Title
On the Capacity of Noncoherent Wireless Networks

Author
Sebastian, Joyson

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Los Angeles

On the Capacity of Noncoherent Wireless Networks

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical Engineering

by

Joyson Sebastian

2018

ABSTRACT OF THE DISSERTATION

On the Capacity of Noncoherent Wireless Networks

by

Joyson Sebastian
Doctor of Philosophy in Electrical Engineering
University of California, Los Angeles, 2018
Professor Suhas N. Diggavi, Chair

Wireless networks are characterized by variation in the network states. In practice the variations are combated by allocating resources for learning the network states. In networks with high mobility users, the variations are fast enough so that allocating separate resources may significantly deplete the resources and quality of communication. In this thesis we study the optimal schemes for nonchoherent networks, where the network channel states are unknown and are changing within given time periods. We address the question on how to optimally allocate the resources for training and communication.

In the first part of the thesis, we consider single flow noncoherent wireless networks, where there is a single information source and a single destination. A simple nontrivial version of this is the noncoherent multiple input multiple output (MIMO) network. We consider the noncoherent MIMO with asymmetric link strengths, which would arise when the antennas are well separated. Examples of this are in the 5G architecture where the basestations can cooperate through a backhaul and when there is device-to-device cooperation through a sidechannel. The study of noncoherent MIMO is also fundamental in understanding the nature of noncoherent networks in the sense that the cut-sets in noncoherent networks form a MIMO. We prove that for single input multiple output (SIMO) and multiple input single input (MISO) networks, it is optimal to use the statistically best antenna. For 2×2 MIMO with symmetric statistics *i.e.,* the direct links have identical statistics and so do the cross links, we derive the generalized degrees of freedom (gDoF) and prove that training-based

schemes are not optimal. For larger $M \times M$ MIMO we prove that in general, a training scheme that learns all the channel parameters is not optimal in gDoF measure. We then proceed to study the noncoherent diamond network (2-relay channel). We prove that in certain regimes it is optimal to perform a relay selection and operate the network. In other regimes where we need to operate both the relays, it is not optimal to learn all the channel states through training. We propose a novel scheme that partially trains the network and combine it with scaling at the relays and quantize-map-forward operation and prove that our scheme is gDoF optimal.

In the second part of the thesis, we consider multiple flow noncoherent wireless networks. We specifically consider the noncoherent 2-user interference channel, where both the transmitters and the receivers do not know the channels strengths, but the statistics are known. For studying this, we first consider the fast fading interference channel (FF-IC) where the transmitters do not know the channel, but the receivers do know the channel. We extend the existing rate-splitting schemes when the channels are known at the receivers, to the fast fading case by performing rate-splitting based on the statistics of the channel. We prove that this scheme achieves the capacity approximately for a wide range of fading models. With this result for the FF-IC, we proceed to the noncoherent IC. We propose a noncoherent scheme with rate-splitting based on the statistics of the channel. We prove that this schemes achieves higher gDoF than a training-based scheme. The results extend to the case of noncoherent IC with feedback, where the outputs at the receivers are fed back to the corresponding transmitter.

The dissertation of Joyson Sebastian is approved.

Songwu Lu

Richard D. Wesel

Christina Panagio Fragouli

Suhas N. Diggavi, Committee Chair

University of California, Los Angeles

2018

Table of Contents

List of Figures

List of Tables

Acknowledgments

I thank God who has led me over the years to this point in completing my Ph.D. It would have been impossible to go through the challenges without his Grace.

I thank my advisor Professor Suhas Diggavi for his support, guidance and patience throughout the last 5 years. His guidance has brought to me understanding on aspects of research that I would have been otherwise unaware of. I thank Professor Christina Fragouli for her guidance in the joint research work and also for serving in my doctoral committee. I especially thank Can Karakus who mentored me in the beginning of the Ph.D work. I thank my labmates Shaunak Mishra, Jad Hachem, Ayan Sengupta, Mehrdad Shoukatbaksh and Yahya Ezzeldin for their companionship and support during this journey.

I would also like to thank Professors Rick Wesel and Songwu Lu for taking the time to serve in my doctoral committee.

I thank National Science Foundation and Guru Krupa Foundation, whose funding supported my doctoral research[1].

I gratefully acknowledge the roles my parents Sebastian and Mariakutty have played in my journey this far. I thank Uncle Bob and Aunt Alice for their support during last five years.

[1]The work in this thesis was supported in part by NSF grants #1314937 and #1514531, and gifts from Guru Krupa Foundation.

Vita

2013-2018	Graduate Student Researcher, Electrical Engineering University of California, Los Angeles
Summer 2016	Research Intern Qualcomm Research, San Diego, CA
June 2015	M.S., Electrical Engineering University of California, Los Angeles
August 2012	B.Tech., Electroics and Electrical Communication Engineering Indian Institute of Technology, Kharagpur
Summer 2011	Research Intern Tata Institute of Fundamental Research (TIFR), Mumbai, India

Publications

Joyson Sebastian and Suhas Diggavi, "Generalized Degrees of Freedom of Noncoherent Diamond Networks", arxiv:1802.02667, *Submitted to IEEE Transactions on Information Theory, 2018.*

Joyson Sebastian, Ayan Sengupta and Suhas Diggavi, "Generalized Degrees Freedom of Noncoherent MIMO with Asymmetric Links", arxiv:1705.07355, *Submitted to IEEE Transactions on Information Theory, 2018.*

Joyson Sebastian, Can Karakus and Suhas Diggavi, "Approximate Capacity of Fast Fading

Interference Channels with No Instantaneous CSIT", arxiv:1706.03659, *Accepted in IEEE Transactions on Communications, 2018.*

Joyson Sebastian, Ayan Sengupta and Suhas Diggavi, "On Capacity of Noncoherent MIMO with Asymmetric Link Strengths", *IEEE International Symposium on Information Theory, Aachen, Germany, Jun. 2017.*

Joyson Sebastian, Can Karakus and Suhas Diggavi, "Approximately achieving the feedback interference channel capacity with point-to-point codes", *IEEE International Symposium on Information Theory, Barcelona, Spain, Jul. 2016.*

P. N. Karthik, Raksha Ramakrishna, Geethu Joseph, Chandra R. Murthy, Joyson Sebastian and Neelesh B. Mehta, "Model-based interference cartography and visualization", *National Conference on Communication (NCC), Guwahati, India, Mar. 2016.*

Joyson Sebastian, Can Karakus, Suhas Diggavi and I-Hsiang Wang, "Rate splitting is approximately optimal for fading Gaussian interference channels", *Allerton Conference on Computing, Communication and Control, Monticello, IL, USA, Oct. 2015.*

Joyson Sebastian and Neelesh B. Mehta, "Optimal, distributed, timer-based best two relay discovery scheme for cooperative systems", *IEEE Global Communications Conference (Globecom), Atlanta, GA, USA, Dec. 2013.*

CHAPTER 1

Introduction

1.1 Noncoherent Networks

Significant progress has been made in the past four decades in the understanding of the capacity of wireless networks. There has been breakthrough with the approximation approach in [ADT11] to obtain the capacity of single flow networks. However, the focus has been on a given wireless networks, *i.e.,* the network states was assumed to be known. The inherent variability in the wireless networks can cause this assumption to not hold in general. One approach is to learn the network states and then communicate through the learned network. However there is a tradeoff between the amount of resources to be allocated in learning network states and that to be allocated for communication, for example as demonstrated in [MH99, ZT02], it is not optimal to learn all the channel strengths of a large multiple input multiple output (MIMO) system. The question we ask in this work is similar, but for networks. We assume that the network topologies and channel statistics are known, however the actual channel strengths are unknown. This is one of the building steps in understanding the capacity of a general time varying wireless network. We consider a block fading noncoherent channel model *i.e.,* the channel remains constant for a block of T symbol periods and are independently distributed across the blocks with a Rayleigh distribution. The channel statistics for a general network could be asymmetric since the location of the nodes are largely variable, *i.e.,* there could be significant variation in the strengths. This consideration moves our work to a generalized degrees of freedom framework (gDoF) in contrast to the degrees of freedom framework (DoF) adopted in the existing works on noncoherent networks. In this work, we first consider single flow noncoherent networks including noncoherent MIMO and noncoherent diamond channels. Then we move on to study multiple flow noncoherent networks: we study the gDoF of a noncoherent interference channel.

1.1.1 Noncoherent MIMO

The noncoherent MIMO with asymmetric link strengths is the first nontrivial model in studying the general noncoherent wireless network. In the study of networks, one can think of the cut-set as a distributed MIMO where the nodes are widely separated, resulting in the noncoherent MIMO with asymmetric link strengths (see Figure 1.1).

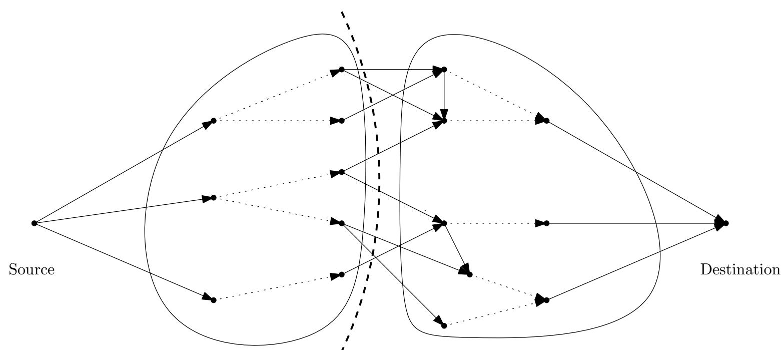


Figure 1.1: Noncoherent MIMO with asymmetric statistics can arise in the analysis of noncoherent networks.

The noncoherent MIMO can arise in the next generation wireless architecture that envisage dense deployment of access points [BLM14] and also in cloud radio access networks (CRAN) [WZH15]. These give rise to multiple access points connected through a (reliable) backhaul and can effectively form a system of widely separated antennas which could be used for coordinated transmission and reception, *e.g.* coordinated multipoint COMP [IDM11]. The widely separated antennas could have disparate average strengths motivating our model, especially for multiple input single output (MISO) and single input multiple output (SIMO) channels with asymmetric link strengths. This is illustrated in Figure 1.2.

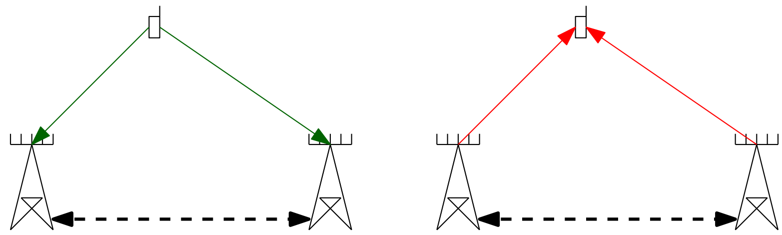


Figure 1.2: Noncoherent SIMO and MISO with asymmetric statistics can arise in COMP architecture where multiple basestations can cooperate through the backhaul.

The MIMO case arises when the receiver could be widely spread (see Figure 1.3) as would be the case when users can cooperate using a separate sidechannel [KD17].

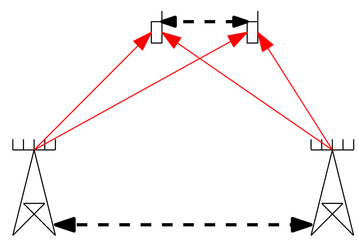


Figure 1.3: Noncoherent MIMO with asymmetric statistics can arise with COMP architecture and device-to-device cooperation.

We study the noncoherent MIMO with a coherence period of T symbols and with channel statistics following independent Gaussian distributions. The average link strengths are assumed to scale with different SNR-exponents in different links. Compared to the MIMO with i.i.d. links, we obtain new structural results on the optimal signalling distribution. We prove that for $T = 1$, the gDoF is zero for MIMO channels with arbitrary link strength distributions, extending the result for MIMO with i.i.d links. We then show that selecting the statistically best antenna is gDoF-optimal for both MISO and SIMO channels.

We also derive the gDoF for the 2×2 MIMO channel with different exponents in the direct and cross links. We develop novel techniques for analyzing the gDoF of the 2×2

MIMO channel. We approximate and discretize the mutual information terms without losing the gDoF for the outer bound. The outer bound optimization problem reduces to a linear program after this discretization. The standard linear programming techniques applied to this scenario yield a distribution with only two mass points as the solution to the optimization problem. Later we use the structure of the problem to reduce the two mass points into a single point. Finding this single point turns out to be a piecewise linear optimization problem which can be solved explicitly. We believe that the outer bounding techniques we develop for the 2×2 MIMO would provide guidelines in studying general noncoherent networks. The approach we take in studying noncoherent diamond network is inspired by the techniques for 2×2 MIMO.

With our new outer bounds we show that it is always necessary to use both the antennas of the 2×2 MIMO with asymmetric link strengths, to achieve the optimal gDoF, in contrast to the results for 2×2 MIMO with identical link distributions. We show that having weaker crosslinks, gives gDoF gain compared to the case with identically distributed links. We observe that it is not optimal to allocate separate training symbols for 2×2 MIMO in general. We extend this observation to larger $M \times M$ MIMO with given SNR-exponents in direct and cross links, by demonstrating a strategy that can achieve larger gDoF than a training-based scheme.

1.1.2 Noncoherent diamond networks

The noncoherent wireless networks with multiple nodes and asymmetric link strengths has not been studied in literature (to the best of our knowledge) from a gDoF perspective. The work in [KK13] considered noncoherent single relay network with identical link strengths and unit coherence time. It was shown that under certain conditions on the fading statistics, the relay does not increase the capacity at high-SNR. In [GY14], similar observations were made for the noncoherent MIMO full-duplex single relay channel with block-fading. They showed that Grassmanian signaling can achieve the degrees of freedom (DoF) without using the relay. Also for certain regimes, decode-and-forward with Grassmanian signaling was shown to approximately achieve the capacity at high-SNR. The nodes being well separated in a network can give rise to significant difference in the channel strengths. Thus, in our work we consider a parallel 2-relay wireless network (diamond network) with asymmetric link strengths and study its gDoF-capacity.

Similar to the coherent diamond network, we demonstrate that it is gDoF-optimal to perform a relay selection in certain regimes of the noncoherent diamond network. This in effect gives a network simplification, similar to [NOF14], where it was shown that by selecting a subset of the relays in an n-relay network, most of the capacity could be achieved. In some of the regimes it is not sufficient to perform a relay selection. We analyze these regimes by considering a modified version of conventional cut-set outer bound. We derive the structure of the distribution that optimizes this outer bound. We discretize the outer bound optimization problem without losing gDoF and simplify it further using linear programming arguments. This helps us reduce the outer bound optimization problem to a bilinear optimization problem for gDoF optimality. We solve the bilinear optimization problem explicitly, its solution yields different regimes of operation for the network. The structure of the solution for the outer bound suggests a nonconcurrent operation of the relays in some regimes. In other regimes it suggests operating both relays, but reducing the power of one of the relays.

Based on the solution for the outer bound, we develop a novel achievability strategy. It involves partially training the network and the relays perform a scaling and quantize-map and forward operation. The scaling at the relays avoid the necessity of knowledge of all the channel states at the destination. We also demonstrate that a training-based scheme that trains the whole network is not gDoF-optimal. This further confirms that allocating separate resources for learning network states is not optimal, extending the observation from noncoherent MIMO.

1.1.3 Noncoherent interference channels

Moving on from single flow networks to multiple flow networks in the noncoherent setting, one of the simple scenarios is the two user noncoherent interference channel (IC). The capacity of the 2-user IC is well understood when the channel strengths are known at the transmitter and the receiver [HK81, CMG08, ETW08, ST11]. The case when channel strengths are not known at the receiver is not well studied in terms of its capacity. There have been a few works that study the DoF behaviour. The DoF region for the MIMO FF-IC was studied in [VV12] and their results showed that when *all users have single antenna*, the DoF region is same for the cases of no CSIT, delayed CSIT and instantaneous CSIT. The results from [TMP13] showed that DoF region for the FF-IC with instantaneous CSIT and no feedback contains the DoF region with output feedback and delayed CSIT.

To proceed to the noncoherent IC we first study the fast fading IC (FF-IC) where the transmitters know the channel statistics and do not have the channel state information, but we assume that the receivers do know the channel states. For the FF-IC we extend the rate-splitting schemes [ETW08, ST11] based on the interference to noise ratio (inr). We perform rate-splitting based on the average inr and demonstrate that this strategy can achieve the approximate capacity for common fading models including Rayleigh fading. The approximate capacity result is obtained for IC with feedback as well as for the IC without feedback.

We use the results for the FF-IC to study the noncoherent IC. The channel statistics are assumed to be known at the transmitter and we propose a noncoherent rate-splitting scheme. The rate-splitting is again performed according to the average inr. We prove that the noncoherent rate-splitting strategy achieves higher gDoF than a training-based scheme that uses 2 symbols to train the channels. We also demonstrate that when the average inr is low, treating interference as noise is better than the training-based scheme and noncoherent rate-splitting. We further consider the noncoherent IC with feedback and study a rate-splitting scheme based on average inr. For the feedback case, our rate-splitting scheme achieves larger gDoF than the case without feedback. Also, we show that the training-based schemes are not gDoF-optimal for the feedback case. Our scheme for the feedback case is better than treating interference as noise even when the average inr is small, for $T \geq 3$.

1.2 Future work

One of the future directions of study include n-relay diamond networks. Our achievability scheme for 2-relay diamond network can be extended to the n-relay case. However, the outer bounds are still an open problem. The more general open problem is the capacity of general noncoherent networks.

Another direction of study is to study backscatter communication systems in a noncoherent setting. Backscatter communication systems typically use a Reader and a radio frequency identification (RFID) tag [Dob12]. Reader transmits a radio frequency (RF) signal; the RFID tag adapts the level of its antenna impedance to vary the reflection coefficient and transmits data via reflecting and modulating the incident signal back to Reader [XYV14, BR14]. We have some preliminary results on backscatter systems with intersymbol-interference. We demonstrate that instead of using a constant carrier sequence, we can optimize the sequence to obtain larger rates or smaller bit error probability. This involves training the channel states. The noncoherent version of the problem will be to consider whether a noncoherent scheme can be designed to outperform the training-based schemes.

1.3 Outline

This thesis is organized as follows. In Chapter 2 we study the noncoherent MIMO with asymmetric link strengths and in Chapter 3 we study the noncoherent diamond channel. In Chapter 4 we study the fast fading interference channels (FF-IC) and in Chapter 5 the results for the FF-IC are extended to the noncoherent interference channel. In Chapter 6 we give our conclusions and the directions for future work.

We would like to point out that most of the results in this work have been accepted or have been submitted for publication. The work on FF-IC [SKD18] has been accepted in IEEE Transactions on Communications. The works on noncoherent MIMO [SD18c] and noncoherent diamond channel [SD18a] have been submitted to IEEE Transactions on Information Theory. The work on backscatter communication [SED18] has been submitted to IEEE Wireless Communications Letters. The work on noncoherent IC [SD18b] is to be submitted to IEEE Transactions on Information Theory.

CHAPTER 2

Noncoherent MIMO

2.1 Introduction

The capacity of fading Multiple Input Multiple Output (MIMO) channels when neither the receiver nor the transmitter knows the fading coefficients was first studied by Marzetta and Hochwald [MH99]. They considered a block fading channel model where the fading gains are identically independent distributed (i.i.d.) Rayleigh random variables and remain constant for T symbol periods. In [ZT02], Zheng and Tse introduced the idea of communication over a Grassmanian manifold for the noncoherent MIMO channel and derived the capacity behavior when the links are i.i.d. and the signal-to-noise-ratio (SNR) is high. Their characterization was tight for the capacity at large SNR, when the coherence time was large compared to number of antennas. In [YDR13], this tight characterization was extended to the case when the number of antennas was large compared to the coherence time. In [NYG17], spatially correlated (temporally flat within a block) noncoherent MIMO broadcast channel with statistical channel state information (CSI) was considered and an achievable Degrees of Freedom (DoF) region was derived.

Some works have especially considered the case with coherence time $T = 1$. The noncoherent single-input-single-output (SISO) channel with $T = 1$ was considered by Taricco and Elia [TE97]. They obtained the capacity behavior in asymptotically low and high SNR regimes. The noncoherent SISO with $T = 1$ was further studied by Abou-Faycal *et al.* [ATS01] and they showed that for any given SNR, the capacity is achieved by an input distribution with a finite number of mass points. For the noncoherent MIMO with

$T = 1$, Lapidoth and Moser [LM03] showed that for high SNR, the capacity behaves double logarithmically with SNR.

Noncoherent MIMO with temporal correlation within each fading block (instead of the constant block fading [MH99, ZT02]) has been considered in some of the recent works. In [MRY13], Morgenshtern *et al.* studied temporally correlated Rayleigh block-fading Single Input Multiple Output (SIMO) channel and showed that at high-SNR, the SIMO channel can have larger DoF than SISO channel, under some mild assumptions on the temporal correlation. Similar results was derived for noncoherent MIMO with temporally correlated block fading in [KRD14], where it was shown that the temporally correlated noncoherent MIMO can have a larger DoF compared to the constant block fading case.

Some works have studied noncoherent networks (with more than two nodes) for the high-SNR capacity behavior. In [Lap05], the zero DoF result for the noncoherent MIMO with $T = 1$ from [LM03] was extended to noncoherent networks with $T = 1$. Koch and Kramer studied the noncoherent single relay network [KK13] and showed that under certain conditions on the fading statistics, the relay does not increase the capacity at high SNR. In [GY14], the noncoherent MIMO full-duplex single relay channel with block-fading was studied, and it was shown that Grassmanian signaling can achieve the DoF without using the relay. Also, the results in [GY14] show that for certain regimes, decode-and-forward with Grassmanian signaling can approximately achieve the capacity at high SNR and the characterization was determined by the number of antennas at the nodes.

To the best of our knowledge, the existing works considers a DoF framework for studying the noncoherent channels, *i.e,* the links in the network scale with same SNR exponent. However, in networks the links could have asymmetry in the channel strengths and a gDoF framework where the link strengths are scaled with different exponents of SNR could better capture the system behavior. We consider noncoherent MIMO with asymmetric link strengths as a first step in the direction of studying the asymmetric noncoherent networks.

2.1.1 Contributions and outline

In this chapter, we consider a noncoherent channel model with coherence time of T symbol periods and asymmetric link distributions, where the link strengths are scaled with different exponents of SNR. In essence, we are moving from the DoF-framework in [MH99, ZT02] to the generalized DoF of noncoherent MIMO channels.

Next generation wireless architecture envisage dense deployment of access points [BLM14]. Another architectural proposal is to use cloud radio access networks (CRAN) [WZH15]. These imply that multiple access points could be connected through a (reliable) backhaul. The implication of this is that of widely separated antennas, which form a virtual antenna array. Such widely separated antennas could be used for coordinated transmission and reception, *e.g.* coordinated multipoint COMP [IDM11]. These widely separated antennas could have disparate average strengths motivating this model (especially SIMO and MISO). This is illustrated in Figure 1.2 on page 3. The MIMO case arises when the receiver could be widely spread (see Figure 1.3 on page 3) as would be the case when users can cooperate using a separate sidechannel [KD17]. Another motivation for this model comes from the study of networks. Here one can think of the cut-set as a distributed MIMO (see Figure 1.1 on page 2) where the nodes are widely separated again resulting in this model. The asymmetric case is also motivated by a fundamental question about the robustness of the results in [MH99, ZT02] to changes in the i.i.d. channel model.

For our channel model with arbitrary (fading) link strengths, we show in Theorem 2.1 that the capacity achieving input distribution is of the form LQ where L is lower triangular and Q is independent of L and is unitary isotropically distributed. This is in contrast to the result for the i.i.d. setting, which yields a diagonal matrix instead of L multiplying Q [MH99]. In Theorem 3.9, we demonstrate that the gDoF of a SIMO channel can be achieved by retaining only the signal received by the best receive antenna. The gDoF result for the SIMO channel is used in Theorem 2.5 to show that for $T = 1$, the gDoF is always zero for a MIMO channel of any size. In Theorem 3.10, we show that the gDoF of the MISO channel can be achieved by only signaling over the statistically best transmit antenna.

In a setting with with N receive antennas, when the exponents in the SNR-scaling are same for all the links (i.i.d. setting), the number of transmit antennas M, required to attain the optimal DoF was shown to be $\min\left(\lfloor T/2 \rfloor, N\right)$, in [ZT02]. They showed that increasing the number of transmit antennas beyond this value reduces the DoF. In this chapter, we provide evidence that this is not the case when the SNR exponents are different: in Theorem 2.7, we show that for a 2×2 MIMO with different exponents in direct and cross links, and $T = 2$, both transmit antennas are required to achieve the optimal gDoF. We also show that having smaller exponents in cross links lead to gDoF gain of $(2/T)\,\gamma_{\text{diff}}$ over the case with same exponents in all the links, where γ_{diff} is the difference in the SNR exponents. In showing this, several novel techniques were needed. In particular we would like to highlight the technique used in Lemma 3.1, where in the optimization problem to find the optimal input distribution for the outer bound, we show that the optimal gDoF can be achieved by a point mass distribution. To arrive at this, we discretized the input distribution without a loss in gDoF, and subsequently used linear programming arguments to show that there exists an optimal distribution with just one mass point. We believe that our techniques for the 2×2 MIMO provide intuitions for studying larger noncoherent networks, especially in analyzing the cut-sets.

Traditional training-based schemes for MIMO systems allocate a training symbol to train each transmit antenna independently. Our results for the 2×2 MIMO also demonstrate that a traditional training based scheme is not gDoF optimal. Our scheme has a gDoF gain of $(2/T)\,\gamma_{\text{diff}}$ compared to a training based scheme. In Theorem 2.8, we extend this observation to larger $M \times M$ MIMO with given SNR exponents in direct and cross links, where we demonstrate a strategy that can achieve larger gDoF than a training based scheme.

Extending our outer bounds to the general MIMO seems a difficult task at the moment; the LQ transformation process used for deriving the outer bound for 2×2 MIMO as done in (2.43), (2.44), (2.45), (2.46) and the subsequent Lemmas (Lemma 3.3, Lemma 2.5 and Lemma 2.6) for bounding the terms in those equations do not easily extend to 3×3 or higher

Table 2.1: Important abbreviations

Abbreviation	Meaning
$\mathcal{CN}$	Circularly symmetric complex Gaussian
Tran	Transpose
DoF	Degrees of freedom
gDoF	Generalized degrees of freedom
SNR	Signal-to-noise ratio
QMF	Quantize-map-forward

MIMOs.

Outline: The rest of this chapter is organized as follows: in Section 2.2, we give the notations and set up the system model; Section 2.3 presents our main results, and Section 3.4 provides some analysis and proofs. Some details of the proofs are deferred to the Appendixes. In some cases, discussion of results in Section 2.3 will refer to lemmas and facts detailed in Section 3.4.

2.2 Notation and system model

2.2.1 Notational conventions

We use the notation $\mathcal{CN}(\mu, \sigma^2)$ for circularly symmetric complex Gaussian distribution with mean μ and variance σ^2. We use the symbol $\sim$ with overloaded meanings: one to indicate that a random variable has a given distribution and second to indicate that two random variables have the same distribution. The logarithm to base 2 is denoted by $\log()$. The notation $A^\dagger$ indicates the Hermitian conjugate of a matrix A and $\mathsf{Tran}(A)$ indicates the transpose of A. We also list the important abbreviations and notations used, in Table 2.1 and in Table 2.2 respectively. The degrees of freedom (DoF) for a network is defined as

Table 2.2: Important notations

Notations	Meaning
$x \sim y$	Random variables x, y have same distribution
$x \sim p$	Random variable x has the distribution p
$A^\dagger$	Hermitian conjugate of a matrix A
$\doteq$	Order equality
$(\mathcal{P})$	Optimal value of an optimization problem $\mathcal{P}$

$$\lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log(\mathsf{SNR})}$$

when the different link strengths in the network scales proportional to SNR, where $C(\mathsf{SNR})$ is the capacity for a given value of SNR. When the different link strengths in the network scales with different SNR exponents, the above formula defines the gDoF. We use the notation $\doteq$ for order equality, *i.e.*, we say $f_1(\mathsf{SNR}) \doteq f_2(\mathsf{SNR})$ if

$$\lim_{SNR\to\infty} \frac{f_1(\mathsf{SNR})}{\log(\mathsf{SNR})} = \lim_{SNR\to\infty} \frac{f_2(\mathsf{SNR})}{\log(\mathsf{SNR})}.$$

The use of symbols $\dot{\leq}, \dot{\geq}$ are defined analogously. The script $\mathcal{P}$ is used to indicate an optimization problem and $(\mathcal{P})$ is used to denote the optimal value of the objective function. We use the notation

$$\text{gDoF}(\mathcal{P}) = \lim_{SNR\to\infty} \frac{(\mathcal{P})}{\log(\mathsf{SNR})}$$

to indicate the scaling of the optimal value of $\mathcal{P}$ when it depends on SNR. We do not set aside separate symbols/notation for constants or random variables or scalars or vectors or matrices, as this is made clear from the context. By default a symbol is a constant scalar, otherwise we define it as random/vector/matrix when it is introduced. When G and X are matrices, GX indicates matrix multiplication. When g is scalar and X is vector/matrix, gX indicates g multiplying each element of X. When we have $g^n = g(1), \ldots, g(n)$ and $X^n = X(1), \ldots, X(n)$ with $g(i)$ being a scalar and $X(i)$ being a vector, then $g^n X^n$ is a short notation for $g(1)X(1), \ldots, g(n)X(n)$. Also, when $\hat{g}^n = \hat{g}(1), \ldots, \hat{g}(n)$ with $\hat{g}(i)$

being a scalar and g^n, X^n being same as previously defined, then $\frac{g^n X^n}{\hat{g}^n}$ is a short notation for $\left(\frac{g(1)}{\hat{g}(1)}\right) X(1), \ldots, \left(\frac{g(1)}{\hat{g}(1)}\right) X(n)$.

2.2.2 System Model

We consider a block-fading MIMO channel with M transmit and N receive antennas, and a coherence time of T symbol durations. The signal flow (over a blocklength T) is given by:

$$Y = GX + W \tag{2.1}$$

where X is the $M \times T$ matrix of transmitted symbols with rows $\underline{X_i}$ corresponding to each transmit antenna; G represents the $N \times M$ channel matrix (which is independently generated every T symbols), and its elements g_{ij} are independent with $g_{ij} \sim \mathcal{CN}\left(0, \rho_{ij}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{ij}}\right)$, where the exponents γ_{ij} are (constant) parameters of the MIMO channel. For convenience, we also use the notation $\underline{\rho}^2(n)$ to denote the vector of channel strengths to n^{th} receiver antenna. The columns of G are $\overline{g_i}$ corresponding to channels from each transmit antenna. The variable Y represents the $N \times T$ matrix of received symbols, with rows corresponding to each receive antenna; and W is an $N \times T$ noise matrix with elements $w_{ij} \sim$ i.i.d. $\mathcal{CN}(0, 1)$. The transmit signals have the average power constraint:

$$\frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \mathbb{E}\left[|x_{mt}|^2\right] = 1. \tag{2.2}$$

2.3 Main results

In this section we go through the main results in our chapter. We first look at general results for the noncoherent MIMO with asymmetric link strengths. We prove a structural result for the optimizing distribution for the noncoherent MIMO in Theorem 2.1. This result has some similarities to that for the noncoherent MIMO with i.i.d. links in the sense that part of the structure is similar. Another general result that we prove is for the noncoherent MIMO with a channel structure that can be decomposed into smaller disjoint channels. We prove similar to the coherent case, that the power can be allocated across the disjoint parts and coding can be done separately among the disjoint parts to achieve the capacity. This result is proved in Theorem 2.2. This result can be used to derive the gDoF of noncoherent parallel channels. This is stated as Corollary 2.3.

Then we look at noncoherent MIMO with specific structures. We consider the noncoherent SIMO in Theorem 3.9 and derive its gDoF. We prove that the gDoF is achieved by using the statistically best antenna. The gDoF result for SIMO can be used to prove that for any MIMO, the gDoF is zero for $T = 1$, by decomposing the MIMO into SIMO parts and constructing a channel with larger capacity. We obtain this result in Theorem 2.5. Next we consider the noncoherent MISO and prove a similar result, that the gDoF can be achieved using the statistically best antenna. This is proved in Theorem 3.10.

The next specific structure we look at is the noncoherent 2×2 MIMO with a given SNR exponent in the direct links and another SNR exponent in the cross links. We handle this in Theorem 2.7. We observe that training-based schemes are not gDoF optimal for the 2×2 MIMO. We extend this observation to larger $M \times M$ MIMO in Theorem 2.8.

2.3.1 Results for general noncoherent MIMO

Theorem 2.1. *The capacity of the noncoherent MIMO system can be achieved with X of the form $X = LQ$ with L being a lower triangular matrix and Q being an isotropically distributed unitary matrix independent of L.*

Proof. The proof is given in Section 2.4.2. □

This theorem is in contrast with the result for the case with G and W having i.i.d. Gaussian elements, where the structure of an optimal X could be written as $X = DQ$ where D is diagonal [MH99]. In our system model only W has i.i.d. elements which ends up restricting the structure to the form LQ.

Theorem 2.2. *Let the channel matrix G of the system be block diagonal as $G = diag\,(G_1, \ldots, G_K)$ where G_i are the diagonal blocks of G, then the capacity $C\,(P, diag\,(G_1, \ldots, G_K))$ of the chan-*

nel for a power P can be achieved by splitting power across the blocks: $C\left(P, \mathit{diag}\left(G_1, \ldots, G_K\right)\right) = \max_{P_1+\cdots+P_K \leq P}\left(C\left(P_1, G_1\right) + \cdots + C\left(P_K, G_K\right)\right)$.

Proof idea. This result holds for coherent MIMO and the proof for noncoherent case is similar. We just need to show $C\left(P, \operatorname{diag}\left(G_1, G_2\right)\right) = \max_{P_1+P_2 \leq P}\left(C\left(P_1, G_1\right) + C\left(P_2, G_2\right)\right)$ because of induction. Let X_{G1}, X_{G2} be the transmitted symbols in G_1 and G_2 of the channel. Similarly Y_{G1}, Y_{G2} be the corresponding received symbols. Now $I\left(X;Y\right) \leq I\left(X_{G1};Y_{G1}\right) + I\left(X_{G2};Y_{G2}\right)$ because $\left(X_{G2}, Y_{G2}\right) - X_{G1} - Y_{G1}$, $\left(X_{G1}, Y_{G1}\right) - X_{G2} - Y_{G2}$ are Markov chains and the desired result easily follows. The detailed steps are given in Appendix A.3. □

Now we have the following corollary from the above theorem.

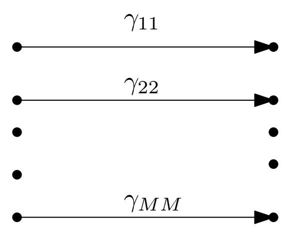


Figure 2.1: Parallel channels with given SNR exponents.

