012.

[70] C.S. Vaze and M.K. Varanasi. The degrees of freedom region of the MIMO interference channel with Shannon feedback. *arXiv preprint arXiv:1109.5779*, 2011.

[71] C.S. Vaze and M.K. Varanasi. The degrees of freedom region and interference alignment for the MIMO interference channel with delayed CSIT. *IEEE Transactions on Information Theory*, 58(7):4396–4417, 2012.

[72] Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai. The capacity region of the gaussian mimo broadcast channel. In *International Symposium on Information Theory (ISIT)*, page 174. IEEE, 2004.

[73] Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai. The capacity region of the gaussian multiple-input multiple-output broadcast channel. *IEEE Transactions on Information Theory*, 52(9):3936–3964, 2006.

[74] Frans M. J. Willems. The feedback capacity region of a class of discrete memoryless multiple access channels. *IEEE Transactions on Information Theory*, 28:93–95, January 1982.

[75] Xinping Yi and David Gesbert. Precoding methods for MISO broadcast channel with delayed CSIT. *arXiv preprint arXiv:1207.2103*, 2012.

[76] Ram Zamir and Meir Feder. On lattice quantization noise. *Information Theory, IEEE Transactions on*, 42(4):1152–1159, 1996.

[77] Ram Zamir, Shlomo Shamai, and Uri Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6):1250–1276, 2002.

[78] Yan Zhu and Dongning Guo. Ergodic fading Z-interference channels without state information at transmitters. *IEEE Transactions on Information Theory*, 57(5):2627–2647, 2011.