Corollary 2.3. *The gDoF of parallel channel system (Figure 2.1)* $G = \mathit{diag}\begin{pmatrix} g_{11} & . & . & g_{MM} \end{pmatrix}$ *with links* $g_{ii} \sim \mathcal{CN}\left(0, \rho_{ii}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{ii}}\right)$ *is* $\sum_i \left(1 - \frac{1}{T}\right)\gamma_{ii}$

Proof. This is obtained by decomposing into individual channels (Theorem 2.2) and using the SISO results from [ZT02]. For SISO with link g_{ii} distributed according to $\mathcal{CN}\left(0, \rho_{ii}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{ii}}\right)$, the gDoF is $\left(1 - \frac{1}{T}\right)\gamma_{ii}$ for any T. □

2.3.2 SIMO and MISO channels

In this subsection we consider the noncoherent SIMO and MISO channels with asymmetric link strengths. The gDoF result for SIMO can be easily derived by extending the results for the case with i.i.d. links. The gDoF result for SIMO can be extended to arbitrary MIMO case, for $T = 1$. For the MISO case, the existing techniques are not sufficient for computing

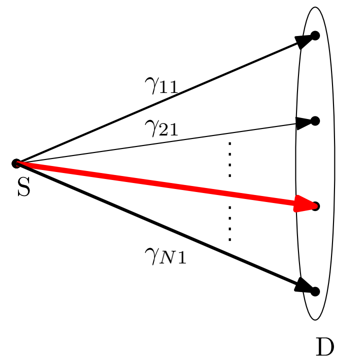


Figure 2.2: Noncoherent SIMO with given SNR exponents.

the outer bound. We develop new techniques, manipulating entropy expressions using Linear Algebra techniques to derive the gDoF of MISO.

Theorem 2.4. *For the noncoherent SIMO channel (Figure 2.2) with* $G = \mathsf{Tran}\left(\begin{bmatrix} g_{11} & . & . & g_{N1} \end{bmatrix}\right)$, *where* $g_{i1} \sim \mathcal{CN}\left(0, \rho_{i1}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{i1}}\right)$, *the gDoF is* $\left(1 - \frac{1}{T}\right) \max_i \gamma_{i1}$, *i.e., the gDoF can be achieved by using only the statistically best receive antenna.*

Proof. We only need to prove the outer bound, since achievability follows by using the statistically best receive antenna. The outer bound can be proved as an extension of results for SIMO with identical links. Let $\rho_*^2 = \max_i \rho_{i1}^2$. Now with W being $T \times 1$ noise vector with i.i.d. $\mathcal{CN}\left(0, 1\right)$ elements, G' being a $1 \times N$ channel matrix with i.i.d. $\mathcal{CN}\left(0, \mathsf{SNR}^{\gamma*}\right)$ elements, W_1 being a noise vector with independent (but not identical) Gaussian elements $w_{1i} \sim \mathcal{CN}\left(0, \rho_*^2 - \rho_{i1}^2\right)$ and K being a constant diagonal matrix with elements $k_{ii} = \frac{\rho_{1i}}{\rho_*}$, we have $K\left(G'X + W + W_1\right)$ with same distribution as

$$Y = GX + W.$$

Hence by data processing inequality $I\left(X; GX + W\right) \leq I\left(X; G'X + W\right)$. Now due to the results for i.i.d. noncoherent MIMO [ZT02], we have $I\left(X; G'X + W\right) \dot{\leq} \left(T - 1\right) \log\left(\rho_*^2\right)$. Hence the required result follows. $\square$

The gDoF result for SIMO can now be used to prove that for $T = 1$, the gDoF is zero for any MIMO.

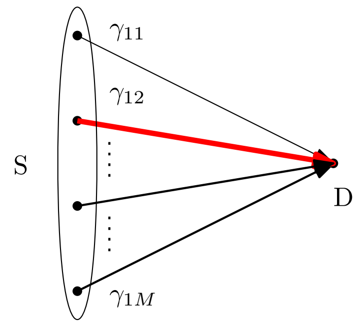


Figure 2.3: Noncoherent MISO with given SNR exponents.

Theorem 2.5. *(gDoF of arbitrary MIMO for $T = 1$) For any G with $T = 1$ the gDoF is zero.*

Proof. This can be easily shown by examining $\overline{g_i}$, the channels and $\underline{X_i}$, the symbols respectively from the i^{th} antenna, from $G = \begin{bmatrix} \overline{g_1} & \overline{g_2} & . & . & \overline{g_N} \end{bmatrix}$, the channel and $X = \mathsf{Tran}\begin{bmatrix} \mathsf{Tran}\left(\underline{X_1}\right) & \ldots & \mathsf{Tran}\left(\underline{X_N}\right) \end{bmatrix}$ the symbols for the whole MIMO. Now consider N SIMO channels $Y_i = \overline{g_i}\underline{X_i} + \frac{W_i}{\sqrt{N}}$, where W_i and W have same distribution but are independent. Now

$$
\begin{aligned}
& I\left(X; GX + W\right) \\
& \quad \leq I\left(X; \overline{g_1}\underline{X_1} + \frac{W_1}{\sqrt{N}}, \ldots, \overline{g_N}\underline{X_N} + \frac{W_N}{\sqrt{N}}\right)
\end{aligned} \tag{2.3}
$$

using data processing inequality since $\sum_{i=1}^{N} \frac{W_i}{\sqrt{N}} \sim W$ and $\sum_{i=1}^{N} \overline{g_i}\underline{X_i} = GX$. This creates a new channel which is decomposable into N SIMO channels and has higher capacity than the original channel. Hence the required result follows using Theorem 2.2 to decompose the SIMO channels and since each SIMO channel has zero gDoF from Theorem 3.9. □

Note that the above result is a generalization of the zero DoF result for MIMO from [LM03]. In their model the channel statistics is fixed and the power of the i.i.d. noise is scaled. But our result is more general in the sense that we allow the fading channel strengths to be scaled with different exponents.

Theorem 2.6. *For the noncoherent MISO channel (Figure 2.3) with* $G = \left[\begin{array}{cccc} g_{11} & . & . & g_{1M} \end{array}\right]$*, the gDoF is* $\left(1 - \frac{1}{T}\right) \max_i \gamma_{1i}$*, i.e., the gDoF can be achieved by only using the statistically best transmit antenna.*

Proof idea. We only need to prove the outer bound. In this case Y is a column vector and $h(Y)$ can be evaluated using Lemma 3.2. Also, we prove that $h(Y|X) \dot{\geq} \mathbb{E}\left[\log\left(1 + \sum_{i=1}^{M} \rho_{1i}^2 \left\|\underline{X_i}\right\|^2\right)\right]$ using Linear Algebra techniques. With these two results, the gDoF result follows. See Section 2.4.3 for details. □

Remark 2.1. *If one were to train the antennas and select the best antenna, the gDoF achievable is* $\left(1 - \frac{M}{T}\right) \max_i \gamma_{1i}$*, this is lower than what we can achieve using the statistically best transmit antenna. In a practical setting for a* 2×1 *MISO system with coherence time* $T = 3$*,* $\mathsf{SNR} = 10$ *dB, and links with average strengths* 0.1 *(*i.e, *-10 dB) and* 0.03 *(*i.e, *-15 dB), we believe that using training to select the best antenna within every coherence period, would be suboptimal. Once training is performed, there is only* 1 *time slot that can be used to communicate data. We believe that using the statistically best antenna only, without training both antennas, would yield a better rate.*

2.3.3 2×2 MIMO

In this subsection we describe the results for the 2×2 MIMO with SNR exponents γ_D in the direct links and γ_{CL} in the crosslinks. We describe our outer bound and obtain a signaling distribution to achieve the outer bound in gDoF. The signaling distribution uses the structure of our solution for outer bound optimization problem.

Theorem 2.7. *For the* 2×2 *symmetric noncoherent MIMO with*

$$G = \left[\begin{array}{cc} g_{11} & g_{12} \\ g_{21} & g_{22} \end{array}\right],$$

where $g_{11} \sim g_{22} \sim \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_D}\right)$ *,* $g_{12} \sim g_{21} \sim \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{CL}}\right)$ *and* $\gamma_D \geq \gamma_{CL}$ *(*$\gamma_D \geq \gamma_{CL}$ *is*

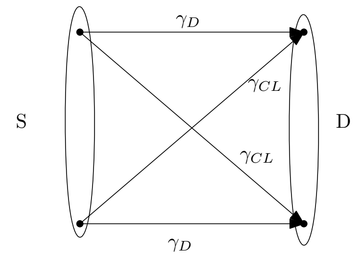


Figure 2.4: 2×2 MIMO with SNR exponents γ_D in the direct links and γ_{CL} in the crosslinks

without loss of generality), the gDoF is given in Table 2.3, and can be achieved by

$$X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ \eta & c & 0 & . & . & 0 \end{bmatrix} Q,$$

where $\eta \sim \mathcal{CN}\left(0, |b|^2\right)$ *independent of the unitary isotropic* Q, $|a|^2 = \mathsf{SNR}^{-\gamma_a}$, $|b|^2 = \mathsf{SNR}^{-\gamma_b}$, $|c|^2 = \mathsf{SNR}^{-\gamma_c}$, *and the values of* $(\gamma_a, \gamma_b, \gamma_c)$ *are as given in Table 2.3.*

Table 2.3: gDoF of 2×2 MIMO with $\gamma_{11} = \gamma_{22} = \gamma_D \geq \gamma_{CL} = \gamma_{12} = \gamma_{21}$

Regime	Solution $(\gamma_a, \gamma_b, \gamma_c)$	gDoF
$T = 2$	$(0, 0, \gamma_{CL})$	$\gamma_D - \frac{1}{2}\gamma_{CL}$
$T \geq 3$	$(0, 0, 0)$	$2\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{1}{T}\gamma_{CL}\right)$

Proof idea. From Theorem 2.1, we have an optimal distribution of the form

$$X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q, \tag{2.4}$$

where Q is unitary isotropically distributed and independent of a, b, c. We first obtain a capacity outer bound as the maximum of the expected value of a function $f\left(|a|^2, |b|^2, |c|^2\right)$. This is using Lemma 3.3, Lemma 2.5 and Lemma 2.6 which help to convert the entropy terms $h()$ into expected values. Then in Lemma 3.1 we prove that the maximization of $\mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right]$ can be achieved with a single mass point of $\left(|a|^2, |b|^2, |c|^2\right)$ for optimal gDoF. Then the gDoF outer bound can be expressed as the maximum of a piecewise linear optimization problem, which yields the solution as above. The detailed proof of the outer bound is in Section 2.4.4. The inner bound can be verified by using the distribution stated in the Theorem to evaluate the mutual information, the calculation is given in Appendix A.4. $\square$

Note that the above result shows that we need to use both antennas for achieving the gDoF for $T = 2$, since with only one antenna we can only achieve $\frac{1}{2}\gamma_D$ from Theorem 3.9. This is in contrast to the result for 2×2 MIMO with i.i.d. links, where the optimal gDoF could be achieved using a single transmit antenna for $T = 2$; also it was shown that using both antennas was sub-optimal [ZT02]. For $T \geq 3$ for a 2×2 MIMO with all exponents γ_D, the gDoF is $2\left(1 - \frac{2}{T}\right)\gamma_D$ [ZT02], whereas in our model with direct link exponents γ_D and cross link exponents γ_{CL}, the gDoF is $2\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{1}{T}\gamma_{CL}\right)$. Thus having weaker crosslinks gives a gDoF gain of $\frac{2}{T}\left(\gamma_D - \gamma_{CL}\right)$. Also as $T \to \infty$ the gDoF achieved is $2\gamma_D$, which agrees with the gDoF result for coherent MIMO [CTK14].

Also it is clear that training-based schemes are suboptimal for the 2×2 MIMO. For $T = 2$, if one were to train the links, one has to use two time slots, which leaves no time for communicating. For $T > 3$, the gDoF achievable after using two time slots to communicate is $2\left(1 - \frac{2}{T}\right)\gamma_D$ which is less than the gDoF $2\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{1}{T}\gamma_{CL}\right)$ that we achieve. The gain in gDoF, that we have, is $\frac{2}{T}\left(\gamma_D - \gamma_{CL}\right)$.

2.3.4 Nonoptimality of training

We observed in the previous subsection that training-based schemes cannot achieve the gDoF for 2×2 MIMO. We can extend this observation to larger MIMO. We specifically consider the $M \times M$ MIMO with exponents γ_D in direct links and γ_{CL} in crosslinks ($\gamma_D > \gamma_{CL}$). Using the following theorem, we prove that training-based schemes can be suboptimal for this case.

Theorem 2.8. *A gDoF of* $M\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{M-1}{T}\gamma_{CL}\right)$ *can be achieved for an* $M \times M$ *MIMO with coherence time* $T > M$ *and with exponents* γ_D *in direct links and* γ_{CL} *in crosslinks*

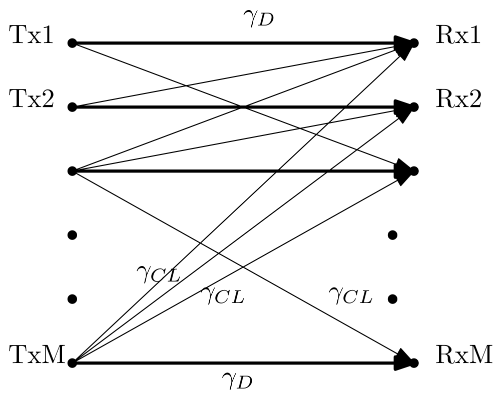


Figure 2.5: $M \times M$ MIMO with exponents in direct links γ_D and in crosslinks γ_{CL}

($\gamma_D > \gamma_{CL}$) (Figure 2.5), by using i.i.d. Gaussian codebooks across antennas and time periods.

Proof. In this case, the channel matrix G has diagonal elements g_{ii} distributed according to $\mathcal{CN}(0, \mathsf{SNR}^{\gamma_D})$ and the rest of the elements are distributed according to $\mathcal{CN}(0, \mathsf{SNR}^{\gamma_{CL}})$. Using Gaussian codebooks, the rate $R \geq I(GX + W; X)$ is achievable with X being an $M \times T$ matrix with i.i.d. Gaussian elements. Analyzing this mutual information yields an achievable gDoF of
$M\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{M-1}{T}\gamma_{CL}\right)$ per symbol. The calculations are given in Appendix A.5. □

Note that the gDoF $M\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{M-1}{T}\gamma_{CL}\right)$ cannot be achieved by a conventional training scheme where all transmitters train independently. This is clear since it requires M symbols in every coherence period for training and the maximum gDoF achievable using the rest of symbols is $M\left(1 - \frac{M}{T}\right)\gamma_D$ [CTK14, (Theorem 2)], assuming the channels are available perfectly due to training. This is smaller than $M\left(\left(1 - \frac{1}{T}\right)\gamma_D - \frac{M-1}{T}\gamma_{CL}\right)$. Thus using Gaussian codebooks and not using training gives a gDoF gain of $\frac{M(M-1)}{T}(\gamma_D - \gamma_{CL})$.

2.4 Analysis

In this section we first state some mathematical preliminaries required for the analysis. Then in Section 2.4.2, we derive the structure of the capacity achieving distribution for noncoherent

MIMO. In Section 2.4.3, we prove the gDoF outer bounds for the noncoherent MISO and in Section 2.4.4 we derive the gDoF outer bounds for the 2×2 MIMO system.

2.4.1 Mathematical Preliminaries

Fact 2.1. *For exponentially distributed random variable ξ with mean μ_ξ and $a \geq 0, b > 0$, $\log(a + b\mu_\xi) - \gamma_E \log(e) \leq \mathbb{E}[\log(a + b\xi)] \leq \log(a + b\mu_\xi)$ where γ_E is the Euler's constant.*

Proof. This follows due to the results given in Section 4.3 (on page 99). □

Fact 2.2. *For chi-squared random variable $\chi^2(k)$ and $a \geq 0, b > 0$,*

$$\log(a + bk) - \frac{\log(e)\,2}{k} + \log\left(1 + \frac{1}{k}\right) \leq \mathbb{E}\left[\log\left(a + b\chi^2(k)\right)\right] \leq \log(a + bk)\,. \tag{2.5}$$

Proof. The result is proved in Section 4.3 (on page 99) for the Gamma distribution and the result for the chi-squared distribution follows as a special case. □

Fact 2.3. *For an exponential random variable ξ with mean μ_ξ and a given constant $b > 0$*

$$\mathbb{E}\left[\frac{b}{b+\xi}\right] = \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \Gamma\left(0, \frac{b}{\mu_\xi}\right) \leq \frac{b}{\mu_\xi} \ln\left(1 + \frac{\mu_\xi}{b}\right) < 1, \tag{2.6}$$

where $\Gamma(0, x)$ is the incomplete gamma function.

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\frac{b}{b+\xi}\right] &= \mathbb{E}\left[\frac{1}{1+\frac{\xi}{b}}\right] && (2.7)\\
&\overset{(i)}{=} \int_0^\infty \frac{b}{\mu_\xi} e^{-\frac{bx}{\mu_\xi}} \frac{1}{1+x} dx && (2.8)\\
&\overset{(ii)}{=} \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \int_1^\infty e^{-\frac{bx}{\mu_\xi}} \frac{1}{x} dx && (2.9)\\
&\overset{(iii)}{=} \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \int_{\frac{b}{\mu_\xi}}^\infty e^{-t} \frac{1}{t} dt && (2.10)\\
&\overset{(iv)}{=} \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \Gamma\left(0, \frac{b}{\mu_\xi}\right) && (2.11)
\end{aligned}$$

where the step (i) is because ξ/b is exponentially distributed with mean μ_ξ/b, the steps (ii), (iii) are by change of variables, and the step (iv) is by the definition of incomplete gamma function. Also $e^x\Gamma(0,x) \le \ln\left(1+\frac{1}{x}\right)$ is obtained after observing the connection with the exponential integral $E_1(x)$ as $\Gamma(0,x) = E_1(x)$ and using the inequalities from [AS64] for $E_1(x)$. This yields $(b/\mu_\xi)\, e^{b/\mu_\xi}\Gamma(0,b/\mu_\xi) \le (b/\mu_\xi)\ln(1+\mu_\xi/b)$. Also $(b/\mu_\xi)\ln(1+\mu_\xi/b) < 1$ because $0 < x\ln(1+1/x) < 1$ for $x > 0$ $\square$

Fact 2.4. *Let H be an isotropically distributed matrix and Φ be a unitary matrix distributed according to any distribution independent of H, then $H, \Phi H, H\Phi$ all have the same distribution. Moreover $\Phi H, H\Phi$ are independent of Φ. See [MH99] for details.*

Lemma 2.1. *Let $[\xi_1, \xi_2, \ldots, \xi_n]$ be an arbitrary complex random vector and Q be an $n \times n$ isotropically distributed unitary matrix independent of ξ_i, then* $h([\xi_1, \xi_2, \ldots, \xi_n]\, Q) = h\left(\sum |\xi_i|^2\right) + (n-1)\,\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log\left(\frac{\pi^n}{\Gamma(n)}\right)$

Proof idea. This is proved by using the fact that in radial coordinates, the distribution of $[\xi_1, \xi_2, \ldots, \xi_n]\, Q$ will be dependent only on the radius. See Appendix A.1 for more details. Note that we can use this Lemma also on $h\left(\xi_1 \overline{q_1}^{(T)}\right)$ with an isotropically distributed unit vector $\overline{q_1}^{(T)}$ by considering the equality $h\left(\xi_1 \overline{q_1}^{(T)}\right) = h([\xi_1, 0, .., 0]\, Q)$, where the isotropically distributed unit vector $\overline{q_1}^{(T)}$ can be taken as the first row of an isotropically distributed unitary matrix Q. $\square$

Corollary 2.9. *Let $[\xi_1, \xi_2, \ldots, \xi_n]$ be an arbitrary complex random vector, ξ be an arbitrary complex random variable and Q be an $n \times n$ isotropically distributed unitary matrix independent of ξ, ξ_i, then* $h([\xi_1, \xi_2, \ldots, \xi_n]\, Q |\, \xi) = h\left(\sum |\xi_i|^2 \middle|\, \xi\right) + (n-1)\,\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log\left(\frac{\pi^n}{\Gamma(n)}\right)$

Proof. This can be proved similar to the previous lemma since the distribution of $h([\xi_1, \xi_2, \ldots, \xi_n]\, Q |\, \xi)$ will be dependent only on the radius. We can use this corollary also on $h\left(\xi_1 \overline{q_1}^{(T)} \middle|\, \xi\right)$, similar to the previous Lemma. $\square$

2.4.2 Properties of transmitted signals that achieve capacity

We now establish the properties of capacity achieving distribution for the noncoherent MIMO with asymmetric statistics. We have our channel model $Y = GX + W$. Now for any $T \times T$ unitary matrix Φ we have $Y\Phi^\dagger = GX\Phi^\dagger + W\Phi^\dagger$. Since w_{ij} are i.i.d. $\mathcal{CN}(0,1)$, $W\Phi^\dagger$ and W have the same distribution, and hence

$$p\left(Y\Phi^\dagger | X\Phi^\dagger\right) = p\left(Y|X\right). \tag{2.12}$$

Now

$$C = \sup_{p(X)} I\left(X;Y\right) \tag{2.13}$$

subject to the average power constraint (2.2) and we have

$$\begin{aligned} I\left(X;Y\right) &= \mathbb{E}\left[\log\left(\frac{p\left(Y|X\right)}{p\left(Y\right)}\right)\right] \\ &= \int dX p\left(X\right) \int dY p\left(Y|X\right) \log\left(\frac{p\left(Y|X\right)}{\int d\tilde{X} p\left(\tilde{X}\right) p\left(Y|\tilde{X}\right)}\right). \end{aligned} \tag{2.14}$$

Lemma 2.2. *(Invariance of $I\left(X;Y\right)$ to post-rotations of X): Suppose that X has a probability density $p_0\left(X\right)$ that generates some mutual information I_0. Then, for any unitary matrix Φ, the "post-rotated" probability density, $p_1\left(X\right) = p_0\left(X\Phi^\dagger\right)$ also generates I_0.*

Proof idea. This is an adaptation of the existing results for MIMO from [MH99, Lemma 1]. The proof proceeds by substituting the post-rotated density $p_1\left(X\right)$ into (2.14), changing the variables of integration and using $p\left(Y\Phi|X\Phi\right) = p\left(Y|X\right)$ from (2.12). □

Lemma 2.3. *The signal of the form $X = LQ$ with L being a lower triangular matrix and Q being an isotropically distributed unitary matrix independent of L, achieves the capacity of the noncoherent MIMO.*

Proof. Let X be a capacity achieving random variable and I_0 be the corresponding mutual information achieved. Now X can be decomposed as $X = L\Phi'$ using the LQ decomposition with L upper diagonal and Φ' unitary, but they could be jointly distributed and Φ' may not be isotropically unitary distributed. Let Θ be an isotropically distributed unitary matrix that is independent of L and Φ'. Now use $X_1 = X\Theta$ for signaling and let Y be the corresponding received signal. Then

$$I\left(X_1; Y|\Theta\right) = I\left(X\Theta; Y|\,\Theta\right) \tag{2.15}$$
$$= I_0 \tag{2.16}$$

using Lemma B.1. Now

$$I\left(X_1; Y\right) + I\left(\Theta; Y|X_1\right) = I\left(\Theta; Y\right) + I\left(X_1; Y|\Theta\right) \tag{2.17}$$
$$I\left(X_1; Y\right) + 0 \overset{(i)}{=} I\left(\Theta; Y\right) + I\left(X_1; Y|\Theta\right) \tag{2.18}$$
$$I\left(X_1; Y\right) \overset{(ii)}{\geq} I\left(X_1; Y|\Theta\right) \tag{2.19}$$
$$= I_0, \tag{2.20}$$

where (i) was because $I\left(\Theta; Y|X_1\right) = 0$ since $\Theta - X_1 - Y$ is a Markov chain and (ii) was because $I\left(X_1; Y|\Theta\right) \geq 0$. Hence without loss of generality the signal of the form $LQ = L\Phi'\Theta$ with $Q = \Phi'\Theta$ achieves the capacity. Now $Q = \Phi'\Theta$ is also unitary isotropically distributed and independent of Φ' using Fact 2.4 on page 25. $\square$

Next, we focus our attention on computing $h\left(Y|X\right)$, which will be necessary in future derivations. Let $Y(n)$ be the n^{th} row of Y. Conditioned on X, the rows of Y are independent Gaussian. Hence:

$$h\left(Y|X\right) = \sum_{n=1}^{N} h\left(Y\left(n\right)|X\right). \tag{2.21}$$

With $\underline{\rho}^2\left(n\right)$ being the vector of channel strengths to n^{th}receiver antenna, we have:

$$\begin{aligned} K_{Y(n)|X} &= \mathbb{E}\left[Q^\dagger L^\dagger g^\dagger\left(n\right) g\left(n\right) LQ\middle|\, LQ\right] + I_T \\ &= Q^\dagger L^\dagger \mathbb{E}\left[g^\dagger\left(n\right) g\left(n\right)\right] LQ + I_T \\ &= Q^\dagger L^\dagger \operatorname{diag}\left(\underline{\rho}^2\left(n\right)\right) LQ + I_T \end{aligned}$$

where I_T is a $T \times T$ identity matrix and $\text{diag}\left(\underline{\rho}^2(n)\right)$ is the diagonal matrix formed from $\underline{\rho}^2(n)$. Hence:

$$
\begin{aligned}
& h\left(Y(n)\,|X\right) \\
&= \mathbb{E}\left[\log\left(\det\left(\pi e K_{Y(n)|X}\right)\right)\right] && (2.22)\\
&= \mathbb{E}\left[\log\left(\det\left(\pi e\left(Q^{\dagger}L^{\dagger}\text{diag}\left(\underline{\rho}^2(n)\right)LQ + I_T\right)\right)\right)\right] && (2.23)\\
&\overset{(i)}{=} \mathbb{E}\left[\log\left(\det\left(\pi e\left(L^{\dagger}\text{diag}\left(\underline{\rho}^2(n)\right)L + I_T\right)\right)\right)\right] && (2.24)
\end{aligned}
$$

where (i) uses the property of determinants to cancel Q and $Q^{\dagger}$. Also, for $T \geq M$, using the lower triangular structure of L with $L_{M\times M}$ being the first $M \times M$ submatrix of L (rest of the elements of L are zero for $T \geq M$) we have:

$$
\begin{aligned}
& h\left(Y(n)\,|X\right) \\
&= \mathbb{E}\left[\log\left(\det\left(\left(L^{\dagger}_{M\times M}\text{diag}\left(\underline{\rho}^2(n)\right)L_{M\times M} + I_M\right)\right)\right)\right] + (T)\log(\pi e). && (2.25)
\end{aligned}
$$

2.4.3 Outer bound for the $M \times 1$ MISO system

We now prove the gDoF outer bound given in Theorem 3.10 for the $M \times 1$ MISO system.We assume that $T > 1$, since for $T = 1$ we have the desired result using Theorem 2.5 on page 19. Also we assume that $T \geq M$ in the following outer bound computations. The case for $T < M$ is easily derived following similar steps reaching the same result and is given in Appendix A.6.

For the capacity achieving distribution, we have the structure $X = \begin{bmatrix} L_{M\times M} & 0_{M\times(T-M)} \end{bmatrix} Q$ (from Theorem 2.1), where

$$
L_{M\times M} = \begin{bmatrix} x_{11} & 0 & 0 & \\ . & . & 0 & 0 \\ . & & . & 0 \\ x_{M1} & . & . & x_{MM} \end{bmatrix}
$$

and $0_{M\times(T-M)}$ is an $M \times (T-M)$ matrix with elements of value zero. For the MISO, we have $Y = GX + W$ with $G = \begin{bmatrix} g_{11} & . & . & g_{1M} \end{bmatrix}$ and W is $1 \times T$ with i.i.d. $\mathcal{CN}(0,1)$

components. We assume $\rho_{11}^2 \geq \rho_{1i}^2$ without loss of generality. Now note that WQ has also same distribution as W and is independent of Q (using the fact that W is isotropically distributed and Fact 2.4). Hence

$$\begin{aligned} h(Y) &= h\left((GX+W)Q\right) \\ &= h\Bigg(\Bigg[\ \left(w_{11} + \textstyle\sum_{i=1}^{M} x_{i1} g_{1i}\right), \quad \left(w_{12} + \textstyle\sum_{i=2}^{M} x_{i2} g_{1i}\right), \ldots \\ &\qquad \ldots, \ \left(w_{1M} + \textstyle\sum_{i=M}^{M} x_{i2} g_{1i}\right), \quad w_{1(M+1)} \quad \ldots, \quad w_{1T}\ \Bigg] Q\Bigg). \end{aligned}$$

Now using Lemma 3.2 on page 66, we have

$$\begin{aligned} h(Y) = h&\left(\sum_{j=1}^{M}\left|w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i}\right|^2 + \sum_{i=M+1}^{T} |w_{1i}|^2\right) \\ &+ (T-1)\,\mathbb{E}\left[\log\left(\sum_{j=1}^{M}\left|w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i}\right|^2 + \sum_{i=M+1}^{T} |w_{1i}|^2\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right) \end{aligned} \tag{2.26}$$

$$\begin{aligned} \overset{(i)}{\leq} h&\left(\sum_{j=1}^{M}\left|w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i}\right|^2 + \sum_{i=M+1}^{T} |w_{1i}|^2\right) \\ &+ (T-1)\,\mathbb{E}\left[\log\left(\sum_{i=1}^{M} \rho_{1i}^2 \left(\sum_{j=1}^{i} |x_{ij}|^2\right) + T - M\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right), \end{aligned} \tag{2.27}$$

where (i) was using Tower property of expectation, Jensen's inequality and $\sum_{j=1}^{M}\sum_{i=j}^{M} |x_{ij}|^2 \rho_{1i}^2 = \sum_{i=1}^{M}\sum_{j=1}^{i} |x_{ij}|^2 \rho_{1i}^2$. Now using (3.52) we have

$$\begin{aligned} h(Y|X) &= \mathbb{E}\left[\log\left(\det\left(L_{M\times M}^{\dagger}\operatorname{diag}\ \left(\rho_{11}^2, \ldots, \rho_{1M}^2\right) L_{M\times M} + I_M\right)\right)\right] \\ &\qquad + (T)\log(\pi e) \end{aligned} \tag{2.28}$$

$$= \mathbb{E}\left[\log\left(\prod_{i=1}^{M}(1+\omega_i)\right)\right] + (T)\log(\pi e) \tag{2.29}$$

where ω_i are the eigenvalues of $L_{M\times M}^{\dagger}\operatorname{diag}\ \left(\rho_{11}^2, \ldots, \rho_{1M}^2\right) L_{M\times M}$. The eigenvalues are non-negative since the matrix is Hermitian. Hence

$$h(Y|X) = \mathbb{E}\left[\log\left(\prod_{i=1}^{M}(1+\omega_i)\right)\right] + (T)\log(\pi e) \tag{2.30}$$

$$\geq \mathbb{E}\left[\log\left(1+\sum\omega_i\right)\right]+(T)\log(\pi e) \tag{2.31}$$

the last step is true because $\omega_i \geq 0$. Now

$$\sum\omega_i = \text{Trace}\left(L^\dagger_{M\times M}\text{diag }\left(\rho^2_{11},\ldots,\rho^2_{1M}\right)L_{M\times M}\right) \tag{2.32}$$

$$= \text{Trace}\left(\text{diag }\left(\rho^2_{11},\ldots,\rho^2_{1M}\right)L_{M\times M}L^\dagger_{M\times M}\right) \tag{2.33}$$

$$= \sum_{i=1}^{M}\rho^2_{1i}\left(\sum_{j=1}^{i}|x_{ij}|^2\right). \tag{2.34}$$

Hence

$$h(Y|X) \geq \mathbb{E}\left[\log\left(1+\sum_{i=1}^{M}\rho^2_{1i}\left(\sum_{j=1}^{i}|x_{ij}|^2\right)\right)\right]+(T)\log(\pi e). \tag{2.35}$$

Hence

$$\begin{aligned} I(X;Y) \leq\ & h\left(\sum_{j=1}^{M}\left|w_{1j}+\sum_{i=j}^{M}x_{ij}g_{1i}\right|^2+\sum_{i=M+1}^{T}|w_{1i}|^2\right) \\ & +(T-1)\,\mathbb{E}\left[\log\left(\sum_{i=1}^{M}\rho^2_{1i}\left(\sum_{j=1}^{i}|x_{ij}|^2\right)+T-M\right)\right] \\ & -\mathbb{E}\left[\log\left(1+\sum_{i=1}^{M}\rho^2_{1i}\left(\sum_{j=1}^{i}|x_{ij}|^2\right)\right)\right] \\ & +\log\left(\frac{\pi^T}{\Gamma(T)}\right)-T\log(\pi e) \end{aligned} \tag{2.36}$$

$$\dot{\leq}\ (T-1)\log\left(\sum_{i=1}^{M}\rho^2_{1i}MT+T\right), \tag{2.37}$$

where the last step was using maximum entropy results and Jensen's inequality. Hence

$$\limsup_{\mathsf{SNR}\to\infty}\frac{1}{T}\frac{I(X;Y)}{\log(\mathsf{SNR})} \leq \left(1-\frac{1}{T}\right)\gamma_{11}. \tag{2.38}$$

2.4.4 Outer bound for the 2×2 MIMO system

In this subsection, we prove the gDoF outer bound from Theorem 2.7 for the 2×2 MIMO with exponents γ_D in the direct links and γ_{CL} in the crosslinks. We have the structure of the optimal distribution as

$$X = \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q$$

from Theorem 2.1 and we have

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix},$$

and $Y = GX + W$, where W is $2 \times T$ vector with i.i.d. $\mathcal{CN}(0,1)$ components. We assume $T \geq 2$, since for $T = 1$ the gDoF is zero due to Theorem 2.5. We have:

$$h(Y) = h\left(G \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} Q + W\right) \tag{2.39}$$

$$\stackrel{(i)}{=} h\left(\left(G \begin{bmatrix} a & 0 & 0 & . & . & 0 \\ b & c & 0 & . & . & 0 \end{bmatrix} + W\right) Q\right) \tag{2.40}$$

$$= h\left(\begin{bmatrix} ag_{11} + bg_{12} + w_{11} & cg_{12} + w_{12} & w_{13} & . & w_{1T} \\ ag_{21} + bg_{22} + w_{21} & cg_{22} + w_{22} & w_{23} & . & w_{2T} \end{bmatrix} Q\right) \tag{2.41}$$

$$\stackrel{(ii)}{=} h\left(\begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} \Phi Q\right) \tag{2.42}$$

$$\stackrel{(iii)}{=} h\left(\begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} Q\right), \tag{2.43}$$

where the step (i) used the fact that W and WQ have the same distribution and is independent of Q. In step (ii), ξ_{ij} arise after LQ transformation (using Gram-Schmidt process)

$$\begin{bmatrix} ag_{11} + bg_{12} + w_{11} & cg_{12} + w_{12} & w_{13} & . & w_{1T} \\ ag_{21} + bg_{22} + w_{21} & cg_{22} + w_{22} & w_{23} & . & w_{2T} \end{bmatrix} = \begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} \Phi$$

$$|\xi_{11}|^2 = |ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \tag{2.44}$$

$$|\xi_{21}|^2 = \frac{\left|(ag_{21} + bg_{22} + w_{21})(ag_{11} + bg_{12} + w_{11})^* + (cg_{22} + w_{22})(cg_{12} + w_{12})^* + \sum_{i=3}^{T} w_{2i} w_{1i}^*\right|^2}{|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2} \tag{2.45}$$

$$
\begin{aligned}
\left|\xi_{22}\right|^2 = &\left|ag_{21}+bg_{22}+w_{21}\right|^2 + \left|cg_{22}+w_{22}\right|^2 + \sum_{i=3}^{T}\left|w_{2i}\right|^2 \\
&- \frac{\left|\left(ag_{21}+bg_{22}+w_{21}\right)\left(ag_{11}+bg_{12}+w_{11}\right)^* + \left(cg_{22}+w_{22}\right)\left(cg_{12}+w_{12}\right)^* + \sum_{i=3}^{T} w_{2i}w_{1i}^*\right|^2}{\left|ag_{11}+bg_{12}+w_{11}\right|^2 + \left|cg_{12}+w_{12}\right|^2 + \sum_{i=3}^{T}\left|w_{1i}\right|^2}
\end{aligned}
\tag{2.46}
$$

where Φ is unitary. In step (iii) we absorb Φ onto Q using Fact 2.4. The Gram-Schmidt process for LQ transformation yields ξ_{ij} as given in (2.44), (2.45) and (2.46).

Also using (3.52), (3.50) we get

$$
\begin{aligned}
&h\left(Y|X\right) \\
&= \mathbb{E}\left[\log\left(|a|^2\rho_{11}^2 + |b|^2\rho_{12}^2 + |c|^2\rho_{12}^2 + |a|^2|c|^2\rho_{11}^2\rho_{12}^2 + 1\right)\right] \\
&\quad + \mathbb{E}\left[\log\left(|a|^2\rho_{21}^2 + |b|^2\rho_{22}^2 + |c|^2\rho_{22}^2 + |a|^2|c|^2\rho_{21}^2\rho_{22}^2 + 1\right)\right] \\
&\quad + 2T\log\left(\pi e\right)
\end{aligned}
\tag{2.47}
$$

For computing $h\left(Y\right)$, let $\overline{q_1}^{(T)}, \overline{q_2}^{(T)}$ be the first two rows of Q. The vectors $\overline{q_1}^{(T)}, \overline{q_2}^{(T)}$ are orthogonal since Q is isotropic unitary. We have

$$
h\left(Y\right) = h\left(\begin{bmatrix} \xi_{11} & 0 & . & . & . & 0 \\ \xi_{21} & \xi_{22} & 0 & . & . & 0 \end{bmatrix} Q\right) \tag{2.48}
$$

$$
= h\left(\xi_{11}\overline{q_1}^{(T)}\right) + h\left(\xi_{21}\overline{q_1}^{(T)} + \xi_{22}\overline{q_2}^{(T)}\middle|\, \xi_{11}\overline{q_1}^{(T)}\right) \tag{2.49}
$$

Now consider $h\left(\xi_{21}\overline{q_1}^{(T)} + \xi_{22}\overline{q_2}^{(T)}\middle|\, \xi_{11}\overline{q_1}^{(T)}\right)$. Since ξ_{11} is nonnegative and $\xi_{11}\overline{q_1}^{(T)}$ is given in the conditioning, the direction $\overline{q_1}^{(T)}$ is known in the conditioning. Hence considering $\xi_{21}\overline{q_1}^{(T)} + \xi_{22}\overline{q_2}^{(T)}$ in a new orthonormal basis with the first basis vector chosen as $\overline{q_1}^{(T)}$ and the rest of the basis vectors chosen arbitrarily, the projection of $\xi_{21}\overline{q_1}^{(T)} + \xi_{22}\overline{q_2}^{(T)}$ onto the first basis vector is ξ_{21}. The projection onto the rest of the $T-1$ vectors forms $\xi_{22}\overline{q_2}^{(T-1)}$ where $\overline{q_1}^{(T-1)}$ is a $T-1$ dimensional isotropically distributed unit vector. Hence

$$
h\left(\xi_{21}\overline{q_1}^{(T)} + \xi_{22}\overline{q_2}^{(T)}\middle|\, \xi_{11}\overline{q_1}^{(T)}\right) = h\left(\left[\xi_{21}, \xi_{22}\overline{q}_2^{(T-1)}\right]\middle|\, \xi_{11}, \overline{q_1}^{(T)}\right) \tag{2.50}
$$

$$
= h\left(\left[\xi_{21}, \xi_{22}\overline{q}_2^{(T-1)}\right]\middle|\, \xi_{11}\right) \tag{2.51}
$$

and

$$h(Y) = h\left(\xi_{11}\overline{q_1}^{(T)}\right) + h\left(\left[\xi_{21}, \xi_{22}\overline{q}_2^{(T-1)}\right]\Big|\, \xi_{11}\right) \tag{2.52}$$

$$\stackrel{(i)}{=} h\left(\xi_{11}\overline{q_1}^{(T)}\right) + h\left(\left[\xi_{21}, \xi_{22}\overline{q}_2^{(T-1)}\right]\Big|\, |\xi_{11}|^2\right) \tag{2.53}$$

$$\leq h\left(\xi_{11}\overline{q_1}^{(T)}\right) + h\left(\xi_{22}\overline{q}_2^{(T-1)}\Big|\, |\xi_{11}|^2\right) + h\left(\xi_{21}|\, |\xi_{11}|^2\right) \tag{2.54}$$

where (i) is because ξ_{11} is non-negative. Note that above equation contains $\xi_{11}, \xi_{22}, \xi_{21}$ which we would like to convert to the form $|\xi_{11}|^2, |\xi_{22}|^2, |\xi_{21}|^2$ which are available from (2.44), (2.45) and (2.46). We handle $h\left(\xi_{21}|\, |\xi_{11}|^2\right)$ with the following claim.

Claim 2.1. $h\left(\xi_{21}|\, |\xi_{11}|^2\right) \leq h\left(|\xi_{21}|^2\big|\, |\xi_{11}|^2\right) + \log(\pi)$

Proof. We have

$$h\left(\xi_{21}|\, |\xi_{11}|^2\right) \stackrel{(i)}{=} h\left(\xi_{21}e^{i\theta}\big|\, |\xi_{11}|^2, \theta\right) \tag{2.55}$$

$$\stackrel{(ii)}{\leq} h\left(\xi_{21}e^{i\theta}\big|\, |\xi_{11}|^2\right) \tag{2.56}$$

$$\stackrel{(iii)}{=} h\left(|\xi_{21}|^2\big|\, |\xi_{11}|^2\right) + \log(\pi) \tag{2.57}$$

where (i) uses an independent $\theta \sim \text{Unif}\,[0, 2\pi]$, (ii) is because conditioning reduces entropy, (iii) is using Lemma (3.2) since given $|\xi_{11}|^2$, $\xi_{21}e^{i\theta}$ is isotropically distributed. $\square$

Using the above Claim, we get

$$h(Y) \leq h\left(\xi_{11}\overline{q_1}^{(T)}\right) + h\left(\xi_{22}\overline{q}_2^{(T-1)}\Big|\, |\xi_{11}|^2\right) + h\left(|\xi_{21}|^2\big|\, |\xi_{11}|^2\right) + \log(\pi) \tag{2.58}$$

$$\begin{aligned} &\stackrel{(i)}{\leq} h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right) \\ &\quad + (T-1)\,\mathbb{E}\left[\log\left(|a|^2\rho_{11}^2 + \left(|b|^2 + |c|^2\right)\rho_{12}^2 + 1\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right) \\ &\quad + h\left(|\xi_{21}|^2\big|\, |\xi_{11}|^2\right) + h\left(|\xi_{22}|^2\big|\, |\xi_{11}|^2\right) + (T-2)\,\mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right] + \log(\pi) \end{aligned} \tag{2.59}$$

where (i) is by applying Lemma 3.2 on $h\left(\xi_{11}\overline{q_1}^{(T)}\right)$ and Corollary 3.8 on $h\left(\xi_{22}\overline{q}_2^{(T-1)}\Big|\, |\xi_{11}|^2\right)$. Now we use the following Lemma to simplify $h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)$ from the previous expression.

Lemma 2.4. *For any given distribution on* (a, b, c),

$$h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)$$

and

$$\mathbb{E}\left[\log\left(|a|^2 \rho_{11}^2 + \left(|b|^2 + |c|^2\right) \rho_{12}^2 + 1\right)\right]$$

have the same gDoF. Similarly for any given distribution on (a, b, c),

$$h\left(|ag_{21} + bg_{22} + w_{21}|^2 + |cg_{22} + w_{22}|^2 + \sum_{i=3}^{T} |w_{2i}|^2\right)$$

and

$$\mathbb{E}\left[\log\left(|a|^2 \rho_{21}^2 + \left(|b|^2 + |c|^2\right) \rho_{22}^2 + 1\right)\right]$$

have the same gDoF.

Proof. The proof proceeds by constructing a noncoherent channel

$$\mathcal{C}_1 : V = |ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \tag{2.60}$$

with inputs a, b, c output V and showing that it does not have any gDoF. The proof uses outer bounding techniques from [LM03]. See Appendix A.7 for details. □

Hence using the previous lemma, we get

$$\begin{aligned} h(Y) \dot{\leq} T\mathbb{E}&\left[\log\left(|a|^2 \rho_{11}^2 + \left(|b|^2 + |c|^2\right) \rho_{12}^2 + 1\right)\right] \\ &+ h\left(|\xi_{21}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2\right) + h\left(|\xi_{22}|^2 |\xi_{11}|^2 \middle| |\xi_{11}|^2\right) \\ &+ (T-2)\,\mathbb{E}\left[\log\left(|\xi_{22}|^2 |\xi_{11}|^2\right)\right] - T\mathbb{E}\left[\log\left(|\xi_{11}|^2\right)\right]. \end{aligned} \tag{2.61}$$

Now we simplify $\mathbb{E}\left[\log\left(|\xi_{11}|^2\right)\right]$ from the previous expression.

$$\begin{aligned} \mathbb{E}\left[\log\left(|\xi_{11}|^2\right)\right] &= \mathbb{E}\left[\log\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\log\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right) \middle| a, b, c\right]\right] \end{aligned}$$

$$\overset{(i)}{\doteq} \mathbb{E}\left[\log\left(|a|^2\rho_{11}^2 + \left(|b|^2+|c|^2\right)\rho_{12}^2 + 1\right)\right], \tag{2.62}$$

where (i) was using Fact 5.1 on page 135 and that $|ag_{11}+bg_{12}+w_{11}|^2, |cg_{12}+w_{12}|^2, |w_{1i}|^2$ are exponentially distributed given a, b, c. Hence

$$\begin{aligned} h(Y) \dot{\leq} h\left(|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right) + h\left(|\xi_{22}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right) \\ + (T-2)\,\mathbb{E}\left[\log\left(|\xi_{22}|^2|\xi_{11}|^2\right)\right] \end{aligned} \tag{2.63}$$

Now we use the following lemmas to further simplify the terms in the above expression for $h(Y)$.

Lemma 2.5. *We have*

$$h\left(|\xi_{22}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right) \doteq h\left(|\xi_{22}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2, a, b, c\right) \leq \mathbb{E}\left[\log\left(e\mathbb{E}\left[|\xi_{22}|^2|\xi_{11}|^2 \,\middle|\, a, b, c\right]\right)\right].$$

Proof. The proof uses similar techniques as that for Lemma 3.3. See Appendix A.8 for details. □

Lemma 2.6. *We have*

$$h\left(|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right) \doteq h\left(|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2, a, b, c\right) \leq \mathbb{E}\left[\log\left(e\mathbb{E}\left[|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, a, b, c\right]\right)\right].$$

Proof. This can be proved similar to the previous lemma. We omit the proof. □

Hence using (2.44), (2.45) and Lemma 2.6 to bound $h\left(|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right)$, we have

$$|\xi_{21}|^2|\xi_{11}|^2 = \left|(ag_{21}+bg_{22}+w_{21})(ag_{11}+bg_{12}+w_{11})^* + (cg_{22}+w_{22})(cg_{12}+w_{12})^* + \sum_{i=3}^{T} w_{2i}w_{1i}^*\right|^2$$

,

$$h\left(|\xi_{21}|^2|\xi_{11}|^2 \,\middle|\, |\xi_{11}|^2\right)$$

$$
\begin{aligned}
&\dot{\leq} \mathbb{E}\left[\log\left(\left(|a|^2\rho_{11}^2+|b|^2\rho_{12}^2+1\right)\left(|a|^2\rho_{21}^2+|b|^2\rho_{22}^2+1\right)\right.\right.\\
&\qquad\left.\left.+2|c|^2|b|^2\rho_{22}^2\rho_{12}^2+\left(|c|^2\rho_{12}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)+T-2\right)\right] \qquad (2.64)\\
&\dot{\leq} \mathbb{E}\left[\log\left(\left(|a|^2\rho_{11}^2+|b|^2\rho_{12}^2+1\right)\left(|a|^2\rho_{21}^2+|b|^2\rho_{22}^2+1\right)\right.\right.\\
&\qquad\left.\left.+\left(|c|^2\rho_{12}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)\right)\right], \qquad (2.65)
\end{aligned}
$$

where the last step followed due to AM-GM inequality $2|c|^2|b|^2\rho_{22}^2\rho_{12}^2 \leq |b|^4\rho_{22}^2\rho_{12}^2+|c|^4\rho_{22}^2\rho_{12}^2$. Similarly using (2.44) and (2.46) we have

$$
\begin{aligned}
&|\xi_{22}|^2|\xi_{11}|^2\\
&=\left(|ag_{21}+bg_{22}+w_{21}|^2+|cg_{22}+w_{22}|^2+\sum_{i=3}^{T}|w_{2i}|^2\right)\left(|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right)\\
&\quad-\left|(ag_{21}+bg_{22}+w_{21})(ag_{11}+bg_{12}+w_{11})^*+(cg_{22}+w_{22})(cg_{12}+w_{12})^*+\sum_{i=3}^{T}w_{2i}w_{1i}^*\right|^2.
\end{aligned}
$$

After some algebraic manipulations, it can be seen that

$$
\begin{aligned}
&\mathbb{E}\left[|\xi_{22}|^2|\xi_{11}|^2\,\middle|\,a,b,c\right]\\
&=(T-2)^2-(T-2)+(T-2)\left(|a|^2\rho_{21}^2+|b|^2\rho_{22}^2+|c|^2\rho_{22}^2+|a|^2\rho_{11}^2+|b|^2\rho_{12}^2+|c|^2\rho_{12}^2+2\right)\\
&\quad+\left(|a|^2\rho_{11}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)+|b|^2\rho_{12}^2+\left(|a|^2\rho_{21}^2+1\right)\left(|c|^2\rho_{12}^2+1\right)+|b|^2\rho_{22}^2. \qquad (2.66)
\end{aligned}
$$

After retaining only the terms that contribute to gDoF from the above equation, we bound $h\left(|\xi_{22}|^2|\xi_{11}|^2\,\middle|\,|\xi_{11}|^2\right)$ using Lemma 2.6 to get

$$
\begin{aligned}
&h\left(|\xi_{22}|^2|\xi_{11}|^2\,\middle|\,|\xi_{11}|^2\right)\\
&\dot{\leq}\mathbb{E}\left[\log\left(e\mathbb{E}\left[|\xi_{22}|^2|\xi_{11}|^2\,\middle|\,a,b,c\right]\right)\right] \qquad (2.67)\\
&\dot{\leq}\mathbb{E}\left[\log\left(\left(|a|^2\rho_{11}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)+|b|^2\left(\rho_{12}^2+\rho_{22}^2\right)\right.\right.\\
&\qquad\left.\left.+\left(|a|^2\rho_{21}^2+1\right)\left(|c|^2\rho_{12}^2+1\right)\right)\right]. \qquad (2.68)
\end{aligned}
$$

Also

$$
\mathbb{E}\left[\log\left(|\xi_{22}|^2|\xi_{11}|^2\right)\right]
$$

$$= \mathbb{E}\left[\log\left(\mathbb{E}\left[|\xi_{22}|^2 |\xi_{11}|^2 \middle| a, b, c\right]\right)\right] \tag{2.69}$$

$$\begin{aligned} &\dot{\leq} \mathbb{E}\Big[\log\Big(\left(|a|^2 \rho_{11}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right) + |b|^2 \left(\rho_{12}^2 + \rho_{22}^2\right) \\ &\qquad + \left(|a|^2 \rho_{21}^2 + 1\right)\left(|c|^2 \rho_{12}^2 + 1\right)\Big)\Big]. \end{aligned} \tag{2.70}$$

Hence using (2.70), (2.68), (2.65) in (2.63) we get

$$\begin{aligned} &h(Y) \\ &\dot{\leq} \mathbb{E}\left[\log\left(\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + 1\right)\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + 1\right) + \left(|c|^2 \rho_{12}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right)\right)\right] \\ &\quad + (T-1)\,\mathbb{E}\Big[\log\Big(\left(|a|^2 \rho_{11}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right) + |b|^2 \left(\rho_{12}^2 + \rho_{22}^2\right) \\ &\qquad + \left(|a|^2 \rho_{21}^2 + 1\right)\left(|c|^2 \rho_{12}^2 + 1\right)\Big)\Big]. \end{aligned} \tag{2.71}$$

Using the above equation and (2.47), we get

$$\begin{aligned} &I(X;Y) \\ &\dot{\leq} \mathbb{E}\left[\log\left(\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + 1\right)\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + 1\right) + \left(|c|^2 \rho_{12}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right)\right)\right] \\ &\quad + (T-1)\,\mathbb{E}\left[\log\left(\left(|a|^2 \rho_{11}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right) + |b|^2 \left(\rho_{12}^2 + \rho_{22}^2\right) + \left(|a|^2 \rho_{21}^2 + 1\right)\left(|c|^2 \rho_{12}^2 + 1\right)\right)\right] \\ &\quad - \mathbb{E}\left[\log\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + |c|^2 \rho_{12}^2 + |a|^2 |c|^2 \rho_{11}^2 \rho_{12}^2 + 1\right)\right] \\ &\quad - \mathbb{E}\left[\log\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + |c|^2 \rho_{22}^2 + |a|^2 |c|^2 \rho_{21}^2 \rho_{22}^2 + 1\right)\right]. \end{aligned} \tag{2.72}$$

Hence a gDoF equivalent outer bound is given by the optimization problem

$$\mathcal{P}_1 : \begin{cases} \underset{\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T}{\text{maximize}} \mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right], \end{cases} \tag{2.73}$$

where

$$\begin{aligned} &f\left(|a|^2, |b|^2, |c|^2\right) \\ &= \log\left(\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + 1\right)\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + 1\right) + \left(|c|^2 \rho_{12}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right)\right) \\ &\quad + (T-1)\log\left(\left(|a|^2 \rho_{11}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right) + |b|^2 \left(\rho_{12}^2 + \rho_{22}^2\right) + \left(|a|^2 \rho_{21}^2 + 1\right)\left(|c|^2 \rho_{12}^2 + 1\right)\right) \\ &\quad - \log\left(\left(1 + |a|^2 \rho_{11}^2\right)\left(1 + |c|^2 \rho_{12}^2\right) + |b|^2 \rho_{12}^2\right) \\ &\quad - \log\left(\left(1 + |a|^2 \rho_{21}^2\right)\left(1 + |c|^2 \rho_{22}^2\right) + |b|^2 \rho_{22}^2\right). \end{aligned} \tag{2.74}$$

Lemma 2.7. *The gDoF achieved in $\mathcal{P}_1$ can be achieved by a point mass distribution, i.e., gDoF$(\mathcal{P}_1)$ = gDoF$(\mathcal{P}_7)$, where $\mathcal{P}_7$ is the following:*

$$\mathcal{P}_7 : \begin{cases} \textit{maximize } f\left(|a|^2, |b|^2, |c|^2\right) \quad \textit{with} \\ |a|^2 \leq T, |b|^2 \leq T, |c|^2 \leq T. \end{cases} \tag{2.75}$$

Proof idea. The proof proceeds in several steps:

Step 1: Show that there exists a discretization (over an infinite set) for any distribution of $\left(|a|^2, |b|^2, |c|^2\right)$ that does not incur a loss in gDoF.

Step 2: Show that the discretization can be limited to a finite set without incurring a loss in gDoF.

Step 3: View the problem as a linear program with 2 constraints, and show that the there is an optimal distribution with just 2 mass points.

Step 4: Show that the 2 mass points can be collapsed to a single point using arguments of symmetry.

The details of the proof are given in Appendix A.2. □

Changing the variables from $\left(|a|^2, |b|^2, |c|^2\right)$ to $(\gamma_a, \gamma_b, \gamma_c)$ with the substitution $|a|^2 = \mathsf{SNR}^{-\gamma_a}$, $|b|^2 = \mathsf{SNR}^{-\gamma_b}$, $|c|^2 = \mathsf{SNR}^{-\gamma_c}$, it is clear that

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_7\right) = \left(\mathcal{P}_8\right),$$

where $\mathcal{P}_8$ is the following

$$\mathcal{P}_8 : \begin{cases} \text{maximize} f_\gamma\left(\gamma_a, \gamma_b, \gamma_c\right) \quad \text{with} \\ \gamma_a \geq 0, \gamma_b \geq 0, \gamma_c \geq 0, \end{cases} \tag{2.76}$$

with

$$\begin{aligned} & f_\gamma\left(\gamma_a, \gamma_b, \gamma_c\right) \\ & = \max\Big(\max\left(-\gamma_a + \gamma_{11}, -\gamma_b + \gamma_{12}, 0\right) + \max\left(-\gamma_a + \gamma_{21}, -\gamma_b + \gamma_{22}, 0\right) \\ & \qquad , \max\left(-\gamma_c + \gamma_{12}, 0\right) + \max\left(-\gamma_c + \gamma_{22}, 0\right)\Big) \end{aligned}$$

$$
\begin{aligned}
&+(T-1)\max\Bigg(\max\left(-\gamma_a+\gamma_{11},0\right)+\max\left(-\gamma_c+\gamma_{22},0\right)\\
&\qquad\qquad ,\gamma_b+\max\left(\gamma_{12},\gamma_{22}\right),\max\left(-\gamma_a+\gamma_{21},0\right)+\max\left(-\gamma_c+\gamma_{12},0\right)\Bigg)\\
&-\max\left(-\gamma_a+\gamma_{11},-\gamma_b+\gamma_{12},-\gamma_c+\gamma_{12},-\gamma_a-\gamma_c+\gamma_{11}+\gamma_{12},0\right)\\
&-\max\left(-\gamma_a+\gamma_{21},-\gamma_b+\gamma_{22},-\gamma_c+\gamma_{22},-\gamma_a-\gamma_c+\gamma_{21}+\gamma_{22},0\right). \qquad (2.77)
\end{aligned}
$$

For symmetric 2×2 MIMO, we have $\gamma_{11}=\gamma_{22}=\gamma_D, \gamma_{CL}=\gamma_{12}=\gamma_{21}$. Also it can be assumed without loss of generality that $\gamma_D>\gamma_{CL}$. By inspection of the optimization problem, it is clear that we can also restrict $\gamma_a\leq\gamma_D, \gamma_b\leq\gamma_D, \gamma_c\leq\gamma_D$ without affecting the solution. With these additional constraints, we can simplify $\mathcal{P}_8$ to $\mathcal{P}_9$ for the symmetric 2×2 MIMO with $\mathcal{P}_9$ defined as the following:

$$
\mathcal{P}_9:\begin{cases}
\text{maximize}\max\left(-2\gamma_a+\gamma_{CL},-\gamma_a+\gamma_D-\gamma_b,-2\gamma_b+\gamma_{CL},-2\gamma_c+\gamma_{CL},-\gamma_c\right)\\
\qquad\qquad +(T-1)\max\left(-\gamma_a+\gamma_D-\gamma_c,-\gamma_b\right)+T\gamma_D-t_1-t_2\\
t_1=\max\left(-\gamma_a+\gamma_D,-\gamma_b+\gamma_{CL},-\gamma_a-\gamma_c+\gamma_D+\gamma_{CL}\right)\\
t_2=\max\left(-\gamma_b+\gamma_D,-\gamma_c+\gamma_D,-\gamma_a-\gamma_c+\gamma_D+\gamma_{CL}\right)\\
0\leq\gamma_a\leq\gamma_D, 0\leq\gamma_b\leq\gamma_D, 0\leq\gamma_c\leq\gamma_D.
\end{cases} \qquad (2.78)
$$

Using standard linear programming arguments, $\mathcal{P}_9$ has a solution for $(\gamma_a,\gamma_b,\gamma_c,t_1,t_2)$ in one of the corner points of the following region:

$$
\mathcal{R}:\left\{\begin{array}{r}
0\leq\gamma_a\leq\gamma_D;\ 0\leq\gamma_b\leq\gamma_D, 0\\
0\leq\gamma_c\leq\gamma_D;\\
t_1\geq-\gamma_a+\gamma_D;\ t_1\geq-\gamma_b+\gamma_{CL}\\
t_1\geq-\gamma_a-\gamma_c+\gamma_D+\gamma_{CL};\ t_2\geq-\gamma_b+\gamma_D\\
t_2\geq-\gamma_a-\gamma_c+\gamma_D+\gamma_{CL};\ t_2\geq-\gamma_c+\gamma_D
\end{array}\right\}
$$

This can be seen by considering case by case for $\mathcal{P}_9$, depending on which term inside the $\max(\cdot)$'s could come out in the objective function, and noting that $\underset{\gamma_a,\gamma_b,\gamma_c,t_1,t_2}{\text{maximize}}\max\left(f_1,f_2\right)$ is same as $\max\left(\underset{\gamma_a,\gamma_b,\gamma_c,t_1,t_2}{\text{maximize}}\left(f_1\right),\underset{\gamma_a,\gamma_b,\gamma_c,t_1,t_2}{\text{maximize}}\left(f_2\right)\right)$ for linear f_1,f_2.

Suppose $-2\gamma_a + \gamma_{CL} = \max\left(-2\gamma_a + \gamma_{CL}, -\gamma_a + \gamma_D - \gamma_b, -2\gamma_b + \gamma_{CL}, -2\gamma_c + \gamma_{CL}, -\gamma_c\right)$ and $-\gamma_a + \gamma_D - \gamma_c = \max\left(-\gamma_a + \gamma_D - \gamma_c, -\gamma_b\right)$, then $\mathcal{P}_9$ has a solution in one of the corner points of of $\mathcal{R}$. This is true for all possible cases of the values of the two max ()'s. Hence $\mathcal{P}_9$ itself has a solution in one of the corner points of $\mathcal{R}$.

We code in Mathematica to find all the corner points of $\mathcal{R}$ and find the maximum across the corner points. Finding all the corner points and the calculations could be written down in the chapter, but this would be uninteresting. So we have deferred it to the software. We believe that this does not compromise the analytical rigor of the proof. We obtain the solution in Table 2.3. Our Mathematica code is available online at `http://www.seas.ucla.edu/~joyson/Documents/Sym_mimo_outerbound.nb`. This code uses $\gamma_D = 1, \gamma_{CL} = 1 - \epsilon$ and we can obtain the general solution with a simple scaling.

CHAPTER 3

Noncoherent Diamond network

3.1 Introduction

The capacity of (fading) wireless networks has been unresolved for over four decades. There has been recent progress on this topic through an approximation approach (see [ADT11] and references therein) as well as a scaling approach (see [GK00, OLT07] and references therein). However, most of the work is on understanding the capacity of a *given* wireless network, *i.e.,* where the network as well as its parameters (including channel gains) are known, at least at the destination. There has been much less attention[1] to the case where the network parameters (channel gains) are unknown to everyone, *i.e.,* the noncoherent wireless network capacity. The study of noncoherent point-to-point multiple-input-multiple-output (MIMO) wireless channels in [MH99, ZT02] etc. and references therein, revealed that there was an important tradeoff between communication and channel learning in such scenarios. In particular, it might be useful not to use all the resources available to communicate, if it costs too much to learn their parameters; for example, one would not use all the antennas in noncoherent MIMO channels. The question we ask in this chapter is similar, but in the context of wireless relay networks, in particular we study when one should use training to learn the channels and if so which links to learn and how to use them. The central question examined in this chapter is the generalized degrees of freedom (gDoF) of noncoherent wireless networks (albeit for specific topologies) when there might be significant (known) statistical variations in the link strengths.

[1]Exceptions include [Lap05, ND10, KK13].

The noncoherent wireless model for MIMO, where neither the receiver nor the transmitter knows the fading coefficients was studied by Marzetta and Hochwald [MH99]. In their channel model, the fading gains remain constant within a block of T symbol periods and the fading gains are identically independent distributed (i.i.d.) Rayleigh random variables across the blocks. The general capacity of a noncoherent MIMO is still unknown, but the behavior at high signal-to-noise-ratio (SNR) for the noncoherent MIMO with i.i.d. links is characterized in [ZT02]. There, the idea of communication over a Grassmanian manifold was used to study the capacity behavior at high SNR. The case with unit coherence time ($T = 1$) for the noncoherent single-input-single-output (SISO) channel was considered by Taricco and Elia [TE97] and they obtained the capacity behavior in asymptotically low and high SNR regimes. Abou-Faycal *et al.* [ATS01] further studied this case; they showed that for any given SNR, the capacity is achieved by an input distribution with a finite number of mass points. Lapidoth and Moser [LM03] showed that for the noncoherent MIMO with $T = 1$, the capacity behaves double logarithmically with the SNR for high SNR and this result was later extended to noncoherent networks [Lap05]. In contrast, the work of Zheng and Tse [ZT02] showed that when there is block-fading (*i.e.,* $T > 1$), then for high SNR, the capacity can scale logarithmically with the SNR. They showed that when the links are i.i.d. with M transmit antennas and N receive antennas, the number of transmit antennas M^*, required to attain the degrees of freedom (DoF) was $\min\left(\lfloor T/2 \rfloor, M, N\right)$. The DoF was shown to be $M^*\left(1 - M^*/T\right)$ in that case. The case of the noncoherent MIMO with asymmetric statistics on the link strengths was studied in Chapter 2. There, we showed that the (generalized degrees of freedom) gDoF for single-input-multiple-output (SIMO) and multiple-input-single-output (MISO) channels can be achieved by using only the strongest link. Also, for the 2×2 MIMO with two different SNR-exponents in the direct-links and cross-links, the gDoF was derived as a function of the SNR-exponents and the coherence time. Also, they showed that several insights from the identical link statistics scenarios of [MH99, ZT02] may not carry over to the case with asymmetric statistics; including the optimality of training and the number of antennas to be used.

The noncoherent single relay network with identical link strengths and unit coherence time was studied in [KK13], where it was shown that the relay does not increase the capacity at high SNR under certain conditions on the fading statistics. Similar observations were made in [GY14] for the noncoherent MIMO full-duplex single relay channel with block-fading, where they showed that Grassmanian signaling can achieve the DoF without using the relay. Also, their results show that for certain regimes, decode-and-forward with Grassmanian signaling can approximately achieve the capacity at high SNR and the characterization is determined by the number of antennas at the nodes. However, the assumption in [KK13, GY14] is that the channel strengths are symmetric in the sense explained below *i.e.,* these papers studied DoF and not gDoF. In many scenarios, the average strengths of the links can be asymmetric, *i.e.,* some links could be significantly weaker than others. This can happen when the relays are well separated, then the channel gains can be very different and this is not captured by the DoF. These differences matter in the high SNR regime, if the channel strengths are significantly different with respect to the operating SNR. This way of accounting for channel strength asymmetry in terms of SNR was introduced in the work of Zheng and Tse [ZT03] for calculating the diversity multiplexing tradeoff for the coherent MIMO[2] and was subsequently used for the coherent interference channel to analyze the gDoF [ETW08]. We use a similar framework to model channel asymmetry for our noncoherent relay problem. Therefore, in this sense, our model is for asymmetric channels (in terms of SNR scaling) in contrast to the symmetric channels (in terms of SNR scaling) studied in [KK13, GY14].

The diamond (parallel relay) network was introduced in [SG00]. Though the single-letter capacity is still unknown, for the coherent network (known channels) it has been characterized to within a constant additive bound (and in some scenarios a constant multiplicative bound) in [ADT11], with improved bounds established in [ND13, SWF14, KOG15]. As mentioned earlier, ours is the noncoherent model, which, to the best of our knowledge, has not

[2]In the diversity multiplexing tradeoff [ZT03], the channel statistics were symmetric and the different channel scaling was introduced to account for channel strength variation in their realization. However, this was not the case for the coherent interference channel [ETW08] where the notion of gDoF was defined.

been studied for the diamond network. We consider a block-fading channel model where the fading gains are i.i.d. Rayleigh distributed and remain constant for T symbol periods. Our model considers the diamond network where the link strength could have different fading distributions. This is naturally motivated when the relay locations are well separated, causing the links to have different average strengths (and therefore different statistics).

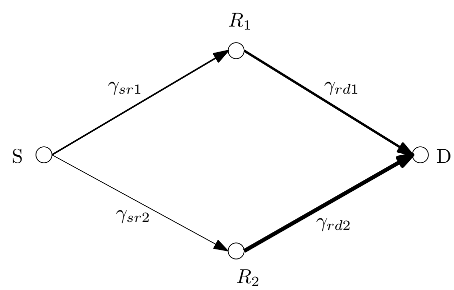


Figure 3.1: The 2-relay diamond network with given SNR exponents of link strengths.

In this chapter we have the following contributions:

1. We develop an outer bound for the block-fading noncoherent diamond network, which does not follow directly from the standard cut-set bound. We derive the structure of the optimal distribution required for this outer bound. We reduce the outer bound to a simpler form preserving the gDoF using novel techniques.

2. We show that any scheme that allocates separate symbols for channel training for each link, cannot always meet the new gDoF outer bound.

3. We develop a new relaying strategy which we term as train-scale quantize-map-forward (TS-QMF) for the noncoherent diamond network, which we show achieves the new gDoF outer bound, and is therefore gDoF-optimal.

4. We demonstrate the tradeoff between network learning and utilization, by showing that there are certain regimes where a simple relay selection is gDoF-optimal and that there are other regimes where we need both the relays. Even in the regimes where both the relays are used, we do not necessarily learn the channel values, as seen in

the TS-QMF scheme. In regimes where we need to operate both the relays, we use a time-sharing random variable to coordinate the relay operation.

We first derive a new outer bound for the noncoherent diamond network[3] in Theorem 3.1. The main issue in not using the standard cut-set bound for the block-fading case is that the transmit symbol block at the relay can depend on the current received symbol block at the relay, due to the block-nature of the model (see Figure 3.3). Therefore, we need to develop a slightly modified form of the cut-set bound to account for the block-fading channel. This is expressed as an optimization problem (akin to the classical cut-set bound which is also expressed as an optimization). In Theorem 3.3, we develop the structure for the distribution that solves the optimization problem of the outer bound. We show that this is of the form LQ, where L is lower triangular and Q is a unitary isotropically distributed matrix independent of L. Next, in Theorem 3.2, we outline some regimes of the network parameters in which a relay selection together with the decode-and-forward strategy is gDoF-optimal. This shows that in the noncoherent case, we might need to use a smaller part of the network, as learning and communicating in the entire network might be suboptimal. In a way, this gives a form of network simplification, similar to that observed for the coherent case [NOF14], where it was shown that (simplified) subnetworks could achieve most of the network capacity. In [NOF14], the authors demonstrated that for the coherent n-relay diamond network, we can always find a subset of k relays that can achieve a fraction $k/(k+1)$ of the total capacity within a constant gap.

Next, we proceed to the more difficult regime in which a simple relay selection is not optimal. For this regime, in Theorem 3.4, we further develop novel outer bounding techniques to simplify the results in Theorem 3.1. The outer bounding techniques in this Chapter is influenced by the methods developed in Chapter 2 for the noncoherent MIMO: there, we discretized the outer bound (without losing gDoF) and used linear programming techniques to further reduce the outer bound. We analyze the outer bound from Theorem 3.1, and show

[3] We also provide a more general version for block-fading noncoherent *acyclic* networks in Appendix B.2.

that the optimization problem of the outer bound is solved (in terms of gDoF) by a joint distribution (of the signals for the source and the relays) which has only two mass points. This is proved in Theorem 3.4 by discretizing the terms in outer bound (without losing gDoF) and using linear programming arguments. Subsequently, in Theorem 3.5, we reduce the optimization problem for choosing the two mass points, to a bilinear optimization problem, and we solve it explicitly. The bilinear optimization does not arise in the noncoherent MIMO case Chapter 2. In Chapter 2, there was only a piecewise linear optimization.

The approximate capacity of the coherent diamond channel (and of general unicast networks) can be achieved by the quantize-map-forward (QMF) strategy [ADT11, ADT15]. Here the strategy is that the relay quantizes the received signal and maps it (uniformly at random) to the transmit codebook. The standard QMF strategy requires the knowledge of the channels at the destination, for this, the links need to be trained. If we use a standard training method for the noncoherent diamond network, we need at least one symbol in every block to train the channels from the source to the relays, and we need at least two symbols in every block to train the channels from the relays to the destination (since there are two variables to be learned at the destination). In Theorem 3.6, we analyze the gDoF (assuming perfect network state knowledge at every nodes) using only the remaining symbols after training and we verify that this cannot achieve our outer bound.

Subsequently, we develop a new relaying strategy, which we call "train-scale QMF" (see Section 3.3.2) which we show is gDoF-optimal with respect to our outer bound, in Theorem 3.7. In the new scheme, we use a combination of training, scaling and QMF schemes to achieve this: the source sends training symbols to the relays, the relays scale the data symbols with the channel estimate obtained from training, then the relays perform QMF on the scaled symbols. The scaling is performed at the relays, so that the destination need not have the knowledge of the channels from the source to the relays. Another important characteristic of our scheme is that the source sends training symbols to the relays, but the relays do not send training symbols to the destination. If the relays need to send training symbols to the destination, we need to set aside two symbols in every block, and this is not gDoF-optimal due to Theorem 3.6.

In certain regimes, the distribution solving the optimization of the outer bound effectively induces a nonconcurrent operation of the two relays: while one relay is ON, the other relay is OFF and vice versa. There are regimes where both the relays are operated simultaneously, but one of the relays is kept at a lower power. These regimes (described in Theorem 3.5) are identified jointly by the channel strengths and the coherence time. Also, similar to the coherent case, Theorem 3.2 identifies regimes in which relay selection is optimal for the noncoherent case. The regimes for relay selection can be identified by channel strengths alone, independent of the coherence time.

The rest of this chapter is organized as follows: in Section 5.2 we set up the system model, Section 3.3 presents our main results and some interpretations along with an outline of the proof ideas, with reference to lemmas and facts detailed in Section 3.4 which provides the main analysis and many of the proofs. Most detailed proofs are deferred to the appendixes.

3.2 System model

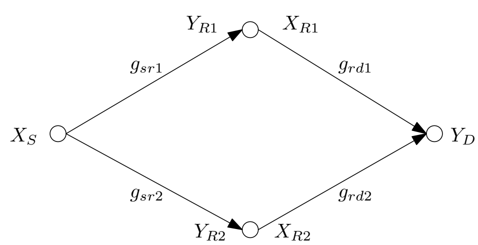


Figure 3.2: Signal flow over the 2-relay diamond network.

We use the same notations as defined in Section 2.2.1 on page 13. We consider a 2-relay diamond network as illustrated in Figure 3.2, with a coherence time of T symbol durations. The signal flow (over a block-length T) is given by:

$$\begin{bmatrix} Y_{\mathrm{R}_1} \\ Y_{\mathrm{R}_2} \end{bmatrix} = \begin{bmatrix} g_{\mathrm{sr1}} \\ g_{\mathrm{sr2}} \end{bmatrix} X_{\mathrm{S}} + \begin{bmatrix} W_{\mathrm{R}_1} \\ W_{\mathrm{R}_2} \end{bmatrix} \tag{3.1}$$

$$Y_{\mathrm{D}} = \begin{bmatrix} g_{\mathrm{rd}1} & g_{\mathrm{rd}2} \end{bmatrix} \begin{bmatrix} X_{\mathrm{R}_1} \\ X_{\mathrm{R}_2} \end{bmatrix} + W_{\mathrm{D}}. \tag{3.2}$$

For succinct notation let

$$X = \begin{bmatrix} X_{\mathrm{S}} \\ X_{\mathrm{R}_1} \\ X_{\mathrm{R}_2} \end{bmatrix}, \quad X_{\mathrm{R}} = \begin{bmatrix} X_{\mathrm{R}_1} \\ X_{\mathrm{R}_2} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_{\mathrm{R}_1} \\ Y_{\mathrm{R}_2} \\ Y_{\mathrm{D}} \end{bmatrix}, \quad Y_{\mathrm{R}} = \begin{bmatrix} Y_{\mathrm{R}_1} \\ Y_{\mathrm{R}_2} \end{bmatrix}, \tag{3.3}$$

$$G = \begin{bmatrix} g_{\mathrm{sr}1} & 0 & 0 \\ g_{\mathrm{sr}2} & 0 & 0 \\ 0 & g_{\mathrm{rd}1} & g_{\mathrm{rd}2} \end{bmatrix}, \quad W = \begin{bmatrix} W_{\mathrm{R}_1} \\ W_{\mathrm{R}_2} \\ W_{\mathrm{D}} \end{bmatrix}. \tag{3.4}$$

Then we have the effective flow

$$Y = GX + W, \tag{3.5}$$

where X_{S} is the $1 \times T$ vector of transmitted symbols from the source, $g_{\mathrm{sr}i}$ is the channel from the source to the relay R_i, W_{R_i} is the $1 \times T$ noise vector at the relay R_i with i.i.d. $\mathcal{CN}(0,1)$ elements, Y_{R_i} is the $1 \times T$ vector of received symbols at the relay R_i, X_{R_i} is the $1 \times T$ vector of transmitted symbols from the relay R_i, $g_{\mathrm{sr}i}$ is the channel from the relay R_i to the destination, W_{D} is the $1 \times T$ noise vector at the destination with its elements $w_{\mathrm{d}j} \sim$ i.i.d. $\mathcal{CN}(0,1)$ and Y_{D} is the $1 \times T$ vector of received symbols at the destination. The channels $g_{\mathrm{sr}i}, g_{\mathrm{rd}i}$ remains constant over block-length T. Every block has independent instances of $g_{\mathrm{sr}i}$,$g_{\mathrm{rd}i}$ with $g_{\mathrm{sr}i} \sim \mathcal{CN}(0, \rho_{\mathrm{sr}i}^2)$ i.i.d. and $g_{\mathrm{rd}i} \sim \mathcal{CN}(0, \rho_{\mathrm{rd}i}^2)$ i.i.d. For gDoF analysis, we define $\gamma_{\mathrm{sr}i}, \gamma_{\mathrm{rd}i}$[4] as

$$\gamma_{\mathrm{sr}i} = \frac{\log(\rho_{\mathrm{sr}i}^2)}{\log(\mathsf{SNR})}, \ \gamma_{\mathrm{rd}i} = \frac{\log(\rho_{\mathrm{rd}i}^2)}{\log(\mathsf{SNR})}. \tag{3.6}$$

The transmitted symbols at each relay are dependent only on the previously received symbols at the relay. The transmit signals have the average power constraint: $(1/T)\,\mathbb{E}\left[\|X_{\mathrm{S}}\|^2\right] = (1/T)\,\mathbb{E}\left[\|X_{\mathrm{R}_1}\|^2\right] = (1/T)\,\mathbb{E}\left[\|X_{\mathrm{R}_2}\|^2\right] = 1.$

[4]For our analysis, we assume that $\gamma_{\mathrm{sr}i} = \log\left(\rho_{\mathrm{sr}i}^2\right)/\log(\mathsf{SNR})$ and $\gamma_{\mathrm{rd}i} = \log\left(\rho_{\mathrm{rd}i}^2\right)/\log(\mathsf{SNR})$ are constants. However, we could define $\hat{\gamma}_{\mathrm{sr}i} = \log\left(\rho_{\mathrm{sr}i}^2\right)/\log(\mathsf{SNR})$, $\hat{\gamma}_{\mathrm{rd}i} = \log\left(\rho_{\mathrm{rd}i}^2\right)/\log(\mathsf{SNR})$ which depends on SNR and consider the limits $\gamma_{\mathrm{sr}i} = \lim_{\mathsf{SNR}\to\infty} \hat{\gamma}_{\mathrm{sr}i}$, $\gamma_{\mathrm{rd}i} = \lim_{\mathsf{SNR}\to\infty} \hat{\gamma}_{\mathrm{rd}i}$ which can be assumed to be constants independent of SNR. However, the gDoF analysis will yield the same result for both cases.

3.3 Main Results

In this section, we derive the gDoF for the noncoherent diamond network. For this purpose, in Theorem 3.1, we first derive a modified version of the cut-set outer bound for the noncoherent diamond network. This outer bound is in the form of an optimization problem. A looser version of this outer bound (that can be easily evaluated) can be used in certain regimes to obtain the gDoF. For achieving the gDoF in these regimes, we use relay selection and the decode-and-forward strategy. These regimes and the details of the achievability scheme are given in Theorem 3.2.

For the regimes that are not handled by Theorem 3.2, we calculate new outer bounds in Section 3.3.1 by simplifying and solving the outer bound optimization problem from Theorem 3.1. The outer bound is developed through Theorem 3.3, Theorem 3.4 and Theorem 3.5. Theorem 3.3 derives the structure of the optimizing distribution for the outer bound from Theorem 3.1. Theorem 3.4 uses this structure to bring the outer bound to a form that can be explicitly solved. The solution is obtained by Theorem 3.5.

In Theorem 3.6, we show that training-based schemes are not optimal in general for the regime considered in Section 3.3.1. Subsequently, we develop a new scheme that meets the outer bound developed in Section 3.3.1. The scheme is described in Section 3.3.2. In Theorem 3.7, this scheme is shown to meet the outer bound.

Theorem 3.1. *For the 2-relay diamond network, the capacity is outer bounded by $\bar{C}$, where*

$$T\bar{C} = \sup_{p(X)} \min \Big\{ I\left(X_S; Y_R\right), I\left(X_S; Y_{R_2}\right) + I\left(X_{R_1}; Y_D \middle| X_{R_2}\right), \\ I\left(X_S; Y_{R_1}\right) + I\left(X_{R_2}; Y_D \middle| X_{R_1}\right), I\left(X_R; Y_D\right) \Big\} \tag{3.7}$$

with X, X_R, Y_R defined in (3.3).

Proof idea. This is a modified version of the cut-set outer bound for noncoherent networks. The conventional cut-set outer bound does not automatically follow for the noncoherent case. The main reason for this is that we have a block-fading model, which means that there is a mismatch between the symbols and the block memoryless nature of the channel. Figure 3.3 illustrates this, where it can be seen that the causal relaying means that the symbols from the current fading block could potentially be used for relaying, causing the mismatch between the block memoryless model and the relaying. The detailed proof is in Appendix B.1. Theorem 3.1 is stated for the 2-relay diamond network, but this can be generalized and a generalized version of the cut-set outer bound for acyclic noncoherent networks is given in Appendix B.2. □

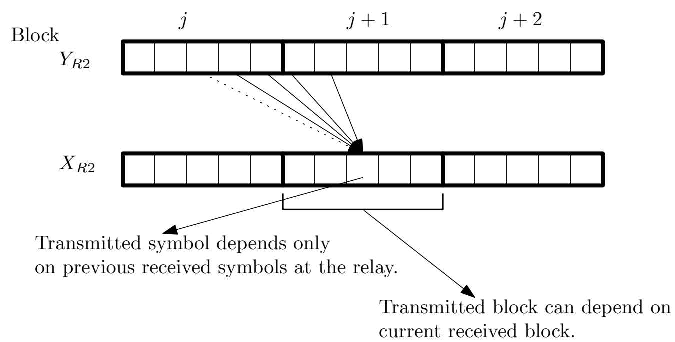


Figure 3.3: The transmitted symbols from the relays depend only on the previously received symbols, including the current fading block. Therefore the transmitted symbol could depend on the received symbols in the current fading block.

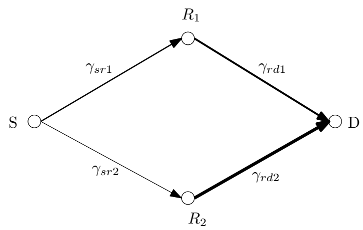
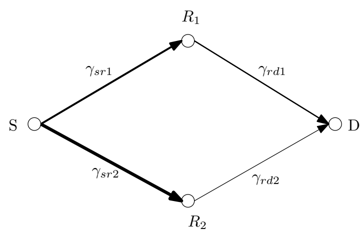
In the next theorem, we explain the regimes in which the gDoF can be achieved by a simple relay selection and the decode-and-forward strategy.

Theorem 3.2. *For the 2-relay diamond network with parameters in the regimes indicated in Table 3.1, the gDoF can be achieved by selecting a single relay as indicated in Table 3.1.*

Proof. For achievability, we use the decode-and-forward strategy by selecting a single relay depending on the regime as indicated in Table 3.1. (The existing noncoherent schemes [ZT02] can be used in each link). For example when $\gamma_{\mathrm{rd1}} \geq \gamma_{\mathrm{sr1}} \geq \gamma_{\mathrm{sr2}}$, we use decode-and-forward using only Relay R_1. The gDoF achievable from the source to Relay R_1 is $(1-1/T)\,\gamma_{\mathrm{sr1}}$ and

[5] *For the figures in the table, the thickness of each arrow is just an illustration consistent with the range of the gamma parameters in the first column of the table. There could be other consistent illustrations.*

Table 3.1: Regimes where a simple relay selection is gDoF-optimal.

Regime	Illustration[5]	Relay selected	gDoF
$\gamma_{\mathrm{rd1}} \geq \gamma_{\mathrm{sr1}} \geq \gamma_{\mathrm{sr2}}$		R_1	$\left(1-\frac{1}{T}\right)\gamma_{\mathrm{sr1}}$
$\gamma_{\mathrm{sr1}} \geq \gamma_{\mathrm{rd1}} \geq \gamma_{\mathrm{rd2}}$		R_1	$\left(1-\frac{1}{T}\right)\gamma_{\mathrm{rd1}}$

gDoF achievable from Relay R_1 to the destination is $(1-1/T)\,\gamma_{\mathrm{rd1}}$ [ZT02]. Each link can be trained using one symbol, the rest of the symbols can be used for data transmission and this will achieve the gDoF for each link [ZT02]. Thus, in this case, the gDoF achievable from the source to the destination evaluates to $\min\{(1-1/T)\,\gamma_{\mathrm{sr1}}, (1-1/T)\,\gamma_{\mathrm{rd1}}\} = (1-1/T)\,\gamma_{\mathrm{sr1}}$. The other case from the last row of Table 3.1 can be similarly evaluated.

Now, we only need to show the outer bound for these cases. We use the outer bound

$$T\bar{C} \leq \min\left\{\sup_{p(X)} I\left(X_{\mathrm{S}};Y_{\mathrm{R}}\right), \sup_{p(X)} I\left(X_{\mathrm{R}};Y_{\mathrm{D}}\right)\right\}. \tag{3.8}$$

This is obtained by loosening (3.7). The above equation consists of a SIMO term and a MISO term. From Chapter 2, the gDoF for SIMO and MISO can be achieved using just the strongest link. Hence the above equation yields the gDoF outer bound

$$\bar{\gamma} \leq \left(1-\frac{1}{T}\right)\min\left\{\max\left\{\gamma_{\mathrm{sr1}},\gamma_{\mathrm{sr2}}\right\}, \max\left\{\gamma_{\mathrm{rd1}},\gamma_{\mathrm{rd2}}\right\}\right\}. \tag{3.9}$$

This equation for the gDoF outer bound reduces to the gDoF term in Table 3.1 in the different regimes as indicated in the table. For example when $\gamma_{\mathrm{rd1}} \geq \gamma_{\mathrm{sr1}} \geq \gamma_{\mathrm{sr2}}$, the right-hand-side (RHS) of (3.9) reduces to $(1-1/T)\,\gamma_{\mathrm{sr1}}$. $\square$

Discussion: The regimes discussed in this theorem arise also in the coherent case and relay selection is optimal for the coherent case in these regimes. These regimes are dictated by the γ parameters alone, independent of T. As we look into other regimes, we will see that the coherence time T will also affect the relay operation and achievability strategies.

The rest of the results are about the nontrivial regimes of the 2-relay diamond network that cannot be handled with the simple outer bound from (3.9). We develop new outer bound techniques and achievability schemes for these cases. The outer bound techniques involve loosening (3.7), discretizing the terms involved in it without losing gDoF and subsequently obtaining a solution for the optimization formulation of the outer bound using linear programming methods. Achievability schemes involve a modification of the QMF strategy [ADT11, ADT15]: the differences from the standard QMF strategy to our scheme are that we only partially train the network and we use a scaling at the relays to avoid the necessity of the knowledge of the entire network parameters at the destination. Also, from (3.9), it is clear that if $T = 1$, the gDoF is zero. Hence we consider $T \geq 2$ for the rest of the chapter.

3.3.1 Nontrivial Regimes of the Diamond Network

In this section, we deal with the diamond network with its parameters lying in the regime that cannot be handled by the decode-and-forward strategy as in Theorem 3.2. This regime has

$$\gamma_{\text{sr1}} > \gamma_{\text{sr2}}, \gamma_{\text{sr1}} > \gamma_{\text{rd1}}, \gamma_{\text{rd2}} > \gamma_{\text{rd1}}, \gamma_{\text{rd2}} > \gamma_{\text{sr2}}. \tag{3.10}$$

With this, all regimes of the diamond network are covered (we exclude the cases which can be obtained by relabeling the relays).

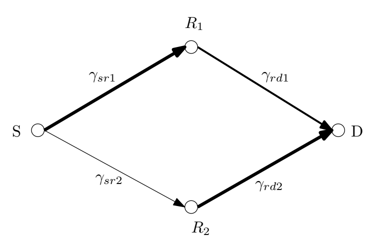


Figure 3.4: Regime with $\gamma_{\text{sr1}} > \gamma_{\text{sr2}}$, $\gamma_{\text{sr1}} > \gamma_{\text{rd1}}$, $\gamma_{\text{rd2}} > \gamma_{\text{rd1}}$ and $\gamma_{\text{rd2}} > \gamma_{\text{sr2}}$.

We now proceed with developing a (tight) gDoF outer bound for this regime.

Theorem 3.3. *The outer bound (3.7) for the diamond network is optimized by input X of the form $X = LQ$ with L being a lower triangular matrix and Q being an isotropically distributed unitary matrix independent of L.*

Proof. This is an adaptation of the existing results for the noncoherent MIMO channel [MH99]. The proof is in Appendix B.3. $\square$

Theorem 3.4. *The outer bound (3.7) can be further upper bounded as*

$$T\bar{C} \dot{\leq} \min\left\{(T-1)\log\left(\rho_{sr1}^2\right), (\mathcal{P}_1)\right\}, \tag{3.11}$$

where $(\mathcal{P}_1)$ is the solution of the optimization problem

$$\mathcal{P}_1 : \begin{cases} \underset{\mathbb{E}\left[|a|^2\right]\leq T, \mathbb{E}\left[|b|^2+|c|^2\right]\leq T}{\text{maximize}} \quad \min\left\{\psi_1, \psi_2\right\} \\ |a|^2, |b|^2, |c|^2 \geq 0 \end{cases} \tag{3.12}$$

with

$$\begin{aligned}\psi_1 =& T\mathbb{E}\left[\log\left(\rho_{rd2}^2|a|^2 + \rho_{rd1}^2|b|^2 + \rho_{rd1}^2|c|^2 + T\right)\right] \\ & - \mathbb{E}\left[\log\left(\rho_{rd2}^2|a|^2 + \rho_{rd1}^2|b|^2 + \rho_{rd1}^2|c|^2 + \rho_{rd1}^2\rho_{rd2}^2|c|^2|a|^2 + 1\right)\right],\end{aligned} \tag{3.13}$$

$$\begin{aligned}\psi_2 =& (T-1)\log\left(\rho_{sr2}^2\right) + \mathbb{E}\left[\log\left(\rho_{rd2}^2|a|^2 + \rho_{rd1}^2|b|^2 + 1\right)\right] \\ & + (T-1)\mathbb{E}\left[\log\left(\rho_{rd1}^2|c|^2 + T - 1\right)\right] \\ & - \mathbb{E}\left[\log\left(\rho_{rd2}^2|a|^2 + \rho_{rd1}^2|b|^2 + \rho_{rd1}^2|c|^2 + \rho_{rd1}^2\rho_{rd2}^2|c|^2|a|^2 + 1\right)\right].\end{aligned} \tag{3.14}$$

Proof. Due to Theorem 3.3, we have

$$\begin{bmatrix} X_{\mathrm{R}_2} \\ X_{\mathrm{R}_1} \\ X_{\mathrm{S}} \end{bmatrix} = \begin{bmatrix} a & 0 & 0 & . & . & . & 0 \\ b & c & 0 & . & . & . & 0 \\ d & e & f & 0 & . & . & 0 \end{bmatrix} Q \tag{3.15}$$

as the structure of the optimizing distribution, with a, b, c, d, e, f as random with unknown distributions and Q being an isotropic unitary $T\times 3$ matrix independent of the other variables.

Hence we have

$$
\begin{aligned}
T\bar{C} &= \sup_{p(a,b,c,d,e,f)} \min \Big\{ I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right), I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right), \\
&\qquad\qquad I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_1}\right) + I\left(X_{\mathrm{R}_2}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}\right), I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right) \Big\} && (3.16) \\
&\leq \sup_{p(a,b,c,d,e,f)} \min \left\{ I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right), I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_1}\right) + I\left(X_{\mathrm{R}_2}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}\right), I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right) \right\} && (3.17) \\
&\leq \min \Bigg\{ \sup_{p(a,b,c,d,e,f)} I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right), \\
&\qquad \sup_{p(a,b,c,d,e,f)} \min \left\{ I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right), I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right) \right\} \Bigg\} && (3.18) \\
&\overset{(i)}{\dot{\leq}} \min \Big\{ (T-1) \log \left(\rho_{\mathrm{sr1}}^2\right), \\
&\qquad \sup_{p(a,b,c,d,e,f)} \min \left\{ I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right), I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right) \right\} \Big\} && (3.19) \\
&\overset{(ii)}{\dot{\leq}} \min \left\{ (T-1) \log \left(\rho_{\mathrm{sr1}}^2\right), \sup_{p(a,b,c)} \min \left\{ \psi_1, \psi_2 \right\} \right\} && (3.20) \\
&\overset{(iii)}{=} \min \left\{ (T-1) \log \left(\rho_{\mathrm{sr1}}^2\right), (\mathcal{P}_1) \right\}. && (3.21)
\end{aligned}
$$

In step (i), we observe that $I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right)$ corresponds to a noncoherent SIMO channel. From Chapter 2, the gDoF of the noncoherent SIMO is achieved by using the strongest link alone. Hence

$$
\sup_{p(a,b,c,d,e,f)} I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right) \dot{\leq} (T-1) \log \left(\rho_{\mathrm{sr1}}^2\right)
$$

follows in step (i). The step (ii) is by showing

$$
I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right) \dot{\leq} \psi_1, \; I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right) \dot{\leq} \psi_2,
$$

where ψ_i are independent of d, e, f. The details of step (ii) are in Section 3.4.2. In step (iii), we defined $(\mathcal{P}_1) = \sup_{p(a,b,c)} \min \{\psi_1, \psi_2\}$. The optimization problem $\mathcal{P}_1$ can be viewed as a tradeoff between between a MISO cut (Figure 3.8 on page 69) and a parallel cut (Figure 3.9 on page 71). The tradeoff arises because the unknown channel from one of the relays acts as an interference to the transmission from the other relay, hence the operation of Relay R_1 and Relay R_2 need to be optimized. □

In the following lemma, we further reduce $\mathcal{P}_1$ into a form that can be solved explicitly.

Lemma 3.1. *The solution of $\mathcal{P}_1$ has the same gDoF as the solution of $\mathcal{P}_9$.*

$$\mathcal{P}_9 : \begin{cases} \textit{maximize min} \Big\{ p_\lambda \left((T-1)\, \gamma_{rd2} \log(\mathsf{SNR}) - \log\left(\mathsf{SNR}^{\gamma_{rd1}} |c_1|^2 + 1 \right) \right) \\ \qquad\qquad + (T-1)(1-p_\lambda)\, \gamma_{rd1} \log(\mathsf{SNR}) \,, \ (T-1)\, \gamma_{sr2} \log(\mathsf{SNR}) \\ \qquad\qquad + (T-2)\, p_\lambda \log\left(\mathsf{SNR}^{\gamma_{rd1}} |c_1|^2 + 1 \right) \\ \qquad\qquad + (T-1)(1-p_\lambda)\, \gamma_{rd1} \log(\mathsf{SNR}) \Big\} \\ |c_1|^2 \leq T, 0 \leq p_\lambda \leq 1, \end{cases} \tag{3.22}$$

i.e.,

$$gDoF(\mathcal{P}_1) = gDoF(\mathcal{P}_9)\,. \tag{3.23}$$

Proof sketch. The proof proceeds in several steps in Appendix B.4. We show in (B.61) that we can discretize the function $\min\{\psi_1, \psi_2\}$ over discrete values of $\left(|a|^2, |b|^2, |c|^2\right)$ without losing gDoF. The discretization is over countably infinite points with the distance between points chosen inversely proportional to the SNR. This is illustrated as the first step in Figure 3.5. We then show that at any SNR the discretization can be limited to a finite number of points without losing gDoF. This is illustrated as the second step in Figure 3.5. With a fixed finite number of points, maximizing $\min\{\psi_1, \psi_2\}$ can be reduced to a linear program with the probabilities at the discrete points as the variables. This linear program together with the total power and probability constraints can be shown to have its optimal solution with just 3 nonzero probability points. This is illustrated as the third step in Figure 3.5. We then collapse 3 nonzero probability points to 2 points using the structure of the objective function. Again we use the structure of the function $\min\{\psi_1, \psi_2\}$ to reduce the problem to an optimization problem over two variables $|c_1|^2, p_\lambda$ as in $\mathcal{P}_9$. The details are in Appendix B.4. $\square$

Discussion: Effectively, $\mathcal{P}_9$ is derived from $\mathcal{P}_1$ with a probability distribution

$$\left(|a|^2, |b|^2, |c|^2\right) = \begin{cases} \left(T, 0, |c_1|^2\right) & \text{w.p. } p_\lambda \\ (0, T/2, T/2) & \text{w.p. } (1 - p_\lambda) \end{cases} \tag{3.24}$$

as the solution and reducing the optimization problem to the variables $p_\lambda, |c_1|^2$. These points are not directly obtained, but the problem is reduced in several steps, to reach the final form containing contribution only from the two points. The existence of the two points in the outer bound suggests the necessity to use a time-sharing random variable to coordinate the two relays to achieve the gDoF. The variable a is associated with relay R_2 and (b, c) is with relay R_1. The mass point $\left(T, 0, |c_1|^2\right)$ needs both relays, however the point $(0, T/2, T/2)$ needs only Relay R_1. After further solving the optimization problem, if $|c_1|^2$ turns out to be zero, the joint distribution would be using a nonconcurrent operation of the relays: while one relay is ON , the other needs to be OFF and vice versa. Though this is in the outer bound, it suggests the (gDoF) optimal system operation through this interpretation.

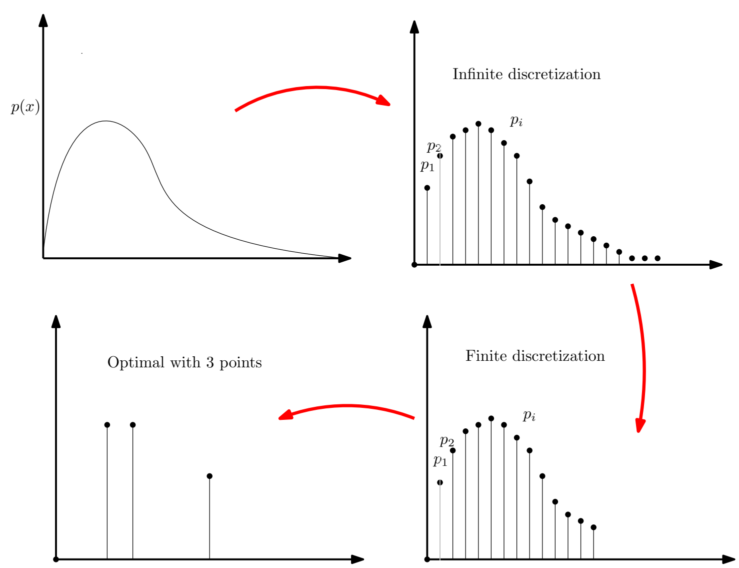


Figure 3.5: Proof methodology by discretization and linear programming arguments for Lemma 3.1.

Theorem 3.5. *The optimization problem $\mathcal{P}_9$ given in (3.22) has the solution as given in Table 3.2.*

Table 3.2: Solution of ($\mathcal{P}_9$) for achieving the gDoF.

<table>
<tr><th colspan="2">Regime</th><th>$\lvert c_1\rvert^2$</th><th>p_λ</th></tr>
<tr><td colspan="2">$(T-2)\,\gamma_{\mathrm{rd2}}-(T-1)\,\gamma_{\mathrm{rd1}}\le 0$</td><td>$0$</td><td>$\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}}$</td></tr>
<tr><td rowspan="2">$(T-2)\,\gamma_{\mathrm{rd2}}-(T-1)\,\gamma_{\mathrm{rd1}}> 0$</td><td>$\gamma_{\mathrm{rd2}}>\gamma_{\mathrm{sr2}}+\gamma_{\mathrm{rd1}}$</td><td>$0$</td><td>$\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}}$</td></tr>
<tr><td>$\gamma_{\mathrm{rd2}}\le\gamma_{\mathrm{sr2}}+\gamma_{\mathrm{rd1}}$</td><td>$\mathsf{SNR}^{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{sr2}}-\gamma_{\mathrm{rd1}}}$</td><td>$1$</td></tr>
</table>

Proof idea. The detailed proof is in Section 3.4.3. We change variables from $|c_1|^2$ to γ_c using the transformation $\rho_{\mathrm{rd1}}^2\,|c_1|^2=\mathsf{SNR}^{\gamma_c}$ with $\gamma_c\le\gamma_{\mathrm{rd1}}$. This yields a bilinear optimization problem in terms of γ_c and p_λ. The bilinear optimization problem gives different solutions depending on the value of coefficients involved, and we tabulate the results. □

Discussion: We have regimes in which $|c_1|^2=0$. Here the optimizing distribution of the outer bound effectively suggests a nonconcurrent operation of the two relays. Also, when $p_\lambda=0$, $|a|^2$ is always zero and hence the Relay R_2 is not used. (This follows due to the structure of the solution (3.15) and the nature of the mass points from (3.24)).

Theorem 3.6. *(Nonoptimality of training schemes) For the regime described in (3.10), training-based schemes cannot always achieve the outer bound (3.21).*

Proof. If only a single relay is used, we need to set aside at least one symbol in every block of length T, to train the channel from the source to the relays and the channel from the relays to the destination. Then the gDoF achievable is

$$\gamma_{1,\mathrm{train}}\times T=(T-1)\max\left\{\min\left\{\gamma_{\mathrm{sr1}},\gamma_{\mathrm{rd1}}\right\},\min\left\{\gamma_{\mathrm{sr2}},\gamma_{\mathrm{rd2}}\right\}\right\}. \tag{3.25}$$

If both the relays are used, for training the channels from the relays to the destination, we need to set aside at least two symbols in every block of length T, since there are two parameters to be learned at the destination. For training the channels from the source to the relays, we need to set aside at least one symbol in every block of length T. After training, we can have super-symbols from the source to the relays with length at most $T-1$, and from the relays to the destination with length at most $T-2$. Now, using the cut-set outer bound with this super-symbols, and assuming perfect network state knowledge at all nodes *i.e.,* using a coherent outer bound, we can upper bound the gDoF $\gamma_{2,\text{train}}$ achievable using training-based scheme as

$$\begin{aligned}\gamma_{2,\text{train}} \times T \leq & \min\{\,(T-1)\,\gamma_{\text{sr}1}, (T-2)\,\gamma_{\text{rd}2}, (T-1)\,\gamma_{\text{sr}2} + (T-2)\,\gamma_{\text{rd}1}, \\ & (T-1)\,\gamma_{\text{sr}1} + (T-2)\,\gamma_{\text{rd}2}\} \end{aligned} \tag{3.26}$$

$$\overset{(i)}{=} \min\{(T-1)\,\gamma_{\text{sr}1}, (T-2)\,\gamma_{\text{rd}2}, (T-1)\,\gamma_{\text{sr}2} + (T-2)\,\gamma_{\text{rd}1}\}, \tag{3.27}$$

where (i) is because $\gamma_{\text{sr}1} > \gamma_{\text{sr}2}, \gamma_{\text{sr}1} > \gamma_{\text{rd}1}, \gamma_{\text{rd}2} > \gamma_{\text{rd}1}, \gamma_{\text{rd}2} > \gamma_{\text{sr}2}$ in this regime.

Now, examining the outer bound (3.21), in order to complete the proof, we just need to give a sample point where

$$\gamma_{1,\text{train}} \times T, \gamma_{2,\text{train}} \times T < \min\{(T-1)\,\gamma_{\text{sr}1}, \text{gDoF}\,(\mathcal{P}_1)\} \tag{3.28}$$

with strict inequality. We give a sample point $T=3, \gamma_{\text{sr}1}=4, \gamma_{\text{sr}2}=1, \gamma_{\text{rd}1}=2, \gamma_{\text{rd}2}=3$. Now with this choice

$$(T-1)\max\{\min\{\gamma_{\text{sr}1}, \gamma_{\text{rd}1}\}, \min\{\gamma_{\text{sr}2}, \gamma_{\text{rd}2}\}\} = 4 \tag{3.29}$$

$$\min\{(T-1)\,\gamma_{\text{sr}1}, (T-2)\,\gamma_{\text{rd}2}, (T-1)\,\gamma_{\text{sr}2} + (T-2)\,\gamma_{\text{rd}1}\} = 3 \tag{3.30}$$

$$\min\{(T-1)\,\gamma_{\text{sr}1}, \text{gDoF}\,(\mathcal{P}_1)\} = 5.33, \tag{3.31}$$

where $\text{gDoF}\,(\mathcal{P}_1)$ is evaluated using Lemma 3.1 and Table 3.2. One can construct several other counterexamples to demonstrate the suboptimality of training. □

3.3.2 Train-Scale Quantize-Map-Forward (TS-QMF) Scheme for the Noncoherent Diamond Network

In this section, we describe our scheme for achieving the gDoF for the nontrivial regime (3.10) of the diamond network. The same scheme can be used to achieve the gDoF in the other regimes, but decode-and-forward is also gDoF-optimal in those regimes. Our scheme is a modification of the QMF scheme developed in [ADT11, OD13, ADT15]. The QMF strategy, introduced in [ADT11] is the following. Each relay first quantizes the received signal, then randomly maps it to a Gaussian codeword and transmits it. The destination then decodes the transmitted message, without requiring the decoding of the quantized values at the relays. The specific scheme that [ADT11] focused on was based on a scalar (lattice) quantizer followed by a mapping to a Gaussian random codebook. In [OD10, OD13], this was generalized to a lattice vector quantizer and [LKG11] generalized it to discrete memoryless networks. Our scheme is illustrated in Figure 3.6 and Figure 3.7. We discuss the modifications compared to the QMF scheme, more details on the QMF scheme can be found in [ADT11, OD13, ADT15]. The modifications compared to the QMF scheme are:

1. The source uses super-symbols of length T and the first symbol of the supersymbol is kept for training the channels from the source to the relays.

2. The relays use the first symbol from every received supersymbol to scale (the scaling is precisely defined in following paragraphs) the rest of the symbols in the received supersymbol, the scaled version (ignoring the first symbol) is quantized and mapped into super-symbols of length T and transmitted.

3. The relays are assumed to have access to a time-sharing random variable Λ. We choose Λ to be a Bernoulli random variable and the codebooks for the relays are generated using a distribution that is joint with the distribution of the time-sharing random variable.

We describe our scheme in more detail in the following paragraphs.

3.3.2.1 Source

The codewords at the source are generated according to a Gaussian distribution $p\left(X_S\right)$, where X_S is a vector of length $(T-1)$. The source encodes the message $m \in \left[1 : 2^{nTR}\right]$ onto

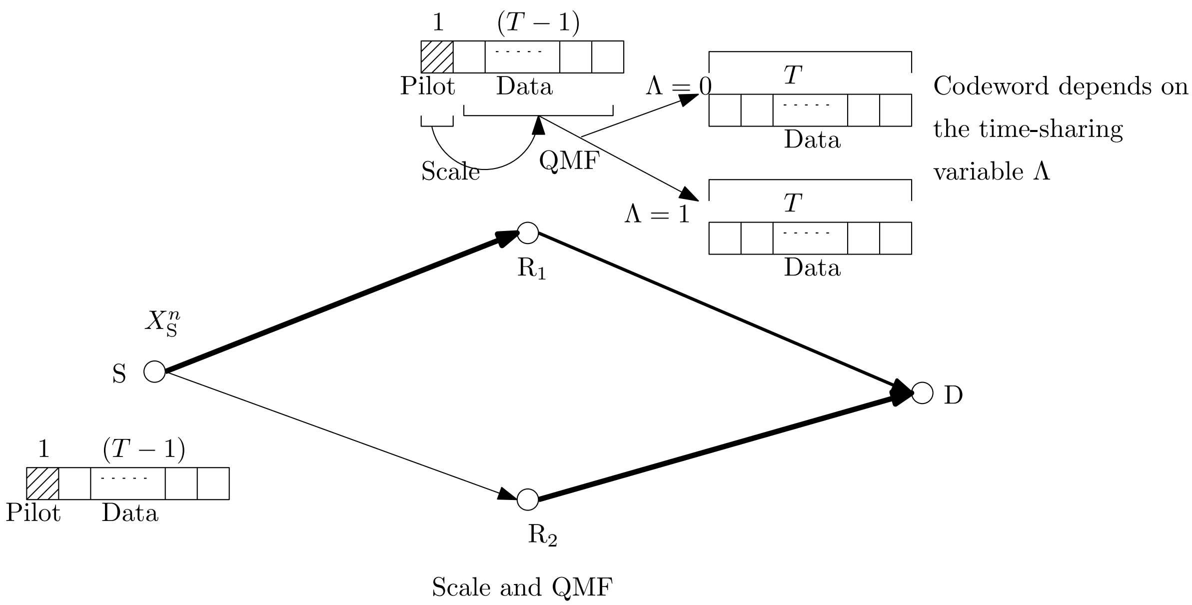


Figure 3.6: Summary of the achievability scheme: the source sends one pilot symbol in every block. The relays scale the data symbols using the pilot and perform QMF operation after scaling. The codewords sent at the relays depend also on the time sharing random variable Λ.

X_{S}^n with $X_{\mathrm{S}}^n = X_{\mathrm{S}}(1) \ldots X_{\mathrm{S}}(n)$ and each $X_{\mathrm{S}}(i)$ is a vector of length $(T-1)$. The source then transmits the sequence

$$[1, X_{\mathrm{S}}(1)], \ldots [1, X_{\mathrm{S}}(i)], \ldots [1, X_{\mathrm{S}}(n)].$$

Thus in every block of length T, the first symbol is for training and the rest of the symbols carry the data.

3.3.2.2 Relays

The relays are assumed to have access to a time-sharing random variable Λ. The codebooks at the relays are generated according to the joint distribution $p(\Lambda)\, p(X_{\mathrm{R}_1}|\Lambda)\, p(X_{\mathrm{R}_2}|\Lambda)$, where $p(X_{\mathrm{R}_i}|\Lambda)$ are Gaussian distributed. The random variables $X_{\mathrm{R}_1}, X_{\mathrm{R}_2}$ are vectors of length T. Hence the codeword sent at each relay depends also on the time-sharing random variable.

Since the source sends a known symbol (e.g., 1) for training at the beginning of every block (of length T), Relay R_1 can obtain $g_{\mathrm{sr1}}^n + w^n$ after n blocks, where $g_{\mathrm{sr1}}^n = g_{\mathrm{sr1}}(1) \ldots g_{\mathrm{sr1}}(n)$ contains the i.i.d. channel realizations across the n blocks and $w^n = w(1) \ldots w(n)$ contains the i.i.d. noise elements with $w(i) \sim \mathcal{CN}(0,1)$. The data symbols are received as $Y_{\mathrm{R}_1}^n = g_{\mathrm{sr1}}^n X_{\mathrm{S}}^n + W_{\mathrm{R}_1}^n$, where W_{R_1} is a noise vector of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$ elements. Relay R_1 scales $Y_{\mathrm{R}_1}^n$ to $Y_{\mathrm{R}_1}^{\prime n} = \frac{Y_{\mathrm{R}_1}^n}{\hat{g}_{\mathrm{sr1}}^n} = \frac{g_{\mathrm{sr1}}^n}{\hat{g}_{\mathrm{sr1}}^n} X_{\mathrm{S}}^n + \frac{W_{\mathrm{R}_1}^n}{\hat{g}_{\mathrm{sr1}}^n}$, where $\hat{g}_{\mathrm{sr1}}$ is obtained from $g_{\mathrm{sr1}} + w$ as

$$\hat{g}_{\mathrm{sr1}} = e^{i\angle(g_{\mathrm{sr1}}+w)} + (g_{\mathrm{sr1}} + w), \tag{3.32}$$

where $\angle(g_{\mathrm{sr1}} + w)$ is the angle of $g_{\mathrm{sr1}} + w$. This scaling is done at the relay using the trained channel, in order to avoid the necessity of having the knowledge of g_{sr1} at the destination. Our scaling uses a modified version $\hat{g}_{\mathrm{sr1}}$ instead of $g_{\mathrm{sr1}} + w$; this is because $1/(g_{\mathrm{sr1}} + w)$ could take infinite magnitude and this problem is avoided by using $1/\hat{g}_{\mathrm{sr1}}$.

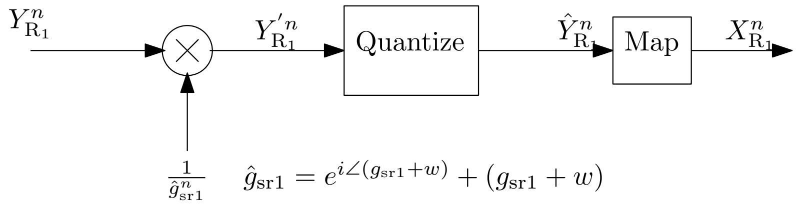


Figure 3.7: Processing at Relay R_1.

Relay R_1 quantizes the scaled version $Y_{\mathrm{R}_1}'^n = \frac{g_{\mathrm{sr}1}^n}{\hat{g}_{\mathrm{sr}1}^n} X_{\mathrm{S}}^n + \frac{W_{\mathrm{R}_1}^n}{\hat{g}_{\mathrm{sr}1}^n}$ into $\hat{Y}_{\mathrm{R}_1}^n = \frac{g_{\mathrm{sr}1}^n}{\hat{g}_{\mathrm{sr}1}^n} X_{\mathrm{S}}^n + \frac{W_{\mathrm{R}_1}^n}{\hat{g}_{\mathrm{sr}1}^n} + Q_{\mathrm{R}_1}^n$ with $Q_{\mathrm{R}_1} \sim \frac{W_{\mathrm{R}_1}}{\hat{g}_{\mathrm{sr}1}}$ independent of all the other variables. The quantized symbols are mapped into $X_{\mathrm{R}_1}^n$ and sent. The transmitted codeword $X_{\mathrm{R}_1}^n$ depends also on the time-sharing random variable. Note that the relays do not train the channels to the destination, as it turns out to be suboptimal as proved in Theorem 3.6.

Relay R_2 does similar processing. It quantizes $Y_{\mathrm{R}_1}'^n = \frac{g_{\mathrm{sr}2}^n}{\hat{g}_{\mathrm{sr}2}^n} X_{\mathrm{S}}^n + \frac{W_{\mathrm{R}_2}^n}{\hat{g}_{\mathrm{sr}2}^n}$ into $\hat{Y}_{\mathrm{R}_1}^n = \frac{g_{\mathrm{sr}2}^n}{\hat{g}_{\mathrm{sr}2}^n} X_{\mathrm{S}}^n + \frac{W_{\mathrm{R}_2}^n}{\hat{g}_{\mathrm{sr}2}^n} + Q_{\mathrm{R}_2}^n$ with $Q_{\mathrm{R}_2} \sim \frac{W_{\mathrm{R}_2}}{\hat{g}_{\mathrm{sr}2}}$ independent of all the other variables. The quantized symbols are mapped into $X_{\mathrm{R}_2}^n$ and sent. Again the transmitted codeword $X_{\mathrm{R}_2}^n$ depends also on the time-sharing random variable.

3.3.2.3 Destination

Using weak typicality decoding [OD10, LKG11, OD13, ADT15], the following rate is achievable:

$$\begin{aligned} TR = \min\Big\{ & I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}}, Y_{\mathrm{D}} \,\middle|\, X_{\mathrm{R}}, \Lambda\right), I\left(X_{\mathrm{R}}, X_{\mathrm{S}}; Y_{\mathrm{D}} \middle| \Lambda\right) - I\left(Y_{\mathrm{R}}'; \hat{Y}_{\mathrm{R}} \,\middle|\, X_{\mathrm{S}}, X_{\mathrm{R}}, Y_{\mathrm{D}}, \Lambda\right), \\ & I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \,\middle|\, X_{\mathrm{R}_2}, \Lambda\right) - I\left(Y_{\mathrm{R}_1}'; \hat{Y}_{\mathrm{R}_1} \,\middle|\, X_{\mathrm{S}}, X_{\mathrm{R}}, \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}}, \Lambda\right), \\ & I\left(X_{\mathrm{S}}, X_{\mathrm{R}_2}; \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}} \,\middle|\, X_{\mathrm{R}_1}, \Lambda\right) - I\left(Y_{\mathrm{R}_2}'; \hat{Y}_{\mathrm{R}_2} \,\middle|\, X_{\mathrm{S}}, X_{\mathrm{R}}, \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}}, \Lambda\right) \Big\} \end{aligned} \tag{3.33}$$

with

$$Y_{\mathrm{R}}' = \begin{bmatrix} Y_{\mathrm{R}_1}' \\ Y_{\mathrm{R}_2}' \end{bmatrix}, \; \hat{Y}_{\mathrm{R}} = \begin{bmatrix} \hat{Y}_{\mathrm{R}_1} \\ \hat{Y}_{\mathrm{R}_2} \end{bmatrix} \tag{3.34}$$

and using a distribution $p(\Lambda)\, p(X_{\mathrm{S}})\, p(X_{\mathrm{R}_1} | \Lambda)\, p(X_{\mathrm{R}_2} | \Lambda)\, p\left(\hat{Y}_{\mathrm{R}_1} \middle| Y_{\mathrm{R}_1}', X_{\mathrm{R}_1} \Lambda\right) p\left(\hat{Y}_{\mathrm{R}_2} \middle| Y_{\mathrm{R}_2}', X_{\mathrm{R}_2} \Lambda\right)$.

We choose the distribution for Λ as

$$\Lambda = \begin{cases} 0 & \text{w.p. } p_\lambda \\ 1 & \text{w.p. } 1 - p_\lambda \end{cases} \tag{3.35}$$

with p_λ being a constant to be chosen. We choose X_{S} as a $(T-1) \times 1$ vector with i.i.d. $\mathcal{CN}(0,1)$ elements, *i.e.,*

$$X_{\mathrm{S}} = [x_{\mathrm{s}}(1), \ldots, x_{\mathrm{s}}(i), \ldots, x_{\mathrm{s}}(T-1)] \tag{3.36}$$

with i.i.d. elements $x_s(i) \sim \mathcal{CN}(0,1)$, and we choose

$$X_{\mathrm{R}_1} = \begin{cases} a_{\mathrm{R10}} X_{\mathrm{R10}} & \text{if } \Lambda = 0 \\ a_{\mathrm{R11}} X_{\mathrm{R11}} & \text{if } \Lambda = 1 \end{cases} \tag{3.37}$$

$$X_{\mathrm{R}_2} = \begin{cases} a_{\mathrm{R20}} X_{\mathrm{R20}} & \text{if } \Lambda = 0 \\ a_{\mathrm{R21}} X_{\mathrm{R21}} & \text{if } \Lambda = 1, \end{cases} \tag{3.38}$$

where $X_{\mathrm{R10}}, X_{\mathrm{R11}}, X_{\mathrm{R20}}, X_{\mathrm{R21}}$ are all $T \times 1$ vectors with i.i.d. $\mathcal{CN}(0,1)$ components, all of them independent of each other, and $a_{\mathrm{R10}}, a_{\mathrm{R11}}, a_{\mathrm{R20}}, a_{\mathrm{R21}}$ are constants to be chosen.

We also have

$$\hat{Y}_{\mathrm{R}_1} = Y'_{\mathrm{R}_1} + Q_{\mathrm{R}_1}, \tag{3.39}$$

where $Y'_{\mathrm{R}_1} = \frac{g_{\mathrm{sr1}}}{\hat{g}_{\mathrm{sr1}}} X_{\mathrm{S}} + \frac{W_{\mathrm{R}_1}}{\hat{g}_{\mathrm{sr1}}}$, $Q_{\mathrm{R}_1} \sim \frac{W_{\mathrm{R}_1}}{\hat{g}_{\mathrm{sr1}}}$ and Q_{R_1} is independent of the other random variables.

Similarly

$$\hat{Y}_{\mathrm{R}_2} = Y'_{\mathrm{R}_2} + Q_{\mathrm{R}_2}, \tag{3.40}$$

where $Y'_{\mathrm{R}_2} = \frac{g_{\mathrm{sr2}}}{\hat{g}_{\mathrm{sr2}}} X_{\mathrm{S}} + \frac{W_{\mathrm{R}_2}}{\hat{g}_{\mathrm{sr2}}}$, $Q_{\mathrm{R}_2} \sim \frac{W_{\mathrm{R}_2}}{\hat{g}_{\mathrm{sr2}}}$ and Q_{R_2} is independent of the other random variables.

Theorem 3.7. *For the diamond network with parameters as described in Section 3.3.1, with the choice*

$$a_{R10} = c_1, a_{R11} = 1, a_{R20} = 1, a_{R21} = 0 \tag{3.41}$$

and choosing the values of $|c_1|^2, p_\lambda$ *from Table 3.2, the gDoF can be achieved.*

Proof sketch. The detailed proof is in Section 3.4.4. In the proof we analyze the expression of the achievable rate from (3.33). Using Theorem 3.13 and due to the relay operation of train-scale-quantization, we first show that the penalty terms $-I\left(Y'_{\mathrm{R}};\hat{Y}_{\mathrm{R}}\middle|X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right)$, $-I\left(Y'_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_1}\middle|X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_2},Y_{\mathrm{D}},\Lambda\right)$ and $-I\left(Y'_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_2}\middle|X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_1},Y_{\mathrm{D}},\Lambda\right)$ do not affect the gDoF when we use Gaussian codebooks with time sharing. Then we show that the terms $I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}}Y_{\mathrm{D}}\middle|X_{\mathrm{R}},\Lambda\right)$, $I\left(X_{\mathrm{S}},X_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_1},Y_{\mathrm{D}}\middle|X_{\mathrm{R}_1},\Lambda\right)$ achieve $(T-1)\,\gamma_{\mathrm{sr1}}\log(\mathsf{SNR})$ in gDoF; hence they achieve part of the outer bound $\min\{(T-1)\,\gamma_{\mathrm{sr1}}\log(\mathsf{SNR}),(\mathcal{P}_1)\}$ from (3.21). Then we show that the terms $I\left(X_{\mathrm{R}}X_{\mathrm{S}};Y_{\mathrm{D}}\middle|\Lambda\right)$, $I\left(X_{\mathrm{S}},X_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_2},Y_{\mathrm{D}}\middle|X_{\mathrm{R}_2},\Lambda\right)$ can be reduced to the same form as in $(\mathcal{P}_1)$ from (3.21). In the inner bound, after using (3.41), we can optimize over $|c_1|^2,p_\lambda$ to achieve the best rates. We show that this optimization problem is same as the one that appeared in Theorem 3.1 in the calculation of the outer bound. Hence choosing the values of $|c_1|^2,p_\lambda$ from the solution of the outer bound from Table 3.2 and using it in the inner bound, we achieve the gDoF. □

Discussion: The specific choices in Theorem 3.7 are designed to exactly match the terms arising in the inner bound with the terms arising in the outer bound. The time-sharing random variable Λ is chosen to have a cardinality of 2, since the outer bound distribution has 2 mass points (3.24). The scaling is performed at the relays so that the penalty terms $-I\left(Y'_{\mathrm{R}};\hat{Y}_{\mathrm{R}}\middle|X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right)$, $-I\left(Y'_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_1}\middle|X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_2},Y_{\mathrm{D}},\Lambda\right)$ and $-I\left(Y'_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_2}\middle|X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_1},Y_{\mathrm{D}},\Lambda\right)$ do not affect the gDoF. A QMF scheme with Gaussian codebooks without the scaling at the relays does not demonstrate this property as we observe in Remark 3.2 on page 82. We train the channels from the source to the relays using a single training symbol, but we do not train the channels from the relays to the destination. The intuition behind this is that using a single training symbol is gDoF-optimal for a SIMO channel, but using two training symbols is not gDoF-optimal for a MISO channel. This intuition is made more precise in Theorem 3.6. Observing the values of $|c_1|^2,p_\lambda$ from Table 3.2, and the network operation as defined in this section, we see three regimes of relay operation. We can interpret these regimes by recalling that the tradeoff in the cut-set outer bound (tradeoff arises as $\mathcal{P}_1$ in the outer bound (3.21)) is between a MISO cut (Figure 3.8 on page 69) and a parallel cut (Figure 3.9 on page 71). The other cuts are already maximized by our choice of a Gaussian codebook at the source. The tradeoff arises in using Relay R_1 or Relay R_2 because the unknown channel from one of the relays acts as an interference to the transmission from the other relay. The three regimes are described below:

1. If $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} \leq 0$, then the relays operate nonconcurrently, Relay R_1 is ON with probability $1-(\gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}})$ and Relay R_2 is ON with probability $\gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}}$. Note that we already have $\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{rd1}}$, so $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} \leq 0$ implies that $\gamma_{\mathrm{rd2}}, \gamma_{\mathrm{rd1}}$ are quite close to each other in their values. In this case, the nonconcurrent operation ensures maximum gDoF across the MISO cut (see Figure 3.8), by avoiding interference between the relay symbols at the destination. The parallel cut (see Figure 3.9) can match the gDoF across MISO cut even when R_1 is not always ON, since the parallel cut has contribution from γ_{sr2}.

2. If $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} > 0$ and $\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{sr2}} + \gamma_{\mathrm{rd1}}$, then the relays again operate nonconcurrently, Relay R_1 is ON with probability $1-\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$ and Relay R_2 is ON with probability $\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$. Here $\gamma_{\mathrm{rd2}}, \gamma_{\mathrm{rd1}}$ are not close to each other, hence for maximum gDoF across the MISO cut (Figure 3.8), Relay R_2 needs to be always ON. The nonconcurrent operation reduces the gDoF across the MISO cut (Figure 3.8). However, since $\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{sr2}} + \gamma_{\mathrm{rd1}}$, the gDoF across the MISO cut (Figure 3.8) can have a lower value to match the parallel cut (Figure 3.9).

3. If $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} > 0$ and $\gamma_{\mathrm{rd2}} \leq \gamma_{\mathrm{sr2}} + \gamma_{\mathrm{rd1}}$, then both the relays operate simultaneously, but Relay R_1 operates with a reduced power, its transmit power is scaled by $\mathsf{SNR}^{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{sr2}}-\gamma_{\mathrm{rd1}}}$. Here R_2 needs to be always ON to get maximum gDoF value across the MISO cut (Figure 3.8) compared to the parallel cut (Figure 3.9), since $\gamma_{\mathrm{rd2}} \leq \gamma_{\mathrm{sr2}} + \gamma_{\mathrm{rd1}}$. Also Relay R_1 operates at a lower power to reduce interference with Relay R_2. Reducing the power of Relay R_1 reduces the gDoF across the parallel cut (Figure 3.9), but this does not affect the overall gDoF because $\gamma_{\mathrm{rd2}} \leq \gamma_{\mathrm{sr2}} + \gamma_{\mathrm{rd1}}$.

We also note that we can get another set of regimes by relabeling the relays (reversing the roles of the relays in Figure 3.4) and this would reverse the roles of Relay R_1 and and Relay R_2 in the modes of operation.

3.4 Analysis

In this section we provide more detailed analysis for the results stated in the previous section. In Section 3.4.1, we state the mathematical preliminaries required for the analysis. This include the results from previous works. In Section 3.4.2, we give the details required for Theorem 3.4 to loosen the outer bound (3.7) to a different form which can be optimized explicitly; this optimization is detailed in Section 3.4.3. In Section 3.4.4, we analyze the rate achievable for the TS-QMF scheme from Theorem 3.7. A subresult required for the analysis of the TS-QMF scheme is described in Section 3.4.5. The TS-QMF scheme requires the relays to perform a scaling followed by QMF operation. We analyze a point-to-point SISO channel in Section 3.4.5, which has similar structure as the effective relay-to-destination channel.

3.4.1 Mathematical Preliminaries

Fact 3.1. *For exponentially distributed random variable ξ and $a \geq 0, b > 0$, $\log(a + b\mu_\xi) - \gamma \log(e) \leq \mathbb{E}[\log(a + b\xi)] \leq \log(a + b\mu_\xi)$, where γ is the Euler's constant.*

Proof. The details are in Section 4.3 (on page 99). □

Lemma 3.2. *Let $[\xi_1, \xi_2, \ldots, \xi_n]$ be an arbitrary complex random vector and Q be an $n \times n$ unitary isotropic distributed random matrix independent of ξ_i, then*

$$h([\xi_1, \xi_2, \ldots, \xi_n] Q) = h\left(\sum |\xi_i|^2\right) + (n-1)\,\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log(\pi^n / \Gamma(n)). \quad (3.42)$$

Proof. See page 25 for details. □

Remark 3.1. *The above lemma can be applied to $h([\xi_1, 0, \ldots, 0] Q) = h(\xi_1 \bar{q})$, where $\bar{q}$ is an n-dimensional isotropically distributed random unit vector ($\bar{q}$ can be taken as the first row of Q), to obtain $h(\xi_1 \bar{q}) = h\left(|\xi_1|^2\right) + (n-1)\,\mathbb{E}\left[\log\left(|\xi_1|^2\right)\right] + \log(\pi^n / \Gamma(n))$.*

Corollary 3.8. *Let* $[\xi_1, \xi_2, \ldots, \xi_n]$ *be an arbitrary complex random vector,* ξ *be an arbitrary complex random variable and* Q *be an* $n \times n$ *unitary isotropic distributed random matrix independent of* ξ, ξ_i*, then* $h([\xi_1, \xi_2, \ldots, \xi_n] Q | \xi) = h\left(\sum |\xi_i|^2 \middle| \xi\right) + (n-1)\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log(\pi^n / \Gamma(n))$.

Fact 3.2. *For an exponentially distributed random variable* ξ *with mean* μ_ξ *and* $a \geq 0, b > 0$, *we have*

$$\mathbb{E}\left[\frac{b}{b+\xi}\right] = \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \Gamma\left(0, \frac{b}{\mu_\xi}\right) \tag{3.43}$$

and

$$\frac{b}{\mu_\xi} \ln\left(1 + \frac{\mu_\xi}{b}\right) \geq \frac{b}{\mu_\xi} e^{\frac{b}{\mu_\xi}} \Gamma\left(0, \frac{b}{\mu_\xi}\right) \geq \frac{b}{2\mu_\xi} \ln\left(1 + \frac{2\mu_\xi}{b}\right), \tag{3.44}$$

where $\Gamma(0, x)$ *is the incomplete gamma function. Note that* $0 \leq x \ln(1 + 1/x) \leq 1$.

Proof. See page 24 for details. □

3.4.1.1 Chi Squared distribution

We will use the properties of chi-squared distribution in our inner bounds for the noncoherent diamond network. If $w_i \sim \mathcal{CN}(0,1)$ i.i.d., then

$$\sum_{i=1}^{T} |w_i|^2 \sim \frac{1}{2}\chi^2(2T),$$

where $\chi^2(k)$ is chi-squared distributed (which is the sum of squared of k standard normal (real) random variables). Also, $\sqrt{\frac{1}{2}\chi^2(2T)}\overline{q}^{(T)}$ will be a T dimensional random vector with i.i.d. $\mathcal{CN}(0,1)$ components, where $\overline{q}^{(T)}$ is a T dimensional isotropically distributed complex unit vector. Also, we have the entropy

$$h\left(\frac{1}{2}\chi^2(2T)\right) = T + \ln((T-1)!) + (1-T)\psi(T), \tag{3.45}$$

where $\psi()$ is the digamma function which satisfies

$$\ln(T) - \frac{1}{T} < \psi(T) < \ln(T) - \frac{1}{2T}. \tag{3.46}$$

Also, from [Bat08] we have

$$\ln\left(T+\frac{1}{2}\right) < \psi\left(T+1\right) < \ln\left(T+e^{-\gamma}\right). \tag{3.47}$$

Chi-squared distribution is related to Gamma distribution as

$$\chi^2\left(k\right) \sim \Gamma\left(\frac{k}{2},2\right). \tag{3.48}$$

Fact 3.3. *For a chi-squared distributed random variable $\chi^2\left(k\right)$ and $a \geq 0, b > 0$,*

$$\log\left(a+bk\right) - \frac{2\log\left(e\right)}{k} + \log\left(1+\frac{1}{k}\right) \leq \mathbb{E}\left[\log\left(a+b\chi^2\left(k\right)\right)\right] \leq \log\left(a+bk\right). \tag{3.49}$$

Proof. The result is proved in Section 4.3 (on page 99) for the Gamma distribution and the result for the chi-squared distribution follows as a special case. □

Fact 3.4. *For a noncoherent $N \times M$ MIMO channel $Y = GX + W$ with X chosen as $X = LQ$, Q being a $T \times T$ isotropically distributed unitary random matrix, L being an $M \times T$ lower triangular random matrix independent of Q, G being the $N \times M$ random channel matrix with independently distributed Gaussian elements and W being the $N \times T$ random noise matrix with i.i.d. $\mathcal{CN}\left(0,1\right)$ elements, we have:*

$$h\left(Y|X\right) = \sum_{n=1}^{N} h\left(Y\left(n\right)|X\right), \tag{3.50}$$

where $Y\left(n\right)$ is the n^{th} row of Y and

$$h\left(Y\left(n\right)|X\right) = \mathbb{E}\left[\log\left(\det\left(\pi e\left(L^{\dagger} diag\left(\rho^2\left(n\right)\right)L + I_T\right)\right)\right)\right], \tag{3.51}$$

where $\rho^2\left(n\right)$ is the vector of channel strengths to n^{th} receiver antenna (i.e., $\rho^2\left(n\right)$ contains the variance of the elements of the n^{th} row of G) and I_T is the identity matrix of size $T \times T$. Also, for $T > M$, using the lower triangular structure of L with $L_{M\times M}$ being the first $M \times M$ submatrix of L, we have:

$$h\left(Y\left(n\right)|X\right) = \mathbb{E}\left[\log\left(\det\left(\left(L^{\dagger}_{M\times M} diag\left(\rho^2\left(n\right)\right)L_{M\times M} + I_M\right)\right)\right)\right] + T\log\left(\pi e\right), \tag{3.52}$$

where I_M is the identity matrix of size $M \times M$.

Proof. This follows by standard calculations for Gaussian random variables and using the properties of determinants and unitary matrices. See page 27 for details. $\square$

Theorem 3.9. *For the non-coherent SIMO channel* $Y = GX + Z$*, where* X *is the* $1 \times T$ *vector of transmitted symbols,* $G = \mathsf{Tran}\left(\left[\begin{array}{ccc} g_{11} & . & . & g_{N1}\end{array}\right]\right)$*,* $g_{i1} \sim \mathcal{CN}\left(0, \rho_{i1}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{i1}}\right)$*, and* W *being an* $N \times T$ *noise matrix with i.i.d.* $\mathcal{CN}\left(0,1\right)$ *elements, the gDoF is* $\left(1 - 1/T\right)\max_i \gamma_{i1}$*, i.e., the gDoF can be achieved by using only the statistically best receive antenna.*

Proof. See page 18 for details. $\square$

Theorem 3.10. *For the non-coherent MISO channel* $Y = GX+Z$*, where* X *is the* $M \times T$ *vector of transmitted symbols,* $G = \left[\begin{array}{cccc} g_{11} & . & . & g_{1M}\end{array}\right]$*,* $g_{1i} \sim \mathcal{CN}\left(0, \rho_{1i}^2\right) = \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{1i}}\right)$*, and* W *being an* $1 \times T$ *noise vector with i.i.d.* $\mathcal{CN}\left(0,1\right)$ *elements, the gDoF is* $\left(1 - 1/T\right)\max_i \gamma_{1i}$*, i.e., the gDoF can be achieved by only using the statistically best transmit antenna.*

Proof. See page 20 for details. $\square$

3.4.2 Proof of Theorem 3.4

We just need to show $I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right) \dot{\leq} \psi_1$, $I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right) \dot{\leq} \psi_2$ to complete the proof, continuing from the outline of proof given on page 53.

3.4.2.1 An outer bound: $I\left(X_{\mathbf{R}}; Y_{\mathbf{D}}\right) \dot{\leq} \psi_1$

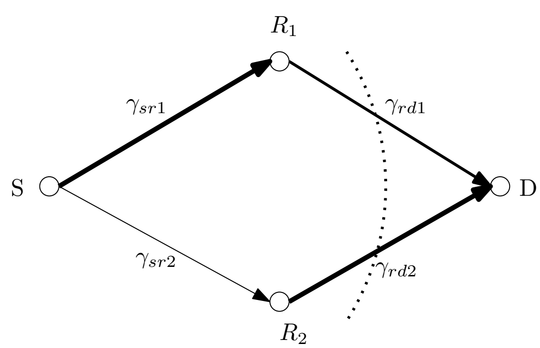


Figure 3.8: The cut corresponding to $I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right)$.

Using the structure of solution from (3.15) on page 53, we have

$$Y_{\mathrm{D}} = \begin{bmatrix} g_{\mathrm{rd1}} & g_{\mathrm{rd2}} \end{bmatrix} \begin{bmatrix} X_{\mathrm{R}_1} \\ X_{\mathrm{R}_2} \end{bmatrix} + W_{\mathrm{D}} \tag{3.53}$$

$$= \begin{bmatrix} g_{\mathrm{rd1}} & g_{\mathrm{rd2}} \end{bmatrix} \begin{bmatrix} b & c & 0 & . & . & . & 0 \\ a & 0 & 0 & . & . & . & 0 \end{bmatrix} Q + W_{\mathrm{D}} \tag{3.54}$$

with $W_{\mathrm{D}} = [w_{\mathrm{d}1}, \ldots, w_{\mathrm{d}T}]$, the elements $w_{\mathrm{d}i}$ being i.i.d. $\mathcal{CN}(0,1)$. Now,

$$h(Y_{\mathrm{D}}) = h\left(\begin{bmatrix} g_{\mathrm{rd1}} & g_{\mathrm{rd2}} \end{bmatrix} \begin{bmatrix} b & c & 0 & . & . & . & 0 \\ a & 0 & 0 & . & . & . & 0 \end{bmatrix} Q + W_{\mathrm{D}}\right) \tag{3.55}$$

$$\overset{(i)}{=} h\left(\left(\begin{bmatrix} g_{\mathrm{rd1}} & g_{\mathrm{rd2}} \end{bmatrix} \begin{bmatrix} b & c & 0 & . & . & . & 0 \\ a & 0 & 0 & . & . & . & 0 \end{bmatrix} + W_{\mathrm{D}}\right) Q\right) \tag{3.56}$$

$$= h\left([a g_{\mathrm{rd2}} + b g_{\mathrm{rd1}} + w_{\mathrm{d1}}, c g_{\mathrm{rd1}} + w_{\mathrm{d2}}, w_{\mathrm{d3}}, \ldots, w_{\mathrm{d}T}]\, Q\right) \tag{3.57}$$

$$\overset{(ii)}{=} h\left(|a g_{\mathrm{rd2}} + b g_{\mathrm{rd1}} + w_{\mathrm{d1}}|^2 + |c g_{\mathrm{rd1}} + w_{\mathrm{d2}}|^2 + \sum_{i=3}^{T} |w_{\mathrm{d}i}|^2\right) \tag{3.58}$$

$$+ (T-1)\,\mathbb{E}\left[\log\left(|a g_{\mathrm{rd2}} + b g_{\mathrm{rd1}} + w_{\mathrm{d1}}|^2 + |c g_{\mathrm{rd1}} + w_{\mathrm{d2}}|^2 + \sum_{i=3}^{T} |w_{\mathrm{d}i}|^2\right)\right]$$

$$+ \log\left(\frac{\pi^T}{\Gamma(T)}\right), \tag{3.59}$$

where (i) is because W_{D} and $W_{\mathrm{D}}Q$ have same the distribution since W_{D} has i.i.d. $\mathcal{CN}(0,1)$ elements and Q is unitary, (ii) is using Lemma 3.2. Now, using (3.52) we can evaluate $h(Y_{\mathrm{D}}|X_{\mathrm{R}})$ to get

$$h(Y_{\mathrm{D}}|X_{\mathrm{R}}) = \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2 \rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1\right)\right] + T \log(\pi e). \tag{3.60}$$

Lemma 3.3. *For any given distribution on (a,b,c), the terms*

$$h\left(|a g_{rd2} + b g_{rd1} + w_{d1}|^2 + |c g_{rd1} + w_{d2}|^2 + \sum_{i=3}^{T} |w_{di}|^2\right)$$

and

$$\mathbb{E}\left[\log\left(\rho_{rd2}^2 |a|^2 + \rho_{rd1}^2 |b|^2 + \rho_{rd1}^2 |c|^2 + T\right)\right]$$

have the same gDoF.

Proof. The proof is in Appendix A.7. □

The following two corollaries follow similar to the above lemma, we omit the proof.

Corollary 3.11. *For any given distribution on* (a,b,c)*, the terms* $h\left(ag_{rd2}+bg_{rd1}+w_{d1}\right)$*,* $h\left(ag_{rd2}+bg_{rd1}+w_{d1}\right|a)$*,* $h\left(\left|ag_{rd2}+bg_{rd1}+w_{d1}\right|^2\right)$*,* $\mathbb{E}\left[\log\left(\rho_{rd2}^2\left|a\right|^2+\rho_{rd1}^2\left|b\right|^2+1\right)\right]$*, all have the same gDoF.*

Corollary 3.12. *For any given distribution on* (a,b,c)*, the terms* $h\left(\left|cg_{rd1}+w_{d2}\right|^2+\sum_{i=3}^{T}\left|w_{di}\right|^2\right)$*,* $h\left(\left|cg_{rd1}+w_{d2}\right|^2+\sum_{i=3}^{T}\left|w_{di}\right|^2\middle|a\right)$*,* $\mathbb{E}\left[\log\left(\rho_{rd1}^2\left|c\right|^2+T-2\right)\right]$*, all have the same gDoF.*

Note that

$$\begin{aligned}&\mathbb{E}\left[\log\left(\left|ag_{\mathrm{rd2}}+bg_{\mathrm{rd1}}+w_{\mathrm{d1}}\right|^2+\left|cg_{\mathrm{rd1}}+w_{\mathrm{d2}}\right|^2+\sum_{i=3}^{T}\left|w_{\mathrm{d}i}\right|^2\right)\right]\\&\doteq\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2\left|a\right|^2+\rho_{\mathrm{rd1}}^2\left|b\right|^2+\rho_{\mathrm{rd1}}^2\left|c\right|^2+T\right)\right]\end{aligned}\tag{3.61}$$

using the Tower property of expectation [WHH05, pp. 380-383] and Fact 5.1. Hence using Lemma 3.3 and the above equation, we get

$$\begin{aligned}I\left(X_{\mathrm{R}};Y_{\mathrm{D}}\right)\doteq&T\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2\left|a\right|^2+\rho_{\mathrm{rd1}}^2\left|b\right|^2+\rho_{\mathrm{rd1}}^2\left|c\right|^2+T\right)\right]\\&-\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2\left|a\right|^2+\rho_{\mathrm{rd1}}^2\left|b\right|^2+\rho_{\mathrm{rd1}}^2\left|c\right|^2+\rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2\left|c\right|^2\left|a\right|^2+1\right)\right]\\=&\psi_1.\end{aligned}\tag{3.62}$$

3.4.2.2 An outer bound: $I\left(X_{\mathbf{S}};Y_{\mathbf{R}_2}\right)+I\left(X_{\mathbf{R}_1};Y_{\mathbf{D}}\middle|X_{\mathbf{R}_2}\right)\dot{\leq}\psi_2$

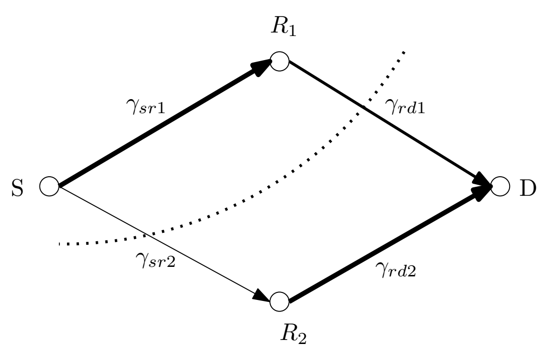


Figure 3.9: The cut corresponding to $I\left(X_{\mathrm{S}};Y_{\mathrm{R}_2}\right)+I\left(X_{\mathrm{R}_1};Y_{\mathrm{D}}\middle|X_{\mathrm{R}_2}\right)$.

We have

$$I\left(X_{\mathrm{S}};Y_{\mathrm{R}_2}\right)\dot{\leq}(T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) \tag{3.63}$$

due to the DoF results for the noncoherent SISO channel [ZT02]. Now,

$$\begin{aligned}
& I\left(X_{\mathrm{R}_1};Y_{\mathrm{D}}\middle|X_{\mathrm{R}_2}\right)\\
& = h\left(Y_{\mathrm{D}}\middle|X_{\mathrm{R}_2}\right)-h\left(Y_{\mathrm{D}}\middle|X_{\mathrm{R}_1},X_{\mathrm{R}_2}\right) && (3.64)\\
& \stackrel{(i)}{=} h\left(\left[\begin{array}{cc} g_{\mathrm{rd2}} & g_{\mathrm{rd1}}\end{array}\right]\left[\begin{array}{cccccc} a & 0 & 0 & . & . & 0\\ b & c & 0 & . & . & 0\end{array}\right]Q+W_{\mathrm{D}}\middle|\left[\begin{array}{cccccc} a & 0 & 0 & . & . & 0\end{array}\right]Q\right)-T\log(\pi e)\\
& \quad -\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2|a|^2+\rho_{\mathrm{rd1}}^2|b|^2+\rho_{\mathrm{rd1}}^2|c|^2+\rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2|c|^2|a|^2+1\right)\right], && (3.65)
\end{aligned}$$

where (i) was using the structure of $X_{\mathrm{R}_1},X_{\mathrm{R}_2}$ from (3.15) on page 53 and by evaluating $h\left(Y_{\mathrm{D}}\middle|X_{\mathrm{R}_1},X_{\mathrm{R}_2}\right)$ using (3.52). Now,

$$\begin{aligned}
& h\left(\left[\begin{array}{cc} g_{\mathrm{rd2}} & g_{\mathrm{rd1}}\end{array}\right]\left[\begin{array}{cccccc} a & 0 & 0 & . & . & 0\\ b & c & 0 & . & . & 0\end{array}\right]Q+W_{\mathrm{D}}\middle|\left[\begin{array}{cccccc} a & 0 & 0 & . & . & 0\end{array}\right]Q\right)\\
& \stackrel{(i)}{=} h\left(\left[\begin{array}{cc} g_{\mathrm{rd2}} & g_{\mathrm{rd1}}\end{array}\right]\left[\begin{array}{cccccc} a & 0 & 0 & . & . & 0\\ b & c & 0 & . & . & 0\end{array}\right]\left[\begin{array}{cc} 1 & 0\\ 0 & Q_{T-1}\end{array}\right]+W_{\mathrm{D}}\middle|a\right) && (3.66)\\
& = h\left(bg_{\mathrm{rd1}}+ag_{\mathrm{rd2}}+w_{\mathrm{d1}},\left[\begin{array}{ccccc} g_{\mathrm{rd1}}c & 0 & . & . & 0\end{array}\right]Q_{T-1}+W_{\mathrm{D},T-1}\middle|a\right) && (3.67)\\
& \stackrel{(ii)}{\leq} h\left(bg_{\mathrm{rd1}}+ag_{\mathrm{rd2}}+w_{\mathrm{d1}}\middle|a\right)+h\left(\left(\left[\begin{array}{ccccc} g_{\mathrm{rd1}}c & 0 & . & . & 0\end{array}\right]+W_{\mathrm{D},T-1}\right)Q_{T-1}\middle|a\right) && (3.68)\\
& \stackrel{(iii)}{\doteq} h\left(bg_{\mathrm{rd1}}+ag_{\mathrm{rd2}}+w_{\mathrm{d1}}\middle|a\right)+h\left(\left|g_{\mathrm{rd1}}c+w_{\mathrm{d2}}\right|^2+\sum_{i=3}^{T}\left|w_{\mathrm{d}i}\right|^2\middle|a\right)\\
& \quad +(T-2)\,\mathbb{E}\left[\log\left(\rho_{\mathrm{rd1}}^2|c|^2+T-1\right)\right] && (3.69)\\
& \stackrel{(iv)}{\doteq} \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2|a|^2+\rho_{\mathrm{rd1}}^2|b|^2+1\right)\right]+(T-1)\,\mathbb{E}\left[\log\left(\rho_{\mathrm{rd1}}^2|c|^2+T-1\right)\right] && (3.70)
\end{aligned}$$

(i) is because by conditioning on $\left[\begin{array}{ccccccc} a & 0 & 0 & . & . & . & 0\end{array}\right]Q$ the first row of Q is known and hence the entropy is evaluated after projecting onto a new orthonormal basis with first basis vector chosen as the first row of Q. The random matrix Q_{T-1} is unitary isotropically distributed in $T-1$ dimensions. Since W_{D} has i.i.d. elements, after this projection, the distribution of W_{D} remains same. The random vector $W_{\mathrm{D},T-1}$ is $T-1$ dimensional with i.i.d. $\mathcal{CN}(0,1)$ elements. The step (ii) is using the fact that conditioning reduces entropy and the fact that $W_{\mathrm{D},T-1}$ has the same distribution as $W_{\mathrm{D},T-1}Q_{T-1}$. The step (iii) is using Lemma 3.2 on $h\left(\left(\begin{bmatrix} g_{\mathrm{rd1}}c & 0 & . & . & 0 \end{bmatrix} + W_{\mathrm{D},T-1}\right) Q_{T-1} \middle| a\right)$ and (iv) is using Corollary 3.11 and Corollary 3.12. Hence we get

$$\begin{aligned}
& I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; Y_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right) \\
& \dot{\leq} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + 1\right)\right] \\
& \quad + (T-1)\,\mathbb{E}\left[\log\left(\rho_{\mathrm{rd1}}^2 |c|^2 + T - 1\right)\right] \\
& \quad - \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1\right)\right] && (3.71) \\
& = \psi_2. && (3.72)
\end{aligned}$$

3.4.3 Solving the Outer Bound Optimization Problem

For the outer bound we have the optimization program:

$$\mathcal{P}_9 : \begin{cases}
\text{maximize } \min \Big\{ p_\lambda \left((T-1)\gamma_{\mathrm{rd2}}\log(\mathsf{SNR}) - \log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd1}}} |c_1|^2 + 1\right)\right) \\
\qquad + (T-1)(1-p_\lambda)\gamma_{\mathrm{rd1}}\log(\mathsf{SNR}),\ (T-1)\gamma_{\mathrm{sr2}}\log(\mathsf{SNR}) \\
\qquad + (T-2)p_\lambda \log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd1}}} |c_1|^2 + 1\right) \\
\qquad + (T-1)(1-p_\lambda)\gamma_{\mathrm{rd1}}\log(\mathsf{SNR}) \Big\} \\
|c_1|^2 \leq T, 0 \leq p_\lambda \leq 1,
\end{cases} \tag{3.73}$$

with

$$\mathrm{gDoF}\left(\mathcal{P}_1\right) = \mathrm{gDoF}\left(\mathcal{P}_9\right) \tag{3.74}$$

due to Lemma 3.1 on page 55 and $\mathcal{P}_1$ is defined in Theorem 3.4 on page 53. Now, with $|c_1|^2 \leq 2T$, we have $0 \leq \log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd1}}} |c_1|^2 + 1\right) \dot{\leq} \gamma_{\mathrm{rd1}} \log(SNR)$. So we change variables by letting $\log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd1}}}|c_1|^2\right)=\gamma_c\log(\mathsf{SNR})$ to get

$$\mathcal{P}_{10}:\begin{cases}\text{maximize}\ \min\big\{p_\lambda\left((T-1)\gamma_{\mathrm{rd2}}-\gamma_c\right)+(T-1)(1-p_\lambda)\gamma_{\mathrm{rd1}},\\ \qquad\qquad (T-1)\gamma_{\mathrm{sr2}}+(T-2)p_\lambda\gamma_c+(T-1)(1-p_\lambda)\gamma_{\mathrm{rd1}}\big\}\\ 0\le\gamma_c\le\gamma_{\mathrm{rd1}},0\le p_\lambda\le 1\end{cases}\tag{3.75}$$

with

$$\mathrm{gDoF}(\mathcal{P}_1)=\mathrm{gDoF}(\mathcal{P}_9)=(\mathcal{P}_{10}).\tag{3.76}$$

Note that we removed scaling by $\log(\mathsf{SNR})$ in $\mathcal{P}_{10}$, so its solution directly yields the gDoF. Following (3.24) on page 56, now $\mathcal{P}_{10}$ gives two mass points for $\left(|a|^2,|b|^2,|c|^2\right)$ for optimal gDoF

$$\left(|a|^2,|b|^2,|c|^2\right)=\begin{cases}\left(T,0,|c_1|^2\right)=\left(T,0,\mathsf{SNR}^{\gamma_c-\gamma_{\mathrm{rd1}}}\right) & \text{w.p. } p_\lambda\\ (0,T/2,T/2) & \text{w.p. } 1-p_\lambda.\end{cases}\tag{3.77}$$

Now $\mathcal{P}_{10}$ is a bilinear optimization problem which we solve explicitly.

3.4.3.1 Solving the bilinear problem

We collect the terms in $\mathcal{P}_{10}$ to rewrite it as

$$\mathcal{P}_{10}:\begin{cases}\text{maximize}\ \min\big\{p_\lambda\left((T-1)(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}})-\gamma_c\right)+(T-1)\gamma_{\mathrm{rd1}},\\ \qquad\qquad (T-1)\gamma_{\mathrm{sr2}}+p_\lambda\left((T-2)\gamma_c-(T-1)\gamma_{\mathrm{rd1}}\right)+(T-1)\gamma_{\mathrm{rd1}}\big\}\\ 0\le\gamma_c\le\gamma_{\mathrm{rd1}},0\le p_\lambda\le 1\end{cases}\tag{3.78}$$

Note that $(T-2)\gamma_c-(T-1)\gamma_{\mathrm{rd1}}<0$. Hence $(T-1)\gamma_{\mathrm{sr2}}+p_\lambda\left((T-2)\gamma_c-(T-1)\gamma_{\mathrm{rd1}}\right)+(T-1)\gamma_{\mathrm{rd1}}$ is decreasing in p_λ. If $(T-1)(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}})-\gamma_c<0$ both the terms inside $\min()$ are decreasing with p_λ and hence the optimal would be achieved at $p_\lambda=0$. However, this value can be achieved in the regime $(T-1)(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}})-\gamma_c\ge 0$ with $p_\lambda=0$ for any value of γ_c. (See Figure 3.10).

Hence it suffices to consider the regime

$$(T-1)(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}})-\gamma_c\ge 0\tag{3.79}$$

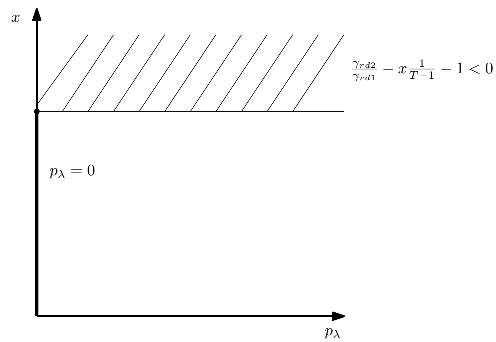


Figure 3.10: For the objective function from (3.78), the regime $(T-1)(\gamma_{\text{rd}2} - \gamma_{\text{rd}1}) - \gamma_c < 0$ is dominated by the line $p_\lambda = 0$.

in $\mathcal{P}_{10}$. In this regime, examining the two terms within the min of $\mathcal{P}_{10}$, $(T-1)\gamma_{\text{sr}2} + p_\lambda((T-2)\gamma_c - (T-1)\gamma_{\text{rd}1}) + (T-1)\gamma_{\text{rd}1}$ is decreasing and $p_\lambda((T-1)(\gamma_{\text{rd}2} - \gamma_{\text{rd}1}) - \gamma_c) + (T-1)\gamma_{\text{rd}1}$ is increasing, as a function of p_λ. Hence the maxmin in terms of p_λ is achieved at the intersection point, if that point is within $[0,1]$. (See Figure 3.11). The intersection point is determined by $(T-1)\gamma_{\text{sr}2} + p_\lambda((T-2)\gamma_c - (T-1)\gamma_{\text{rd}1}) + (T-1)\gamma_{\text{rd}1} = p_\lambda((T-1)(\gamma_{\text{rd}2} - \gamma_{\text{rd}1}) - \gamma_c) + (T-1)\gamma_{\text{rd}1}$, which gives the intersection point to be

$$p'_\lambda = \frac{\gamma_{\text{sr}2}}{\gamma_{\text{rd}2} - \gamma_c}. \tag{3.80}$$

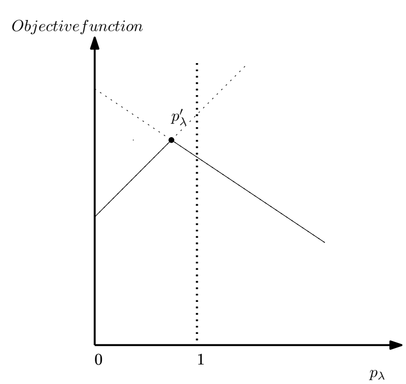


Figure 3.11: Behavior of the bilinear program from (3.78) as a function of p_λ for any $\gamma_c \leq (T-1)(\gamma_{\text{rd}2} - \gamma_{\text{rd}1})$.

Now, we claim that it is sufficient to consider the regime $p'_\lambda \leq 1 \iff \gamma_{\text{sr}2}/(\gamma_{\text{rd}2} - \gamma_c) \leq 1 \iff \gamma_c \leq \gamma_{\text{rd}2} - \gamma_{\text{sr}2}$. Otherwise $p'_\lambda > 1 \iff \gamma_c > \gamma_{\text{rd}2} - \gamma_{\text{sr}2}$, and in this regime, the maxmin in terms of p_λ is achieved by $p_\lambda = 1$ (see Figure 3.12), and the maxmin value is given by $1 \times ((T-1)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}) - \gamma_c) + (T-1)\gamma_{\mathrm{rd1}} = (T-1)\gamma_{\mathrm{rd2}} - \gamma_c$. But a greater

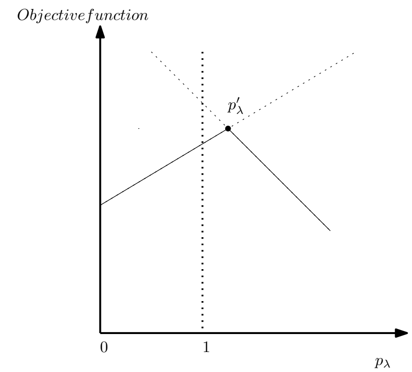


Figure 3.12: Behavior of the bilinear program from (3.78) as a function of p_λ when $p'_\lambda > 1$.

value can be achieved by choosing $\gamma_c = \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}$ (instead of $\gamma_c > \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}$) at $p_\lambda = 1$, and that value is given by $(T-1)\gamma_{\mathrm{rd2}} - (\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}})$. Hence it suffices to consider the regime with

$$\gamma_c \leq \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}. \tag{3.81}$$

Now, using the extra constraints (3.79), (3.81) and substituting the optimal $p'_\lambda = \gamma_{\mathrm{sr2}} / (\gamma_{\mathrm{rd2}} - \gamma_c)$ in $\mathcal{P}_{10}$ (3.78), we get the equivalent problem

$$\underset{0 \leq \gamma_c \leq (T-1)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}), \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}, \gamma_{\mathrm{rd1}}}{\text{maximize}} \quad (T-1)\gamma_{\mathrm{sr2}} + \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}} - \gamma_c}\left((T-2)\gamma_c - (T-1)\gamma_{\mathrm{rd1}}\right) + (T-1)\gamma_{\mathrm{rd1}}. \tag{3.82}$$

Now it can be verified that

$$\begin{aligned} &\frac{d}{d\gamma_c}\left[\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}} - \gamma_c}\left((T-2)\gamma_c - (T-1)\gamma_{\mathrm{rd1}}\right)\right] \\ &= \frac{\gamma_{\mathrm{sr2}}}{(\gamma_{\mathrm{rd2}} - \gamma_c)^2}\left((T-2)\gamma_{\mathrm{rd2}} - (T-1)\gamma_{\mathrm{rd1}}\right). \end{aligned}$$

Hence if $(T-2)\gamma_{\mathrm{rd2}} - (T-1)\gamma_{\mathrm{rd1}} \leq 0$, the maximum in (3.82) is achieved at $\gamma_c = 0$, otherwise the maximum is achieved at $\gamma_c = \min\{\gamma_{\mathrm{rd1}}, (T-1)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}), \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}\}$. With the following claim, we show that if $(T-2)\gamma_{\mathrm{rd2}} - (T-1)\gamma_{\mathrm{rd1}} > 0$, then $\min\{\gamma_{\mathrm{rd1}}, (T-1)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}), \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}\}$ is same as $\min\{\gamma_{\mathrm{rd1}}, \gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}\}$.

Claim 3.1. *If $(T-2)\gamma_{rd2} - (T-1)\gamma_{rd1} > 0$, then $\min\{\gamma_{rd1}, (T-1)(\gamma_{rd2} - \gamma_{rd1}), \gamma_{rd2} - \gamma_{sr2}\} = \min\{\gamma_{rd1}, \gamma_{rd2} - \gamma_{sr2}\}$*

Proof. To prove this, it suffices to show that $(T-1)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}) > \gamma_{\mathrm{rd1}}$. We have

$$\begin{aligned}
&(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} > 0 \\
\Leftrightarrow\ &(T-1)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} > \gamma_{\mathrm{rd2}} \\
\Leftrightarrow\ &(T-1)\,(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{rd1}}) > \gamma_{\mathrm{rd2}}
\end{aligned}$$

Now the required result follows, because $\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{rd1}}$ in the regime under consideration (see Figure 3.4). □

Now, we go through the different regimes that give different solutions.

3.4.3.2 $(T-2)\,\gamma_{\mathbf{rd2}} - (T-1)\,\gamma_{\mathbf{rd1}} \leq 0$

In this case maximum is achieved at

$$\gamma_c^* = 0,\ p_\lambda^* = \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}} - \gamma_c^*} = \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}}. \tag{3.83}$$

Hence following (3.77), we have the solution $\left(T, 0, |c_1|^2\right) = \left(T, 0, \mathsf{SNR}^{-\gamma_{\mathrm{rd1}}}\right)$ with probability $p_\lambda = \gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}}$ and $(0, T/2, T/2)$ with probability $(1-p_\lambda) = 1 - \gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}}$. Effectively we can choose $(T, 0, 0)$ (since $|c_1|^2 = \mathsf{SNR}^{-\gamma_{\mathrm{rd1}}}$ causes the link g_{rd1} to contribute zero gDoF) with probability $p_\lambda = \gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}}$ and $(0, T/2, T/2)$ with probability $(1-p_\lambda) = 1 - \gamma_{\mathrm{sr2}}/\gamma_{\mathrm{rd2}}$. Note that this regime with $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} \leq 0$ disappears as $T \to \infty$, since we already have $\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{rd1}}$ ($\gamma_{\mathrm{rd2}} > \gamma_{\mathrm{rd1}}$ from the description in Section 3.3.1). Following (3.77) we tabulate the optimal distribution for (a, b, c) in Table 3.3.

Table 3.3: Solution with $(T-2)\,\gamma_{\mathrm{rd2}} - (T-1)\,\gamma_{\mathrm{rd1}} \leq 0$

(a, b, c)	Probability
$\left(\sqrt{T}, 0, 0\right)$	$p_\lambda = \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}}$
$\left(0, \sqrt{T/2}, \sqrt{T/2}\right)$	$(1-p_\lambda) = 1 - \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}}$

In this case, we calculate the gDoF by substituting the solution in (3.82) and scaling with

$1/T$. The outer bound gDoF for the network is

$$\frac{1}{T}\left((T-1)\gamma_{\mathrm{sr2}}+\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}}\left(-(T-1)\gamma_{\mathrm{rd1}}\right)+(T-1)\gamma_{\mathrm{rd1}}\right)$$
$$=\left(1-\frac{1}{T}\right)\left(\gamma_{\mathrm{sr2}}+\gamma_{\mathrm{rd1}}-\frac{\gamma_{\mathrm{sr2}}\gamma_{\mathrm{rd1}}}{\gamma_{\mathrm{rd2}}}\right).$$

3.4.3.3 $(T-2)\gamma_{\mathbf{rd2}}-(T-1)\gamma_{\mathbf{rd1}}>0$

In this case the optimal value is achieved by

$$\gamma_c^*=\min\left\{\gamma_{\mathrm{rd1}},\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{sr2}}\right\},\ p_\lambda^*=\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}-\gamma_c^*}. \tag{3.84}$$

Case 1, $\gamma_c^*=\gamma_{\mathbf{rd1}}=\min\left\{\gamma_{\mathbf{rd1}},\gamma_{\mathbf{rd2}}-\gamma_{\mathbf{sr2}}\right\}$ We have the solution

$$\left(T,0,|c_1|^2\right)=\left(T,0,\mathsf{SNR}^{\gamma_c-\gamma_{\mathrm{rd1}}}\right)=(T,0,1) \tag{3.85}$$

with probability $p_\lambda=\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$ and $(0,T/2,T/2)$ with probability $(1-p_\lambda)=1-\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$. For gDoF, we can effectively have $(T,0,T)$ with probability $p_\lambda=\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$ and $(0,T/2,T/2)$ with probability $(1-p_\lambda)=1-\gamma_{\mathrm{sr2}}/\left(\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}\right)$. The result is tabulated in Table 3.4.

Table 3.4: Solution for case 1

(a,b,c)	Probability
$\left(\sqrt{T},0,\sqrt{T}\right)$	$p_\lambda=\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}}$
$\left(0,\sqrt{T/2},\sqrt{T/2}\right)$	$(1-p_\lambda)=1-\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}}$

By substituting the solution in (3.82) and scaling with $1/T$, the outer bound gDoF for the network in this case is

$$\frac{1}{T}\left((T-1)\gamma_{\mathrm{sr2}}+\frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}}\left((T-2)\gamma_{\mathrm{rd1}}-(T-1)\gamma_{\mathrm{rd1}}\right)+(T-1)\gamma_{\mathrm{rd1}}\right)$$
$$=\left(1-\frac{1}{T}\right)\left(\gamma_{\mathrm{sr2}}+\gamma_{\mathrm{rd1}}\right)-\left(\frac{1}{T}\right)\frac{\gamma_{\mathrm{sr2}}\gamma_{\mathrm{rd1}}}{\gamma_{\mathrm{rd2}}-\gamma_{\mathrm{rd1}}}$$

Case 2, $\gamma_c^* = \gamma_{\mathbf{rd2}} - \gamma_{\mathbf{sr2}} = \min\{\gamma_{\mathbf{rd1}}, \gamma_{\mathbf{rd2}} - \gamma_{\mathbf{sr2}}\}$ With this value of x^* we get the point

$$\left(T, 0, |c_1|^2\right) = \left(T, 0, \mathsf{SNR}^{\gamma_c - \gamma_{\mathrm{rd1}}}\right) = \left(T, 0, \mathsf{SNR}^{\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}} - \gamma_{\mathrm{rd1}}}\right) \tag{3.86}$$

with probability $p_\lambda = \gamma_{\mathrm{sr2}} / \left(\gamma_{\mathrm{rd2}} - \gamma_c^*\right) = 1$. The result is tabulated in Table 3.5.

Table 3.5: Solution for case 2

(a, b, c)	Probability
$\left(\sqrt{T}, 0, \sqrt{\mathsf{SNR}^{\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}} - \gamma_{\mathrm{rd1}}}}\right)$	$p_\lambda = 1$

By substituting the solution in (3.82) and scaling with $1/T$, the outer bound gDoF for the network in this case is

$$\begin{aligned}
&\frac{1}{T}\left((T-1)\gamma_{\mathrm{sr2}} + \frac{\gamma_{\mathrm{sr2}}}{\gamma_{\mathrm{rd2}} - (\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}})}\left((T-2)(\gamma_{\mathrm{rd2}} - \gamma_{\mathrm{sr2}}) - (T-1)\gamma_{\mathrm{rd1}}\right) + (T-1)\gamma_{\mathrm{rd1}}\right) \\
&= \frac{1}{T}\gamma_{\mathrm{sr2}} + \left(1 - \frac{2}{T}\right)(\gamma_{\mathrm{rd2}}).
\end{aligned}$$

3.4.4 Achievability Scheme

Here we discuss the gDoF-optimality of our achievability scheme. We analyze the rate expression

$$\begin{aligned}
TR = \min\Big\{ & I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}} Y_{\mathrm{D}} \middle| X_{\mathrm{R}}, \Lambda\right), I\left(X_{\mathrm{R}} X_{\mathrm{S}}; Y_{\mathrm{D}} \middle| \Lambda\right) - I\left(Y'_{\mathrm{R}}; \hat{Y}_{\mathrm{R}} \middle| X_{\mathrm{S}}, X_{\mathrm{R}}, Y_{\mathrm{D}}, \Lambda\right), \\
& I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) - I\left(Y'_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_1} \middle| X_{\mathrm{S}}, X_{\mathrm{R}}, \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}}, \Lambda\right), \\
& I\left(X_{\mathrm{S}}, X_{\mathrm{R}_2}; \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}, \Lambda\right) - I\left(Y'_{\mathrm{R}_2}; \hat{Y}_{\mathrm{R}_2} \middle| X_{\mathrm{S}}, X_{\mathrm{R}}, \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}}, \Lambda\right) \Big\}
\end{aligned} \tag{3.87}$$

from (3.33) arising out of the QMF decoding.

We first note that there is penalty of the form $I\left(Y'_{\mathrm{R}}; \hat{Y}_{\mathrm{R}} \middle| X_{\mathrm{S}}, X_{\mathrm{R}}, Y_{\mathrm{D}}, \Lambda\right)$ in the rate expression (3.87). The following theorem helps to show that the penalty does not contribute to a penalty in gDoF, while still having the terms of the form $I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}} Y_{\mathrm{D}} \middle| X_{\mathrm{R}}, \Lambda\right)$ which roughly behaves as $I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}}\right)$ to achieve full gDoF.

Theorem 3.13. *Let $Y = gX + W$, with X being a vector of length $(T-1)$ with i.i.d. $\mathcal{CN}(0,1)$ elements and W also being a vector of length $(T-1)$ with i.i.d. $\mathcal{CN}(0,1)$ elements and $g \sim \mathcal{CN}(0,\rho^2)$. We define a scaled version of Y as $Y' = \frac{g}{\hat{g}}X + \frac{W}{\hat{g}}$ with*

$$\hat{g} = e^{i\angle(g+w)} + (g+w), \tag{3.88}$$

where $\angle(g+w')$ is the angle of $g+w'$ with $w' \sim \mathcal{CN}(0,1)$. And Y' is quantized to $\hat{Y} = \frac{g}{\hat{g}}X + \frac{W}{\hat{g}} + Q$ with $Q \sim \frac{W}{\hat{g}}$. With this setting we claim:

$$I\left(\hat{Y};X\right) \dot{\geq} (T-1)\log\left(\rho^2\right) \tag{3.89}$$

and

$$I\left(\hat{Y};Y'\middle|X\right) \dot{\leq} 0 \tag{3.90}$$

and hence $I\left(\hat{Y};X\right) - I\left(\hat{Y};Y'\middle|X\right) \dot{\geq} (T-1)\log\left(\rho^2\right)$.

Proof. The proof is in Section 3.4.5. □

Also, due to the relay operation described in Section 3.3.2 (see also Figure 3.7 on page 62), the relays-to-destination channel behaves like a MISO channel with independently distributed symbols from the transmit antennas. In the following theorem, we analyze an entropy expression arising from such a channel.

Theorem 3.14. *For a MISO channel $Y = \left[\begin{array}{cc} g_{11} & g_{12} \end{array}\right] X + W_{1\times T}$ with $g_{11} \sim \mathcal{CN}(0,\rho_{11}^2)$, $g_{12} \sim \mathcal{CN}(0,\rho_{12}^2)$, $W_{1\times T}$ being a $1\times T$ vector with i.i.d. $\mathcal{CN}(0,1)$ elements and X chosen as*

$$X = \left[\begin{array}{c} a_1X_1 \\ a_2X_2 \end{array}\right], \tag{3.91}$$

where X_1 and X_2 are $1\times T$ vectors with i.i.d. $\mathcal{CN}(0,1)$ elements, we have

$$h\left(Y|X\right) \dot{\leq} \log\left(\left(1+\rho_{11}^2\left|a_1\right|^2\right)\left(1+\rho_{21}^2\left|a_2\right|^2\right)\right). \tag{3.92}$$

Proof. See Appendix B.5. □

Now, we analyze the penalty terms from the rate expression of (3.87). We first look at the term $I\left(Y'_{\mathrm{R}};\hat{Y}_{\mathrm{R}}\middle| X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right)$.

$$\begin{aligned}
I\left(Y'_{\mathrm{R}};\hat{Y}_{\mathrm{R}}\middle| X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right) &= h\left(\hat{Y}_{\mathrm{R}}\middle| X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right)-h\left(\hat{Y}_{\mathrm{R}}\middle| Y'_{\mathrm{R}},X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right) && (3.93)\\
&\overset{(i)}{\leq} h\left(\hat{Y}_{\mathrm{R}}\middle| X_{\mathrm{S}}\right)-h\left(\hat{Y}_{\mathrm{R}}\middle| Y'_{\mathrm{R}},X_{\mathrm{S}},X_{\mathrm{R}},Y_{\mathrm{D}},\Lambda\right) && (3.94)\\
&\overset{(ii)}{=} h\left(\hat{Y}_{\mathrm{R}}\middle| X_{\mathrm{S}}\right)-h\left(\left[Q_{\mathrm{R}_1},Q_{\mathrm{R}_2}\right]\right) && (3.95)\\
&\overset{(iii)}{=} h\left(\hat{Y}_{\mathrm{R}_1}\middle| X_{\mathrm{S}}\right)-h\left(Q_{\mathrm{R}_1}\right)+h\left(\hat{Y}_{\mathrm{R}_2}\middle| X_{\mathrm{S}}\right)-h\left(Q_{\mathrm{R}_2}\right) && (3.96)\\
&= I\left(Y'_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_1}\middle| X_{\mathrm{S}}\right)+I\left(Y'_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_2}\middle| X_{\mathrm{S}}\right) && (3.97)\\
&\overset{(iv)}{\dot{\leq}} 0, && (3.98)
\end{aligned}$$

where (i) is using the fact that conditioning reduces entropy, (ii) is because of the choice of the quantizer (3.39),(3.40) with quantization noise independent of the other random variables, (iii) is because $Q_{\mathrm{R}_1},Q_{\mathrm{R}_1}$ are independent of each other and $Y_{\mathrm{R}_2},Y_{\mathrm{R}_1}$ are independent of each other given X_{S}, and (iv) is using (3.89) from Theorem 3.13.

Similarly $I\left(Y'_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_1}\middle| X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_2},Y_{\mathrm{D}},\Lambda\right)\dot{\leq}0$ and $I\left(Y'_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_2}\middle| X_{\mathrm{S}},X_{\mathrm{R}},\hat{Y}_{\mathrm{R}_1},Y_{\mathrm{D}},\Lambda\right)\dot{\leq}0$. Hence for our scheme, the rates

$$\begin{aligned}
TR\,\dot{\leq}\,\min\Big\{ & I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}}Y_{\mathrm{D}}\middle| X_{\mathrm{R}},\Lambda\right), I\left(X_{\mathrm{R}}X_{\mathrm{S}};Y_{\mathrm{D}}\middle|\Lambda\right),\\
& I\left(X_{\mathrm{S}},X_{\mathrm{R}_1};\hat{Y}_{\mathrm{R}_2},Y_{\mathrm{D}}\middle| X_{\mathrm{R}_2},\Lambda\right), I\left(X_{\mathrm{S}},X_{\mathrm{R}_2};\hat{Y}_{\mathrm{R}_1},Y_{\mathrm{D}}\middle| X_{\mathrm{R}_1},\Lambda\right)\Big\} \qquad (3.99)
\end{aligned}$$

are achievable.

Remark 3.2. *For a standard QMF scheme [ADT11] with Gaussian codebooks without training and scaling, we can calculate that the penalty terms of the form* $I\left(Y'_{R};\hat{Y}_{R}\middle| X_{S},X_{R},Y_{D},\Lambda\right)$ *cause a loss in gDoF for the noncoherent diamond network. To understand this with a simple example, consider* $Y=gX+W$ *with* X *being a vector of length* T *with i.i.d.* $\mathcal{CN}(0,1)$ *elements and* W *being a vector of length* T *with i.i.d.* $\mathcal{CN}(0,1)$ *elements,* $g\sim\mathcal{CN}(0,\rho^2)$ *and* Y *is quantized to* $\hat{Y}=gX+W+Q$ *with* $Q\sim W$*. Then in this case, it can be easily verified that* $I\left(\hat{Y};X\right)\,\dot{\geq}\,(T-1)\log(\rho^2)$*, but* $I\left(\hat{Y};Y\middle| X\right)=h\left(Y|X\right)-h\left(Y|\hat{Y},X\right)\doteq\log(\rho^2)$

(in contrast to (3.90) for our scheme). However, we demonstrate that for our scheme with training, scaling and using Gaussian codebooks, the penalty terms do not affect gDoF.

Now, we simplify the four terms in (3.99). The first term

$$
\begin{aligned}
I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}}Y_{\mathrm{D}}\middle|X_{\mathrm{R}},\Lambda\right) &\geq I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}}\middle|X_{\mathrm{R}},\Lambda\right) && (3.100)\\
&\overset{(i)}{=} I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}}\right) && (3.101)\\
&\geq \max\left\{I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}_1}\right), I\left(X_{\mathrm{S}};\hat{Y}_{\mathrm{R}_2}\right)\right\} && (3.102)\\
&\overset{(ii)}{\dot{\geq}} (T-1)\max\left\{\log\left(\rho_{\mathrm{sr1}}^2\right),\log\left(\rho_{\mathrm{sr2}}^2\right)\right\} && (3.103)\\
&\overset{(iii)}{=} (T-1)\log\left(\rho_{\mathrm{sr1}}^2\right), && (3.104)
\end{aligned}
$$

where (i) is because X_{R},Λ are independently distributed of $X_{\mathrm{S}},\hat{Y}_{\mathrm{R}}$, (ii) is using (3.89) from Theorem 3.13 and (iii) is because the regime of the parameters of the network has $\gamma_{\mathrm{sr1}}>\gamma_{\mathrm{sr2}}$.

Now, we consider the second term in (3.99), recalling the choice of X_{R_1},X_{R_2} from (3.35)-(3.38) on page 63.

$$
\begin{aligned}
I\left(X_{\mathrm{R}}X_{\mathrm{S}};Y_{\mathrm{D}}\middle|\Lambda\right) \geq& I\left(X_{\mathrm{R}};Y_{\mathrm{D}}\middle|\Lambda\right) && (3.105)\\
=& h\left(Y_{\mathrm{D}}\middle|\Lambda\right)-h\left(Y_{\mathrm{D}}\middle|X_{\mathrm{R}},\Lambda\right) && (3.106)\\
\overset{(i)}{\dot{\geq}}& p_\lambda h\left(g_{\mathrm{rd1}}a_{\mathrm{R10}}X_{\mathrm{R10}}+g_{\mathrm{rd2}}a_{\mathrm{R20}}X_{\mathrm{R20}}+W_{1\times T}\right)\\
&+(1-p_\lambda)h\left(g_{\mathrm{rd1}}a_{\mathrm{R11}}X_{\mathrm{R11}}+g_{\mathrm{rd2}}a_{\mathrm{R21}}X_{\mathrm{R21}}+W_{1\times T}\right)\\
&-p_\lambda\log\left(\left(1+\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R10}}\right|^2\right)\left(1+\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R20}}\right|^2\right)\right)\\
&-(1-p_\lambda)\log\left(\left(1+\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R11}}\right|^2\right)\left(1+\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R21}}\right|^2\right)\right) && (3.107)\\
\overset{(ii)}{\dot{\geq}}& p_\lambda T\log\left(\max\left\{\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R10}}\right|^2,\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R20}}\right|^2\right\}\right)\\
&+(1-p_\lambda)T\log\left(\max\left\{\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R11}}\right|^2,\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R21}}\right|^2\right\}\right)\\
&-p_\lambda\log\left(\left(1+\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R10}}\right|^2\right)\left(1+\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R20}}\right|^2\right)\right)\\
&-(1-p_\lambda)\log\left(\left(1+\rho_{\mathrm{rd1}}^2\left|a_{\mathrm{R11}}\right|^2\right)\left(1+\rho_{\mathrm{rd2}}^2\left|a_{\mathrm{R21}}\right|^2\right)\right), && (3.108)
\end{aligned}
$$

where (i) is using Theorem 3.14 to evaluate $h\left(Y_{\mathrm{D}}\middle|X_{\mathrm{R}},\Lambda\right)$. Also, $W_{1\times T}$ is noise vector of length T with i.i.d. $\mathcal{CN}(0,1)$ elements. The step (ii) is using the fact that conditioning reduces entropy and the fact that $X_{\mathrm{R}ij}$ has i.i.d. $\mathcal{CN}(0,1)$ elements (refer to (3.35)-(3.38) on page 63).

Now, the third term in (3.99)

$$
\begin{aligned}
& I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) \\
&= I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) + I\left(X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{S}}, X_{\mathrm{R}_2}, \Lambda\right) && (3.109)\\
&\geq I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}_2} \middle| X_{\mathrm{R}_2}, \Lambda\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{S}}, X_{\mathrm{R}_2}, \Lambda\right) && (3.110)\\
&\overset{(i)}{=} I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) && (3.111)\\
&\overset{(ii)}{\doteq} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) && (3.112)\\
&= (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + h\left(Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) - h\left(Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}, X_{\mathrm{R}_2}, \Lambda\right) && (3.113)\\
&\overset{(iii)}{\dot{\geq}} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) \\
&\quad + p_\lambda h\left(g_{\mathrm{rd1}} a_{\mathrm{R10}} X_{\mathrm{R10}} + g_{\mathrm{rd2}} a_{\mathrm{R20}} X_{\mathrm{R20}} + W_{1\times T} \middle| a_{\mathrm{R20}} X_{\mathrm{R20}}\right) \\
&\quad + (1-p_\lambda) h\left(g_{\mathrm{rd1}} a_{\mathrm{R11}} X_{\mathrm{R11}} + g_{\mathrm{rd2}} a_{\mathrm{R21}} X_{\mathrm{R21}} + W_{1\times T} \middle| a_{\mathrm{R21}} X_{\mathrm{R21}}\right) \\
&\quad - p_\lambda \log\left(\left(1+\rho_{\mathrm{rd1}}^2 |a_{\mathrm{R10}}|^2\right)\left(1+\rho_{\mathrm{rd2}}^2 |a_{\mathrm{R20}}|^2\right)\right) \\
&\quad - (1-p_\lambda) \log\left(\left(1+\rho_{\mathrm{rd1}}^2 |a_{\mathrm{R11}}|^2\right)\left(1+\rho_{\mathrm{rd2}}^2 |a_{\mathrm{R21}}|^2\right)\right), && (3.114)
\end{aligned}
$$

where (i) is because $X_{\mathrm{R}_2}, \Lambda$ are distributed independently of $X_{\mathrm{S}}, \hat{Y}_{\mathrm{R}_2}$, and X_{S} is distributed independently of $X_{\mathrm{R}_1}, Y_{\mathrm{D}}$, (ii) is using (3.89) from Theorem 3.13 to evaluate $I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}_2}\right)$ and (iii) is using Theorem 3.14 to evaluate $h\left(Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}, X_{\mathrm{R}_2}, \Lambda\right)$. Also $W_{1\times T}$ is the noise vector of length T with i.i.d. $\mathcal{CN}(0,1)$ elements. Now,

$$
\begin{aligned}
& h\left(g_{\mathrm{rd1}} a_{\mathrm{R10}} X_{\mathrm{R10}} + g_{\mathrm{rd2}} a_{\mathrm{R20}} X_{\mathrm{R20}} + W_{1\times T} \middle| a_{\mathrm{R20}} X_{\mathrm{R20}}\right) \\
&\overset{(i)}{=} h\left(g_{\mathrm{rd1}} a_{\mathrm{R10}} x_{\mathrm{R10}} + g_{\mathrm{rd2}} a_{\mathrm{R20}} \|X_{\mathrm{R20}}\| + w_1, g_{\mathrm{rd1}} X_{\mathrm{R10}\times(T-1)} + W_{1\times(T-1)} \middle| a_{\mathrm{R20}} X_{\mathrm{R20}}\right) && (3.115)\\
&\geq h\left(g_{\mathrm{rd1}} a_{\mathrm{R10}} x_{\mathrm{R10}} + g_{\mathrm{rd2}} a_{\mathrm{R20}} \|X_{\mathrm{R20}}\| + w_1 \middle| a_{\mathrm{R20}}, x_{\mathrm{R10}}, \|X_{\mathrm{R20}}\|\right)
\end{aligned}
$$

$$+ h\left(a_{\mathrm{R10}} g_{\mathrm{rd1}} X_{\mathrm{R10}\times(T-1)} + W_{1\times(T-1)} \middle| a_{\mathrm{R20}}, g_{\mathrm{rd1}}\right) \tag{3.116}$$

$$\stackrel{(ii)}{\dot{\geq}} \mathbb{E}\left[\log\left(\rho_{\mathrm{rd1}}^2 \left|a_{\mathrm{R10}}\right|^2 \left|x_{\mathrm{R10}}\right|^2 + \rho_{\mathrm{rd2}}^2 \left|a_{\mathrm{R20}}\right|^2 \left\|X_{\mathrm{R20}}\right\|^2 + 1\right)\right]$$

$$+ (T-1)\,\mathbb{E}\left[\log\left(\left|a_{\mathrm{R10}}\right|^2 \left|g_{\mathrm{rd1}}\right|^2 + 1\right)\right] \tag{3.117}$$

$$\stackrel{(iii)}{\doteq} \log\left(\left|a_{\mathrm{R10}}\right|^2 \rho_{\mathrm{rd1}}^2 + \left|a_{\mathrm{R20}}\right|^2 \rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(\left|a_{\mathrm{R10}}\right|^2 \rho_{\mathrm{rd1}}^2 + 1\right), \tag{3.118}$$

where (i) is by projecting $g_{\mathrm{rd1}} a_{\mathrm{R10}} X_{\mathrm{R10}} + g_{\mathrm{rd2}} a_{\mathrm{R20}} X_{\mathrm{R20}} + W_{1\times T}$ onto a new orthonormal basis with first basis vector chosen in the direction of X_{R20} and rest of the basis vectors with an arbitrary choice. The direction of X_{R20} is known from $a_{\mathrm{R20}} X_{\mathrm{R20}}$ given in the conditioning since a_{R20} is a known constant. Note that X_{R10} having i.i.d. $\mathcal{CN}(0,1)$ elements, projected onto any direction independent of X_{R10} gives a $\mathcal{CN}(0,1)$ random variable which is x_{R10} in step (i), and X_{R10} projected to rest of the $T-1$ basis vectors gives a vector $X_{\mathrm{R10}\times(T-1)}$ of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$ elements. Also w_1 is i.i.d. $\mathcal{CN}(0,1)$ noise and $W_{1\times(T-1)}$ is the noise vector of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$ elements. The step (ii) is using the property of Gaussians and (iii) is using Fact 5.1 and Fact 3.3 on page 68. Similarly,

$$h\left(g_{\mathrm{rd1}} a_{\mathrm{R11}} X_{\mathrm{R11}} + g_{\mathrm{rd2}} a_{\mathrm{R21}} X_{\mathrm{R21}} + W_{1\times T} \middle| a_{\mathrm{R21}} X_{\mathrm{R21}}\right)$$

$$\dot{\geq} \log\left(\left|a_{\mathrm{R11}}\right|^2 \rho_{\mathrm{rd1}}^2 + \left|a_{\mathrm{R21}}\right|^2 \rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(\left|a_{\mathrm{R11}}\right|^2 \rho_{\mathrm{rd1}}^2 + 1\right). \tag{3.119}$$

Hence, by substituting (3.119), (3.118) in (3.114), we get

$$I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right)$$

$$\dot{\geq} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right)$$

$$+ p_\lambda\left(\log\left(\left|a_{\mathrm{R10}}\right|^2 \rho_{\mathrm{rd1}}^2 + \left|a_{\mathrm{R20}}\right|^2 \rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(\left|a_{\mathrm{R10}}\right|^2 \rho_{\mathrm{rd1}}^2 + 1\right)\right)$$

$$+ (1-p_\lambda)\left(\log\left(\left|a_{\mathrm{R11}}\right|^2 \rho_{\mathrm{rd1}}^2 + \left|a_{\mathrm{R21}}\right|^2 \rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(\left|a_{\mathrm{R11}}\right|^2 \rho_{\mathrm{rd1}}^2 + 1\right)\right)$$

$$- p_\lambda \log\left(\left(1 + \rho_{\mathrm{rd1}}^2 \left|a_{\mathrm{R10}}\right|^2\right)\left(1 + \rho_{\mathrm{rd2}}^2 \left|a_{\mathrm{R20}}\right|^2\right)\right)$$

$$- (1-p_\lambda)\log\left(\left(1 + \rho_{\mathrm{rd1}}^2 \left|a_{\mathrm{R11}}\right|^2\right)\left(1 + \rho_{\mathrm{rd2}}^2 \left|a_{\mathrm{R21}}\right|^2\right)\right). \tag{3.120}$$

The fourth term in (3.99) is symmetric with the third term, hence it can be simplified similar to the above equation to obtain

$$\begin{aligned}
& I\left(X_{\mathrm{S}}, X_{\mathrm{R}_2}; \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}, \Lambda\right) \\
& \dot{\geq} (T-1)\log\left(\rho_{\mathrm{sr1}}^2\right) \\
& \quad + p_\lambda\left(\log\left(|a_{\mathrm{R10}}|^2\rho_{\mathrm{rd1}}^2 + |a_{\mathrm{R20}}|^2\rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(|a_{\mathrm{R20}}|^2\rho_{\mathrm{rd2}}^2 + 1\right)\right) \\
& \quad + (1-p_\lambda)\left(\log\left(|a_{\mathrm{R11}}|^2\rho_{\mathrm{rd1}}^2 + |a_{\mathrm{R21}}|^2\rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(|a_{\mathrm{R21}}|^2\rho_{\mathrm{rd2}}^2 + 1\right)\right) \\
& \quad - p_\lambda\log\left(\left(1+\rho_{\mathrm{rd1}}^2|a_{\mathrm{R10}}|^2\right)\left(1+\rho_{\mathrm{rd2}}^2|a_{\mathrm{R20}}|^2\right)\right) \\
& \quad - (1-p_\lambda)\log\left(\left(1+\rho_{\mathrm{rd1}}^2|a_{\mathrm{R11}}|^2\right)\left(1+\rho_{\mathrm{rd2}}^2|a_{\mathrm{R21}}|^2\right)\right). \qquad (3.121)
\end{aligned}$$

Since we are dealing with the case from Section 3.3.1, from our choice (looking at (3.121) and (3.104)) it follows that

$$I\left(X_{\mathrm{S}}; \hat{Y}_{\mathrm{R}}Y_{\mathrm{D}} \middle| X_{\mathrm{R}}, \Lambda\right), I\left(X_{\mathrm{S}}, X_{\mathrm{R}_2}; \hat{Y}_{\mathrm{R}_1}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}, \Lambda\right) \dot{\geq} (T-1)\log\left(\rho_{\mathrm{sr1}}^2\right) \qquad (3.122)$$

$$\doteq (T-1)\,\gamma_{\mathrm{sr1}}\log(\mathsf{SNR}) \qquad (3.123)$$

for any $a_{\mathrm{R10}}, a_{\mathrm{R11}}, a_{\mathrm{R20}}, a_{\mathrm{R21}}$. Now, we choose

$$a_{\mathrm{R10}} = c_1, a_{\mathrm{R11}} = 1, a_{\mathrm{R20}} = 1, a_{\mathrm{R21}} = 0 \qquad (3.124)$$

and substitute in (3.120) to get

$$\begin{aligned}
I\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}; \hat{Y}_{\mathrm{R}_2}, Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}, \Lambda\right) & \dot{\geq} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) \\
& \quad + p_\lambda\left(\log\left(|c_1|^2\rho_{\mathrm{rd1}}^2 + \rho_{\mathrm{rd2}}^2 + 1\right) + (T-1)\log\left(|c_1|^2\rho_{\mathrm{rd1}}^2 + 1\right)\right) \\
& \quad + (1-p_\lambda)\left(\log\left(\rho_{\mathrm{rd1}}^2 + 1\right) + (T-1)\log\left(\rho_{\mathrm{rd1}}^2 + 1\right)\right) \\
& \quad - p_\lambda\log\left(\left(1+\rho_{\mathrm{rd1}}^2|c_1|^2\right)\left(1+\rho_{\mathrm{rd2}}^2\right)\right) \\
& \quad - (1-p_\lambda)\log\left(\left(1+\rho_{\mathrm{rd1}}^2\right)\right) \qquad (3.125) \\
& \overset{(i)}{\doteq} (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) \\
& \quad + p_\lambda\left((T-2)\log\left(|c_1|^2\rho_{\mathrm{rd1}}^2 + 1\right)\right) \\
& \quad + (1-p_\lambda)\left((T-1)\log\left(\rho_{\mathrm{rd1}}^2 + 1\right)\right) \qquad (3.126) \\
& \doteq (T-1)\,\gamma_{\mathrm{sr2}}\log(\mathsf{SNR})
\end{aligned}$$

$$+ p_\lambda \left((T-2) \log \left(|c_1|^2 \mathsf{SNR}^{\gamma_{\text{rd1}}} + 1 \right) \right)$$
$$+ (1-p_\lambda)(T-1)\gamma_{\text{rd1}} \log(\mathsf{SNR}), \tag{3.127}$$

where (i) was using $|c_1|^2 \rho_{\text{rd1}}^2 < \rho_{\text{rd2}}^2$ since $\rho_{\text{rd1}}^2 < \rho_{\text{rd2}}^2$ and $|c_1|^2$ is power constrained. And similarly on substituting $a_{\text{R10}} = c_1, a_{\text{R11}} = 1, a_{\text{R20}} = 1, a_{\text{R21}} = 0$ in (3.108), we get

$$\begin{aligned} I(X_{\text{R}} X_{\text{S}}; Y_{\text{D}} | \Lambda) &\dot{\geq} p_\lambda (T-1) \log \left(\rho_{\text{rd2}}^2 \right) \\ &\quad + (1-p_\lambda)(T-1) \log \left(\rho_{\text{rd1}}^2 \right) \\ &\quad - p_\lambda \log \left(\left(1 + \rho_{\text{rd1}}^2 |c_1|^2 \right) \right) \qquad (3.128) \\ &\doteq p_\lambda (T-1) \gamma_{\text{rd2}} \log(\mathsf{SNR}) + (1-p_\lambda)(T-1)\gamma_{\text{rd1}} \log(\mathsf{SNR}) \\ &\quad - p_\lambda \log \left(\left(1 + \mathsf{SNR}^{\gamma_{\text{rd1}}} |c_1|^2 \right) \right). \qquad (3.129) \end{aligned}$$

Now, substituting (3.123), (3.127) and (3.129) into (3.99), we get that the rates

$$\begin{aligned} TR \,\dot{\leq}\, \min \Big\{ & (T-1)\gamma_{\text{sr1}} \log(\mathsf{SNR}), (T-1)\gamma_{\text{sr2}} \log(\mathsf{SNR}) + (1-p_\lambda)(T-1)\gamma_{\text{rd1}} \log(\mathsf{SNR}) \\ & + p_\lambda \left((T-2) \log \left(|c_1|^2 \mathsf{SNR}^{\gamma_{\text{rd1}}} + 1 \right) \right), \ (1-p_\lambda)(T-1)\gamma_{\text{rd1}} \log(\mathsf{SNR}) \\ & + p_\lambda (T-1) \gamma_{\text{rd2}} \log(\mathsf{SNR}) - p_\lambda \log \left(\left(1 + \mathsf{SNR}^{\gamma_{\text{rd1}}} |c_1|^2 \right) \right) \Big\} \end{aligned} \tag{3.130}$$

are achievable. Thus with

$$\mathcal{P}_9 : \begin{cases} \text{maximize} \min \Big\{ (T-1)\gamma_{\text{sr2}} \log(\mathsf{SNR}) + (T-1)(1-p_\lambda)\gamma_{\text{rd1}} \log(\mathsf{SNR}) \\ \qquad + (T-2) p_\lambda \log \left(\mathsf{SNR}^{\gamma_{\text{rd1}}} |c_1|^2 + 1 \right), \\ \qquad (T-1)(1-p_\lambda)\gamma_{\text{rd1}} \log(\mathsf{SNR}) \\ \qquad + p_\lambda \left((T-1)\gamma_{\text{rd2}} \log(\mathsf{SNR}) - \log \left(\mathsf{SNR}^{\gamma_{\text{rd1}}} |c_1|^2 + 1 \right) \right) \Big\} \\ |c_1|^2 \leq 2T, 0 \leq p_\lambda \leq 1, \end{cases} \tag{3.131}$$

the rates

$$TR \,\dot{\leq}\, \min \left\{ (T-1)\gamma_{\text{sr1}} \log(\mathsf{SNR}), (\mathcal{P}_9) \right\} \tag{3.132}$$

are achievable. And from Lemma 3.1, the solution of $\mathcal{P}_9$ has the same gDoF as the solution of the optimization problem $\mathcal{P}_1$, where $\mathcal{P}_1$ appeared in the outer bound as

$$T\overline{C} \dot{\leq} \min \left\{ (T-1)\, \gamma_{\text{sr1}} \log \left(\mathsf{SNR} \right), (\mathcal{P}_1) \right\} \tag{3.133}$$

in (3.21). Hence the outer bound can be achieved, using the optimal values of $p_\lambda, |c_1|^2$ for $\mathcal{P}_9$ from Table 3.2, in the input distribution as described in (3.35)- (3.38) and (3.41).

3.4.5 Study of Train-Scale and Quantizing for a Simple Channel

We consider $Y = gX + W$ with X being a vector of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$ elements and W being a vector of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$ elements, $g \sim \mathcal{CN}(0, \rho^2)$. It is scaled to $Y' = (g/\hat{g})\, X + W/\hat{g}$, where we choose $\hat{g} = e^{i\angle(g+w)} + (g+w)$,where $\angle(g+w')$ is the angle of $g + w'$. Note that $|\hat{g}| = 1 + |g+w|$ and

$$1 + |g+w|^2 \leq |\hat{g}|^2 \leq 2(1 + |g+w|^2). \tag{3.134}$$

Now Y' is quantized to

$$\hat{Y} = \frac{g}{\hat{g}} X + \frac{W}{\hat{g}} + Q \tag{3.135}$$

with $Q \sim W/\hat{g}$. For X, being a vector of length $T-1$ with i.i.d. $\mathcal{CN}(0,1)$, we can equivalently use

$$X = \alpha \overline{q}^{(T-1)\dagger}, \tag{3.136}$$

where $\overline{q}^{(T-1)}$ is a $T-1$ dimensional isotropically distributed unitary vector and

$$\alpha \sim \sqrt{\frac{1}{2} \chi^2 \left(2\left(T-1\right)\right)}, \tag{3.137}$$

where $\chi^2(k)$ is chi-squared distributed. (See Section 3.4.1.1 on page 67 for details on chi-squared distribution).

Now, through the rest of this section we claim that $I\left(\hat{Y}; X\right) - I\left(\hat{Y}; Y' \middle| X\right) \dot{\geq} (T-1)\log\left(\rho^2\right)$ by first showing $I\left(\hat{Y}; X\right) \dot{\geq} (T-1)\log\left(\rho^2\right)$ and then showing $I\left(\hat{Y}; Y' \middle| X\right) \dot{\leq} 0$.

3.4.5.1 Analysis of $I\left(\hat{Y}; X\right)$

$$I\left(\hat{Y}; X\right) = h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\right) - h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\, \alpha\overline{q}^{(T-1)\dagger}\right) \tag{3.138}$$

$$\begin{aligned} &\overset{(i)}{\geq} h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger}\middle|\, \frac{g}{\hat{g}}\right) \\ &\quad - h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\, \alpha\overline{q}^{(T-1)\dagger}\right) \end{aligned} \tag{3.139}$$

$$\begin{aligned} &\overset{(ii)}{=} h\left(\left|\frac{g}{\hat{g}}\alpha\right|^2\middle|\, \frac{g}{\hat{g}}\right) + (T-2)\,\mathbb{E}\left[\log\left(\left|\frac{g}{\hat{g}}\alpha\right|^2\right)\right] + \log\left(\frac{\pi^{T-1}}{\Gamma(T-2)}\right) \\ &\quad - h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\, \alpha\overline{q}^{(T-1)\dagger}\right), \end{aligned} \tag{3.140}$$

where (i) is using the fact that conditioning reduces entropy and (ii) is using the result from Corollary 3.8.

Now consider $h\left((g/\hat{g})\,\alpha\overline{q}^{(T-1)\dagger} + W/\hat{g} + Q\middle|\, \alpha\overline{q}^{(T-1)\dagger}\right)$. By projecting $(g/\hat{g})\,\alpha\overline{q}^{(T-1)\dagger} + W/\hat{g} + Q$ onto a new orthonormal basis with the first basis vector taken as $\overline{q}^{(T-1)\dagger}$, which is known in conditioning, we get

$$\begin{aligned} &h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\, \alpha\overline{q}^{(T-1)\dagger}\right) \\ &\overset{(i)}{=} h\left(\frac{g}{\hat{g}}\alpha + \frac{w}{\hat{g}} + q, \frac{W'_{1\times T-2}}{\hat{g}} + Q'_{1\times T-2}\middle|\, \alpha\right) \end{aligned} \tag{3.141}$$

$$\begin{aligned} &\overset{(ii)}{\leq} h\left(\frac{g}{\hat{g}}\alpha + \frac{w}{\hat{g}} + q\middle|\, \alpha\right) \\ &\quad + (T-2)\log\left(\pi e\mathbb{E}\left[\left|\frac{w}{\hat{g}} + q\right|^2\right]\right) \end{aligned} \tag{3.142}$$

$$\begin{aligned} &\overset{(iii)}{=} h\left(\frac{g}{\hat{g}}\alpha + \frac{w}{\hat{g}} + q\middle|\, \alpha\right) \\ &\quad + (T-2)\log\left(\pi e\mathbb{E}\left[2\left|\frac{w}{\hat{g}}\right|^2\right]\right) \end{aligned} \tag{3.143}$$

$$\begin{aligned} &\overset{(iv)}{\leq} h\left(\left(\frac{g}{\hat{g}} - 1\right)\alpha + \frac{w}{\hat{g}} + q\middle|\, \alpha\right) \\ &\quad + (T-2)\log\left(\pi e\frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \end{aligned} \tag{3.144}$$

$$\stackrel{(v)}{\leq} \log\left(\pi e \mathbb{E}\left[\left|\left(\frac{g}{\hat{g}}-1\right)\alpha + \frac{w}{\hat{g}} + q\right|^2\right]\right)$$

$$+ (T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \tag{3.145}$$

$$\stackrel{(vi)}{=} \log\left(\pi e \mathbb{E}\left[\left|\left(\frac{g}{\hat{g}}-1\right)\alpha\right|^2 + 2\left|\frac{w}{\hat{g}}\right|^2\right]\right)$$

$$+ (T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right), \tag{3.146}$$

where in step (i) $W'_{1\times T-2}/\hat{g}, Q'_{1\times T-2}$ are independent vectors of length $(T-2)$ with i.i.d. elements distributed according to $w/\hat{g}$, $w \sim \mathcal{CN}(0,1)$ and $q \sim w/\hat{g}$. This step is similar to that in (3.115). The step (ii) is using the fact that conditioning reduces entropy, maximum entropy results and the fact that $W'_{1\times T-2}/\hat{g}, Q'_{1\times T-2}$ have i.i.d. elements distributed according to $w/\hat{g}, q$, (iii) is because $w/\hat{g}, q$ are i.i.d., (iv) is by subtracting α in the first term, since α is known and using Fact 3.2 from page 67 on $\mathbb{E}\left[|w/\hat{g}|^2\right] \leq \mathbb{E}\left[|w|^2/\left(1+|g+w'|^2\right)\right] = \mathbb{E}\left[1/\left(1+|g+w'|^2\right)\right]$, (v) is using maximum entropy results and (vi) is using the fact that $w/\hat{g} \sim q$.

Hence

$$h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\,\alpha\overline{q}^{(T-1)\dagger}\right)$$

$$\leq \log\left(\pi e \mathbb{E}\left[\left|\left(\frac{g}{\hat{g}}-1\right)\alpha\right|^2 + 2\left|\frac{w}{\hat{g}}\right|^2\right]\right)$$

$$+ (T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \tag{3.147}$$

$$\stackrel{(i)}{\leq} \log\left(\pi e \mathbb{E}\left[\left|\left(\frac{g}{e^{i\angle(g+w')}+(g+w')}-1\right)\alpha\right|^2 + 2\left|\frac{w}{e^{i\angle(g+w')}+(g+w')}\right|^2\right]\right)$$

$$+ (T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \tag{3.148}$$

$$\stackrel{(ii)}{=} \log\left(\pi e \mathbb{E}\left[\left|\left(\frac{e^{i\angle(g+w')}+w'}{e^{i\angle(g+w')}+(g+w')}\right)\right|^2 (T-1) + 2\left|\frac{w}{e^{i\angle(g+w')}+(g+w')}\right|^2\right]\right)$$

$$+ (T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \tag{3.149}$$

$$\overset{(iii)}{\leq} \log\left(\pi e \mathbb{E}\left[\frac{2+2\left|w'\right|^2}{1+\left|g+w'\right|^2}(T-1)+\frac{2}{1+\left|g+w'\right|^2}\right]\right) \tag{3.150}$$

$$+(T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right) \tag{3.151}$$

$$\overset{(iv)}{\leq} \log\left(\pi e \mathbb{E}\left[\frac{2\left|w'\right|^2}{1+\left|g+w'\right|^2}(T-1)+\frac{2T}{\rho^2+1}\ln\left(2+\rho^2\right)\right]\right)$$

$$+(T-2)\log\left(\pi e \frac{2}{\rho^2+1}\ln\left(2+\rho^2\right)\right), \tag{3.152}$$

where (i) is using $\hat{g}=e^{i\angle(g+w')}+(g+w')$, (ii) is using $\mathbb{E}\left[|\alpha|^2\right]=T-1$ with α independent of everything else (α was chosen in (3.137)), (iii) is using $\left|e^{i\angle(g+w')}+w'\right|^2 \leq 2\left(1+|w'|^2\right)$, $\mathbb{E}\left[|w|^2\right] = 1$ and $\left|e^{i\angle(g+w')}+(g+w')\right|^2 \geq 1+|g+w'|^2$ and (iv) is using Fact 3.2 on $\mathbb{E}\left[1/\left(1+|g+w'|^2\right)\right]$.

Now, for $\mathbb{E}\left[|w'|^2/\left(1+|g+w'|^2\right)\right]$, we use the following fact.

Lemma 3.4. $\log\left(\mathbb{E}\left[\frac{|w|^2}{1+|g+w|^2}\right]\right) \dot{\leq} \log\left(\frac{1}{\rho^2}\right)$

Proof. See Appendix B.6. □

Hence using the previous lemma on 3.152, it follows that

$$h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger}+\frac{W}{\hat{g}}+Q\middle|\alpha\overline{q}^{(T-1)\dagger}\right) \dot{\leq} (T-1)\log\left(\frac{1}{\rho^2}\right). \tag{3.153}$$

Now, substituting (3.153) in (3.140), we get

$$I\left(\hat{Y};X\right) \dot{\geq} h\left(\left|\frac{g}{\hat{g}}\alpha\right|^2\middle|\frac{g}{\hat{g}}\right)+(T-2)\mathbb{E}\left[\log\left(\left|\frac{g}{\hat{g}}\alpha\right|^2\right)\right]-(T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.154}$$

$$=h\left(|\alpha|^2\right)+(T-1)\mathbb{E}\left[\log\left(\left|\frac{g}{\hat{g}}\right|^2\right)\right]+(T-2)\mathbb{E}\left[\log\left(|\alpha|^2\right)\right]$$

$$-(T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.155}$$

$$\overset{(i)}{\doteq}(T-1)\mathbb{E}\left[\log\left(\left|\frac{g}{\hat{g}}\right|^2\right)\right]-(T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.156}$$

$$\overset{(ii)}{\geq}(T-1)\mathbb{E}\left[\log\left(|g|^2\right)\right]-(T-1)\mathbb{E}\left[\log\left(2\left(1+|g+w'|^2\right)\right)\right] \tag{3.157}$$

$$- (T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.158}$$

$$\overset{(iii)}{\doteq} (T-1)\log\left(\frac{\rho^2}{2\left(2+\rho^2\right)}\right) - (T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.159}$$

$$\doteq (T-1)\log\left(\rho^2\right), \tag{3.160}$$

where (i) is because $\alpha \sim \sqrt{\frac{1}{2}\chi^2\left(2\left(T-1\right)\right)}$ and using properties of chi-squared random variables (see Section 3.4.1.1 on page 67), (ii) is using $|\hat{g}|^2 \leq 2\left(1+|g+w'|^2\right)$, (iii) is using Fact 5.1 from page 135 for $\mathbb{E}\left[\log\left(1+|g+w'|^2\right)\right]$. Hence we have

$$I\left(\hat{Y};X\right) \dot{\geq} (T-1)\log\left(\rho^2\right).$$

3.4.5.2 Analysis of $I\left(\hat{Y};Y'\middle|X\right)$

$$I\left(\hat{Y};Y'\middle|X\right) = h\left(\hat{Y}\middle|X\right) - h\left(\hat{Y}\middle|Y',X\right) \tag{3.161}$$

$$= h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\alpha\overline{q}^{(T-1)\dagger}\right)$$
$$- h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}}, \alpha\overline{q}^{(T-1)\dagger}\right) \tag{3.162}$$

$$= h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q\middle|\alpha\overline{q}^{(T-1)\dagger}\right) - h(Q) \tag{3.163}$$

$$h(Q) = h\left(\frac{W}{\hat{g}}\right) \tag{3.164}$$

$$\geq h\left(\frac{W}{\hat{g}}\middle|\hat{g}\right) \tag{3.165}$$

$$\overset{(i)}{=} (T-1)\times h\left(\frac{w}{\hat{g}}\middle|\hat{g}\right) \tag{3.166}$$

$$\overset{(ii)}{\geq} (T-1)\left(\mathbb{E}\left[\log\left(\frac{1}{2\left(1+|g+w'|^2\right)}\right)\right] + h(w)\right) \tag{3.167}$$

$$\overset{(iii)}{\doteq} (T-1)\log\left(\frac{1}{\rho^2}\right), \tag{3.168}$$

where (i) is using the fact that W is a vector of length $(T-1)$ with i.i.d. elements distributed as $w \sim \mathcal{CN}(0,1)$, (ii) is using the structure of $\hat{g}$ and (iii) is using Fact 5.1 from page 135 for $\mathbb{E}\left[\log\left(1+|g+w'|^2\right)\right]$ and $h(w) \doteq 0$. Hence

$$I\left(\hat{Y}; Y' \middle| X\right) \dot{\leq} h\left(\frac{g}{\hat{g}}\alpha\overline{q}^{(T-1)\dagger} + \frac{W}{\hat{g}} + Q \middle| \alpha\overline{q}^{(T-1)\dagger}\right) - (T-1)\log\left(\frac{1}{\rho^2}\right) \tag{3.169}$$

We had already shown $h\left((g/\hat{g})\,\alpha\overline{q}^{(T-1)\dagger} + W/\hat{g} + Q \middle| \alpha\overline{q}^{(T-1)\dagger}\right) \dot{\leq} (T-1)\log\left(1/\rho^2\right)$ in (3.153). Hence we have

$$I\left(\hat{Y}; Y' \middle| X\right) \dot{\leq} 0. \tag{3.170}$$

CHAPTER 4

Fast Fading Interference Channel

4.1 Introduction

The 2-user Gaussian IC is a simple model that captures the effect of interference in wireless networks. Significant progress has been made in understanding the capacity of the Gaussian IC [HK81, CMG08, ETW08, ST11]. In practice the links in the channel could be time-varying rather than static. Characterizing the capacity of FF-IC without CSIT has been an open problem. In this chapter we make progress in this direction by obtaining the capacity region of certain classes of FF-IC without instantaneous CSIT within a constant gap.

4.1.1 Related work

Previous works have characterized the capacity region to within a constant gap for the IC where the channel is known at the transmitter and receiver. The capacity region of the 2-user IC without feedback was characterized to within 1 bit/s/Hz in [ETW08]. In [ST11], Suh and Tse characterized the capacity region of the IC with feedback to within 2 bits per channel use. These results were based on the Han-Kobayashi scheme [HK81], where the transmitters use superposition coding splitting their messages into common and private parts, and the receivers use joint decoding. Other variants of wireless networks based on the IC model have been studied in literature. The interference relay channel (IRC), which is obtained by adding a relay to the 2-user interference channel (IC) setup, was introduced in [SE07] and was further studied in [TY11, MDG12, BPY15, GCS16]. In [WT11a], Wang and Tse studied the IC with receiver cooperation. The IC with source cooperation was studied in

[PV11, WT11b].

When the channels are time varying, most of the existing techniques for IC cannot be used without CSIT. In [Far13], Farsani showed that if each transmitter of FF-IC has knowledge of the *inr* to the non-corresponding receiver[1], the capacity region can be achieved within 1 bit/s/Hz. Lalitha *et al.* [SSE11] derived sum-capacity results for a subclass of FF-IC with perfect CSIT. The idea of interference alignment [CJ08] has been extended to FF-IC to obtain the degrees of freedom (DoF) region for certain cases. The degrees of freedom region for the MIMO interference channel with delayed CSIT was studied in [VV12]. Their results show that when *all users have single antenna*, the DoF region is same for the cases of no CSIT, delayed CSIT and instantaneous CSIT. The results from [TMP13] show that DoF region for FF-IC with output feedback and delayed CSIT is contained in the DoF region for the case with instantaneous CSIT and no feedback. Kang and Choi [KC13] considered interference alignment for the K-user FF-IC with delayed channel state feedback and showed a result of $2K/(K+2)$ DoF. They also showed the same DoF can be achieved using a scaled output feedback, but without channel state feedback. Therefore, the above works have characterizations for DoF for several fading scenarios, and also show that for single antenna systems, feedback is not very effective in terms of DoF. However, as we show in this chapter, the situation changes when we look for more than DoF, and for approximate optimality of the entire capacity region. In particular, we allow for arbitrary channel gains, and do not limit ourselves to SNR-scaling results[2]. In particular, we show that though the capacity region is same (within a constant) for the cases of no CSIT, delayed CSIT and instantaneous CSIT, there is a significant difference with output feedback. When there is output feedback and delayed CSIT the capacity region is larger than that for the case with no feedback and instantaneous CSIT in contrast to the DoF result from [TMP13]. This gives

[1]For Tx1 the non-corresponding receiver is Rx2 and similarly for Tx2 the non-corresponding receiver is Rx1

[2]However, we can also use our results to get the *generalized* DoF studied in [ETW08] for the FF-IC. This shows that for generalized DoF, feedback indeed helps, as shown in our results.

us a finer understanding of the role of CSIT as well as feedback in FF-IC with arbitrary (and potentially asymmetric) link strengths, and is one of the main contributions of this chapter.

Some simplified fading models have been introduced to characterize the capacity region of the FF-IC in the absence of CSIT. In [WSD13], Wang *et al.* considered the bursty IC, where the presence of interference is governed by a Bernoulli random state. The capacity of one-sided IC under ergodic layered erasure model, which considers the channel as a time-varying version of the binary expansion deterministic model [ADT11], was studied in [ASC09, ZG11]. The binary fading IC, where the channel gains, the transmit signals and the received signals are in the binary field was studied in [VMA14, VMA17] by Vahid *et al.* In spite of these efforts, the capacity region of FF-IC without CSIT is still unknown, and this chapter presents what we believe to be the first approximate characterization of the capacity region of FF-IC without CSIT, for a class of fading models satisfying the regularity condition, defined as the finite *logarithmic Jensen's gap*.

4.1.2 Contribution and outline

In this chapter we first introduce the notion of *logarithmic Jensen's gap* for fading models. This is defined in Section 4.3 as a number calculated for a fading model depending on the probability distribution for the channel strengths. It is effectively the supremum of $\log\left(\mathbb{E}\left[\text{link strength}\right]\right)-\mathbb{E}\left[\log\left(\text{link strength}\right)\right]$ over all links and operating regimes of the system. We show that common fading models including Rayleigh and Nakagami fading have finite logarithmic Jensen's gap, but some fading models (like bursty fading [WSD13]) have infinite logarithmic Jensen's gap. Subsequently, we show the usefulness of logarithmic Jensen's gap in obtaining approximate capacity regions of FF-ICs without instantaneous CSIT. We show that Han-Kobayashi type rate-splitting schemes [HK81, CMG08, ETW08, ST11] based on *inr*, when extended to rate-splitting schemes based on $\mathbb{E}\left[inr\right]$ for the FF-ICs, give the capacity gap as a function of logarithmic Jensen's gap, yielding the approximate capacity characterization for fading models that have finite logarithmic Jensen's gap. Since our rate-splitting is based on $\mathbb{E}\left[inr\right]$, it does not need instantaneous CSIT. The constant gap capacity result is first obtained for FF-IC without feedback or instantaneous CSIT. We also show that for the FF-IC without feedback, instantaneous CSIT cannot improve the capacity region over the case with no instantaneous CSIT, except for a constant gap. We subsequently study FF-IC with feedback and delayed CSIT to obtain a constant gap capacity result. In this case, having instantaneous CSIT cannot improve the capacity region over the case with delayed CSIT. We show that our analysis of FF-IC can easily be extended to fading interference MAC channel to yield an approximate capacity result.

The usefulness of logarithmic Jensen's gap is further illustrated by analyzing a scheme based on point-to-point codes for a class of FF-IC with feedback, where we again obtain capacity gap as a function of logarithmic Jensen's gap. Our scheme is based on amplify-and-forward relaying, similar to the one proposed in [ST11]. It effectively induces a 2-tap inter-symbol-interference (ISI) channel for one of the users and a point-to-point feedback channel for the other user. The work in [ST11] had similarly shown that an amplify-and-forward based feedback scheme can achieve the symmetric rate point, without using rate-splitting. Our scheme can be considered as an extension to this scheme, which enables us to approximately achieve the entire capacity region. Our analysis also yields a capacity bound for a 2-tap fading ISI channel, the tightness of the bound again determined by the logarithmic Jensen's gap.

The chapter is organized as follows. In section 5.2 we describe the system setup and the notations. In section 4.3 we develop the logarithmic Jensen's gap characterization for fading models. We illustrate a few applications of logarithmic Jensen's gap characterization in the later sections: in section 4.4, by obtaining approximate capacity region of FF-IC without feedback, in section 4.5, by obtaining approximate capacity region of FF-IC with feedback and delayed CSIT, and in section 4.6, by developing point-to-point codes for a class of FF-IC with feedback.

4.2 Model and Notation

We consider the two-user FF-IC (Figure 4.1)

$$Y_1(l) = g_{11}(l)X_1(l) + g_{21}(l)X_2(l) + Z_1(l) \tag{4.1}$$

$$Y_2(l) = g_{12}(l)X_1(l) + g_{22}(l)X_2(l) + Z_2(l), \tag{4.2}$$

where $Y_i(l)$ is the channel output of receiver i (Rxi) at time l, $X_i(l)$ is the input of transmitter i (Txi) at time l, $Z_i(l) \sim \mathcal{CN}(0,1)$ is complex AWGN noise process at Rxi, and $g_{ij}(l)$ is the time-variant random channel gain. The channel gain processes $\{g_{ij}(l)\}$ are constant over a block of size T and independent across blocks and across links (i,j). Without loss of generality we assume block size $T = 1$ for the fading, our results can be easily extended for arbitrary T case by coding across the blocks. The transmitters are assumed to have no knowledge of the channel gain realizations, but the receivers do have full knowledge of their corresponding channels. We assume that $|g_{ij}(l)|^2$ is a random variable with a known distribution. We assume average power constraint $(1/N)\sum_{l=1}^{N}|X_i(l)|^2 \leq 1, i = 1,2$ at the transmitters, and assume Txi has a message $W_i \in \left\{1,.,2^{NR_i}\right\}$, for a block length of N, intended for Rxi for $i = 1,2$, and W_1, W_2 are independent of each other. We denote $SNR_i := \mathbb{E}\left[|g_{ii}|^2\right]$ for $i = 1,2$, and $INR_i := \mathbb{E}\left[|g_{ij}|^2\right]$ for $i \neq j$. For the instantaneous interference channel gains we use $inr_i := |g_{ij}|^2$, $i \neq j$. Note that we allow for arbitrary channel gains, and do not limit ourselves to SNR-scaling results, but get an approximate characterization of the FF-IC capacity region.

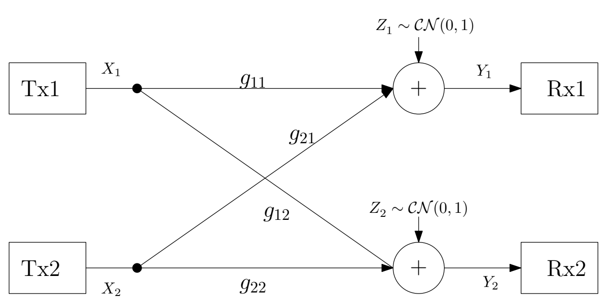


Figure 4.1: The channel model without feedback.

Under the feedback model (Figure 4.2), after each reception, each receiver reliably feeds back the received symbol and the channel states to its corresponding transmitter[3]. For example, at time l, Tx1 receives $(Y_1(l-1), g_{11}(l-1), g_{21}(l-1))$ from Rx1. Thus $X_1(l)$ is allowed to be a function of $\left(W_1, \{Y_1(k), g_{11}(k), g_{21}(k)\}_{k<l}\right)$.

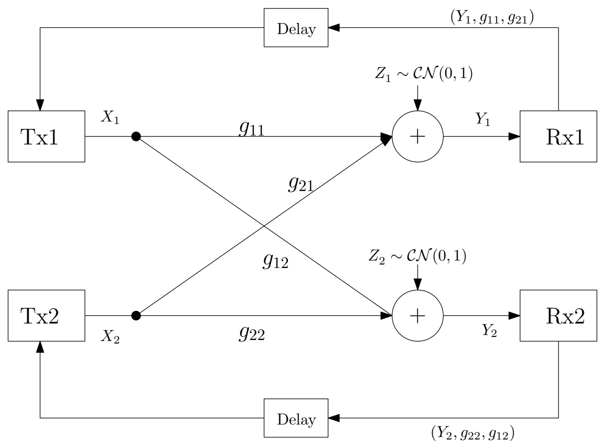


Figure 4.2: The channel model with feedback.

We define symmetric FF-IC to be a FF-IC such that $g_{11} \sim g_{22} \sim g_d$ and $g_{12} \sim g_{21} \sim g_c$ (we use the symbol $\sim$ to indicate random variables following same distribution), all of them being independent. Here g_d and g_c are dummy random variables according to which the direct links and cross links are distributed. We denote $SNR := \mathbb{E}\left[|g_d|^2\right]$, and $INR := \mathbb{E}\left[|g_c|^2\right]$, for the symmetric case.

We use the vector notation $\underline{g}_1 = [g_{11}, g_{21}]$, $\underline{g}_2 = [g_{22}, g_{12}]$ and $\underline{g} = [g_{11}, g_{21}, g_{22}, g_{12}]$. For schemes involving multiple blocks (phases) we use the notation $X_k^{(i)N}$, where k is the user index, i is the block (phase) index and N is the number of symbols per block. The notation $X_k^{(i)}(j)$ indicates the j^{th} symbol in the i^{th} block (phase) of k^{th} user. We explain this in Figure 4.3.

The natural logarithm is denoted by $\ln()$ and the logarithm with base 2 is denoted by $\log()$. Also we define $\log^+(\cdot) := \max(\log(\cdot), 0)$. For obtaining approximate capacity region of ICs, we say that a rate region $\mathcal{R}$ achieves a capacity gap of δ if for any $(R_1, R_2) \in \mathcal{C}$,

[3]IC with rate limited feedback is considered in [VSA12] where outputs are quantized and fed back. Our schemes can also be extended for such cases.

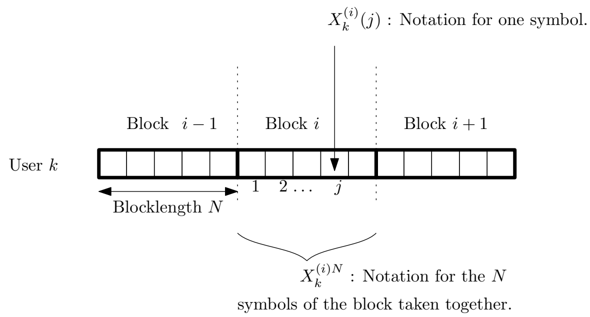


Figure 4.3: The notation for schemes involving multiple blocks (phases).

$(R_1 - \delta, R_2 - \delta) \in \mathcal{R}$, where $\mathcal{C}$ is the capacity region of the channel.

4.3 A logarithmic Jensen's gap characterization for fading models

Definition 4.1. For a given fading model, let $\Phi = \left\{\phi : |g_{ij}|^2 \sim \phi, \text{ for some } i, j \in \{1, 2\}\right\}$ be the set of all probability density functions, that the fading model induce on the channel link strengths $|g_{ij}|^2$, across all operating regimes of the system. We define logarithmic Jensen's gap c_{JG} of the fading model to be

$$\begin{aligned} c_{JG} = \sup_{a \in \mathbb{R}^+, W \sim \phi \in \Phi} & \left(\log\left(a + \mathbb{E}\left[W\right]\right)\right. \\ & \left.-\mathbb{E}\left[\log\left(a + W\right)\right]\right). \end{aligned} \tag{4.3}$$

In other words it is the smallest value of c such that

$$\log\left(a + \mathbb{E}\left[W\right]\right) - \mathbb{E}\left[\log\left(a + W\right)\right] \leq c, \tag{4.4}$$

for any $a \geq 0$, for *any* $\phi \in \Phi$, with W distributed according to ϕ.

The following lemma converts requirement in Definition 4.1 to a simpler form.

Lemma 4.1. *The requirement* $\log\left(a + \mathbb{E}\left[W\right]\right) - \mathbb{E}\left[\log\left(a + W\right)\right] \leq c$ *for any* $a \geq 0$*, is equivalent to* $\log\left(\mathbb{E}\left[W\right]\right) - \mathbb{E}\left[\log\left(W\right)\right] = -\mathbb{E}\left[\log\left(W'\right)\right] \leq c$*, where* $W' = \frac{W}{\mathbb{E}[W]}$*.*

Proof. We first note that letting $a = 0$ in the requirement $\log\left(a + \mathbb{E}\left[W\right]\right) - \mathbb{E}\left[\log\left(a + W\right)\right] \leq c$ shows that $\log\left(\mathbb{E}\left[W\right]\right) - \mathbb{E}\left[\log\left(W\right)\right] = -\mathbb{E}\left[\log\left(W'\right)\right] \leq c$ is necessary.

To prove the converse, note that $\xi(a) = \log(a + \mathbb{E}[W]) - \mathbb{E}[\log(a + W)] \geq 0$ due to Jensen's inequality. Taking derivative with respect to a and again using Jensen's inequality we get

$$(\ln 2)\,\xi'(a) = (a + \mathbb{E}[W])^{-1} - \mathbb{E}\left[(a + W)^{-1}\right] \leq 0. \tag{4.5}$$

Hence, $\xi(a)$ achieves the maximum value at $a = 0$ in the range $[0, \infty)$. Hence we have the equivalent condition

$$\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] \leq c, \tag{4.6}$$

which is equivalent to $-\mathbb{E}[\log(W')] \leq c$. □

Hence, it follows that for any distribution that has a point mass at 0 (for example, bursty interference model [WSD13]), we do not have a finite logarithmic Jensen's gap, since it has $\mathbb{E}[\log(W')] = -\infty$. Now we discuss a few distributions that can be easily shown to have a finite logarithmic Jensen's gap. Note that any finite c, which satisfies Equation (4.4), is an upper bound to the logarithmic Jensen's gap c_{JG}.

4.3.1 Gamma distribution

Gamma distribution generalizes some of the commonly used fading models, including Rayleigh and Nakagami fading. The probability density function for Gamma distribution is given by

$$f(w) = w^{k-1} e^{-\frac{w}{\theta}} / \left(\theta^k \Gamma(k)\right) \tag{4.7}$$

for $w > 0$, where $k > 0$ is the shape parameter, and $\theta > 0$ is the scale parameter.

Proposition 4.2. *If the elements of Φ are Gamma distributed with shape parameter k, they satisfy Equation* (4.4) *with constant $c = \log(e)/k - \log(1 + 1/(2k))$.*

Proof. Using Lemma 4.1, it is sufficient to prove $\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] \leq \log(e)/\alpha - \log(1 + 1/(2\alpha))$. It is known for the Gamma distribution that $\mathbb{E}[W] = k\theta$ and $\mathbb{E}[\ln(W)] = \psi(k) + \ln(\theta)$, where ψ is the digamma function. Therefore,

$$\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] = \log(e)(\ln(k) - \psi(k)) \quad . \tag{4.8}$$

We first use the following property of digamma function $\psi(k) = \psi(k+1) - 1/k$, and then use the inequality $\ln(k+1/2) < \psi(k+1)$from [Bat08, Lemma 1.7]. Hence,

$$\begin{aligned}
&\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] \\
&< \log(e)(\ln(k) - \ln(k+1/2) + 1/k) \qquad (4.9) \\
&= \log(e)/k - \log(1 + 1/(2k)). \qquad (4.10)
\end{aligned}$$

□

Corollary 4.3. *If the elements of* Φ *are exponentially distributed (which corresponds to Rayleigh fading), they satisfy Equation* (4.4) *with constant* $c = 0.86$*.*

Proof. In Rayleigh fading model the $|g_{ij}|^2$ is exponentially distributed. The exponential distribution itself is a special case of Gamma distribution with $k = 1$. Substituting $\alpha = 1$ in (4.10) we get $\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] < 0.86$. □

Nakagami fading can be obtained as a special case of the Gamma distribution; in this case the logarithmic Jensen's gap will depend upon the parameters used in the model.

4.3.2 Weibull distribution

The probability density function for Weibull distribution is given by

$$f(w) = (k/\lambda)(w/\lambda)^{k-1} e^{-(w/\lambda)^k} \qquad (4.11)$$

for $x > 0$ with $k, \lambda > 0$.

Proposition 4.4. *If the elements of* Φ *are Weibull distributed with parameter* k*, they satisfy Equation* (4.4) *with* $c = \gamma \log(e)/k + \log(\Gamma(1 + 1/k))$*, where* γ *is Euler's constant.*

Proof. For Weibull distributed W, we have $\mathbb{E}[W] = \lambda\Gamma\left(1 + \frac{1}{k}\right)$ and $\mathbb{E}[\ln(W)] = \ln(\lambda) - \frac{\gamma}{k}$, where $\Gamma(\cdot)$ denotes the gamma function and γ is the Euler's constant. Hence, it follows that $\log(\mathbb{E}[W]) - \mathbb{E}[\log(W)] \leq \gamma \log(e)/k + \log(\Gamma(1 + 1/k))$. Using Lemma 4.1 concludes the proof. □

Note that exponential distribution can be specialized from Weibull distribution as well, by setting $k = 1$. Hence, we get the tighter gap in the following corollary.

Corollary 4.5. *If the elements of* Φ *are exponentially distributed, they satisfy Equation* (4.4) *with constant* $c = 0.83$.

In the following table we summarize the values we obtain as upper bound on logarithmic Jensen's gap, according to Definition 4.1 and Equation (4.4) for different fading models.

Table 4.1: Upper bound of logarithmic Jensen's gap for different fading models

Fading Model	c
Rayleigh	0.83
Gamma $k = 1$	0.86
Gamma $k = 2$	0.40
Gamma $k = 3$	0.26
Weibull $k = 1$	0.83
Weibull $k = 2$	0.24
Weibull $k = 3$	0.11

4.3.3 Other distributions

Here we give a lemma that can be used together with Lemma 4.1 to verify whether a given fading model has a finite logarithmic Jensen's gap.

Lemma 4.2. *If the cumulative distribution function* $F(w)$ *of* W *satisfies* $F(w) \leq aw^b$ *over* $w \in [0, \epsilon]$ *for some* $a \geq 0$, $b > 0$, *and* $0 < \epsilon \leq 1$, *then*

$$\mathbb{E}\left[\ln(W)\right] \geq \ln(\epsilon) + a\epsilon^b \ln(\epsilon) - \left(a\epsilon^b\right)/b. \tag{4.12}$$

Proof. The condition in this lemma ensures that the probability density function $f(w)$ grows slow enough as $w \to 0^-$ so that $f(w)\ln(w)$ is integrable at 0. Also the behavior for large values of w is not relevant here, since we are looking for a lower bound on $\mathbb{E}[\ln(W)]$. The detailed proof is in Appendix C.3. □

Hence, if the cumulative distribution of the channel gain grows polynomially in a neighborhood of 0, the resulting logarithm becomes integrable, and thus it is possible to find a finite constant c for the Equation (4.4).

4.4 Approximate Capacity Region of FF-IC without feedback

In this section we make use of the logarithmic Jensen's gap characterization to obtain the approximate capacity region of FF-IC with neither feedback nor instantaneous CSIT.

Theorem 4.6. *For a non-feedback FF-IC with a finite logarithmic Jensen's gap c_{JG}, the rate region $\mathcal{R}_{NFB}$ described by* (4.13) *is achievable with $\lambda_{pk} = \min\left(\frac{1}{INR_k}, 1\right)$:*

$$R_1 \le \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 1 \tag{4.13a}$$

$$R_2 \le \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + \lambda_{p1}|g_{12}|^2\right)\right] - 1 \tag{4.13b}$$

$$\begin{aligned} R_1 + R_2 \le\ & \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 2 \end{aligned} \tag{4.13c}$$

$$\begin{aligned} R_1 + R_2 \le\ & \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + |g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \lambda_{p2}|g_{22}|^2 + \lambda_{p1}|g_{12}|^2\right)\right] - 2 \end{aligned} \tag{4.13d}$$

$$\begin{aligned} R_1 + R_2 \le\ & \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + |g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \lambda_{p2}|g_{22}|^2 + |g_{12}|^2\right)\right] - 2 \end{aligned} \tag{4.13e}$$

$$\begin{aligned} 2R_1 + R_2 \le\ & \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + |g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \lambda_{p2}|g_{22}|^2 + |g_{12}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 3 \end{aligned} \tag{4.13f}$$

$$R_1 + 2R_2 \le \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right]$$

$$+\mathbb{E}\left[\log\left(1+\lambda_{p1}\left|g_{11}\right|^2+\left|g_{21}\right|^2\right)\right]$$
$$+\mathbb{E}\left[\log\left(1+\lambda_{p2}\left|g_{22}\right|^2+\lambda_{p1}\left|g_{12}\right|^2\right)\right]-3 \tag{4.13g}$$

and the region $\mathcal{R}_{NFB}$ has a capacity gap of at most $c_{JG}+1$ bits per channel use.

Proof. This region is obtained by a rate-splitting scheme that allocates the private message power proportional to $\frac{1}{\mathbb{E}[inr]}$. The analysis of the scheme and outer bounds are similar to that in [ETW08]. See subsection 4.4.2 for details. □

Remark 4.1. *For the case of Rayleigh fading we obtain a capacity gap of* 1.83 *bits per channel use, following Table 4.1.*

Corollary 4.7. *For FF-IC with finite logarithmic Jensen's gap c_{JG}, instantaneous CSIT cannot improve the capacity region of except by a constant.*

Proof. Our outer bounds in subsection 4.4.2 for the non-feedback IC are valid even when there is instantaneous CSIT. These outer bounds are within constant gap of the rate region $\mathcal{R}_{NFB}$ achieved without instantaneous CSIT. □

Corollary 4.8. *Delayed CSIT cannot improve the capacity region of the FF-IC except by a constant.*

Proof. This follows from the previous corollary, since instantaneous CSIT is always better than delayed CSIT. □

Remark 4.2. *The previous two corollaries are for FF-IC with 2 users and single antennas. It does not contradict the results for MISO broadcast channel, X-channel, MIMO IC and multi-user IC where delayed CSIT or instantaneous CSIT can improve capacity region by more than a constant [MT10, MJS12, KC13, NGJ12, VV12].*

Corollary 4.9. *Within a constant gap, the capacity region of the FF-IC (with finite logarithmic Jensen's gap c_{JG}) can be proved to be same as the capacity region of IC (without fading) with equivalent channel strengths $SNR_i := \mathbb{E}\left[\left|g_{ii}\right|^2\right]$ for $i=1,2,$ and $INR_i := \mathbb{E}\left[\left|g_{ij}\right|^2\right]$ for $i \neq j$.*

Proof. This is an application of the logarithmic Jensen's gap result. The proof is given in Appendix C.4. □

4.4.1 Discussion

It is useful to view Theorem 4.6 in the context of the existing results for the ICs. It is known that for ICs, one can approximately achieve the capacity region by performing superposition coding and allocating a power to the private symbols that is inversely proportional to the strength of the interference caused at the unintended receiver. Consequently, the received interference power is at the noise level, and the private symbols can be safely treated as noise, incurring only a constant rate penalty. At first sight, such a strategy seems impossible for the fading IC, where the transmitters do not have instantaneous channel information. What Theorem 4.6 reveals (with the details in subsection 4.4.2) is that if the fading model has finite logarithmic Jensen's gap, it is sufficient to perform power allocation based on the inverse of average interference strength to approximately achieve the capacity region.

We compare the symmetric rate point achievable for the non-feedback symmetric FF-IC in Figure 4.4. The fading model used is Rayleigh fading. The inner bound in numerical simulation is from Equation (5.5) (which is slightly tighter than (4.13) since some terms in (5.5) are simplified and bounded with the worst case values to obtain (4.13)) in subsection 4.4.2 according to the choice of distributions given in the same subsection. The outer bound is plotted by simulating Equation (4.17) in subsection 4.4.2. The SNR is varied after fixing $\alpha = \frac{\log(INR)}{\log(SNR)}$. The simulation yields a capacity gap of 1.48 bits per channel use for $\alpha = 0.5$ and a capacity gap of 1.51 bits per channel use for $\alpha = 0.25$. Our theoretical analysis for FF-IC gives a capacity gap of $c_{JG} + 1 \leq 1.83$ bits per channel use independent of α, using data from Table 4.1 in Section 4.3.

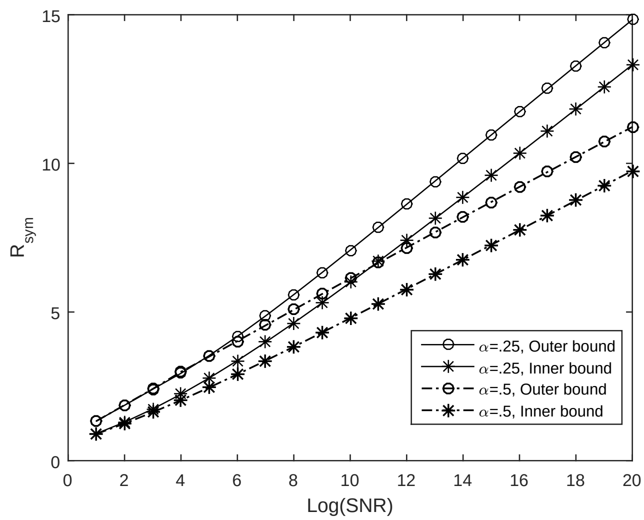


Figure 4.4: Comparison of outer and inner bounds with given $\alpha = \frac{\log(INR)}{\log(SNR)}$ for non-feedback symmetric FF-IC at the symmetric rate point. For high SNR, the capacity gap is approximately 1.48 bits per channel use for $\alpha = 0.5$ and 1.51 bits per channel use for $\alpha = 0.25$ from the numerics. Our theoretical analysis yields gap as 1.83 bits per channel use independent of α.

4.4.2 Proof of Theorem 4.6

From [CMG08] we obtain that a Han-Kobayashi scheme for IC can achieve the following rate region for all $p(u_1)\,p(u_2)\,p(x_1|u_1)\,p(x_2|u_2)$. Note that we use $\left(Y_i, \underline{g_i}\right)$ instead of (Y_i) in the actual result from [CMG08] to account for the fading.

$$R_1 \leq I\left(X_1; Y_1, \underline{g_1}|U_2\right) \tag{4.14a}$$

$$R_2 \leq I\left(X_2; Y_2, \underline{g_2}|U_1\right) \tag{4.14b}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I\left(X_2, U_1; Y_2, \underline{g_2}\right) \\ & + I\left(X_1; Y_1, \underline{g_1}|U_1, U_2\right) \end{aligned} \tag{4.14c}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I\left(X_1, U_2; Y_1, \underline{g_1}\right) \\ & + I\left(X_2; Y_2, \underline{g_2}|U_1, U_2\right) \end{aligned} \tag{4.14d}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I\left(X_1, U_2; Y_1, \underline{g_1}|U_1\right) \\ & + I\left(X_2, U_1; Y_2, \underline{g_2}|U_2\right) \end{aligned} \tag{4.14e}$$

$$\begin{aligned} 2R_1 + R_2 \leq\ & I\left(X_1, U_2; Y_1, \underline{g_1}\right) + I\left(X_1; Y_1, \underline{g_1}|U_1, U_2\right) \\ & + I\left(X_2, U_1; Y_2, \underline{g_2}|U_2\right) \end{aligned} \tag{4.14f}$$

$$\begin{aligned} R_1 + 2R_2 \leq\ & I\left(X_2, U_1; Y_2, \underline{g_2}\right) + I\left(X_2; Y_2, \underline{g_2}|U_1, U_2\right) \\ & + I\left(X_1, U_2; Y_1, \underline{g_1}|U_1\right). \end{aligned} \tag{4.14g}$$

Now similar to that in [ETW08], choose mutually independent Gaussian input distributions U_k, X_{pk} to generate X_k.

$$U_k \sim \mathcal{CN}\left(0, \lambda_{ck}\right), \quad X_{pk} \sim \mathcal{CN}\left(0, \lambda_{pk}\right), \qquad k \in \{1, 2\} \tag{4.15}$$

$$X_1 = U_1 + X_{p1}, \quad X_2 = U_2 + X_{p2}, \tag{4.16}$$

where $\lambda_{ck} + \lambda_{pk} = 1$ and $\lambda_{pk} = \min\left(1/INR_k, 1\right)$. Here we introduced the rate-splitting using the average inr. On evaluating the region described by (5.5) with this choice of input distribution, we get the region described by (4.13); the calculations are deferred to Appendix C.1.

Claim 4.1. *An outer bound for the non-feedback case is given by* (4.17)*:*

$$R_1 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2\right)\right] \tag{4.17a}$$

$$R_2 \leq \mathbb{E}\left[\log\left(1+|g_{22}|^2\right)\right] \tag{4.17b}$$

$$\begin{aligned} R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(1+|g_{22}|^2+|g_{12}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{11}|^2\left(1+|g_{12}|^2\right)^{-1}\right)\right] \end{aligned} \tag{4.17c}$$

$$\begin{aligned} R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(1+|g_{11}|^2+|g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{22}|^2\left(1+|g_{21}|^2\right)^{-1}\right)\right] \end{aligned} \tag{4.17d}$$

$$\begin{aligned} R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(1+|g_{21}|^2+\frac{|g_{11}|^2}{1+|g_{12}|^2}\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{12}|^2+\frac{|g_{22}|^2}{1+|g_{21}|^2}\right)\right] \end{aligned} \tag{4.17e}$$

$$\begin{aligned} 2R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(1+|g_{11}|^2+|g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{12}|^2+\frac{|g_{22}|^2}{1+|g_{21}|^2}\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{11}|^2\left(1+|g_{12}|^2\right)^{-1}\right)\right] \end{aligned} \tag{4.17f}$$

$$\begin{aligned} R_1 + 2R_2 \leq\ & \mathbb{E}\left[\log\left(1+|g_{22}|^2+|g_{12}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{21}|^2+\frac{|g_{11}|^2}{1+|g_{12}|^2}\right)\right] \\ & + \mathbb{E}\left[\log\left(1+|g_{22}|^2\left(1+|g_{21}|^2\right)^{-1}\right)\right]. \end{aligned} \tag{4.17g}$$

Proof. The outer bounds (4.17a) and (4.17b) are easily derived by removing the interference from the other user by providing it as side-information.

The outer bound in Equation (4.17e) follows from [ETW08, Theorem 1]. Those in Equation (4.17f) and Equation (4.17g) follow from [ETW08, Theorem 4]. We just need to modify the theorems from [ETW08] for the fading case by treating $\left(Y_i, \underline{g}_i\right)$ as output, and using the i.i.d property of the channels. We illustrate the procedure for Equation (4.17g) in Appendix C.2. Equation (4.17e) and Equation (4.17f) can be derived similarly.

The derivation of outer bounds (4.17c) and (4.17d) uses similar techniques as for Equation (4.17g). We derive Equation (4.17d) in Appendix C.2. Equation (4.17c) follows due to symmetry. □

Claim 4.2. *The gap between the inner bound* (4.13) *and the outer bound* (4.17) *for the non-feedback case is at most* $c_{JG} + 1$ *bits per channel use.*

Proof. The proof for the capacity gap uses the logarithmic Jensen's gap property of the fading model. Denote the gap between the first outer bound (4.17a) and first inner bound (4.13a) by δ_1, for the second pair denote the gap by δ_2, and so on. By inspection $\delta_1 \leq 1$ and $\delta_2 \leq 1$. Now

$$\begin{aligned}
\delta_3 &= \mathbb{E}\left[\log\left(1 + |g_{11}|^2\left(1 + |g_{12}|^2\right)^{-1}\right)\right] \\
&\quad - \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] + 2 && (4.18)\\
&\overset{(a)}{\leq} \mathbb{E}\left[\log\left(1 + |g_{11}|^2 (1 + INR_1)^{-1}\right)\right] \\
&\quad - \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2\right)\right] + 2 + c_{JG} && (4.19)\\
&\overset{(b)}{\leq} 2 + c_{JG}. && (4.20)
\end{aligned}$$

The step (a) follows from Jensen's inequality and logarithmic Jensen's gap property of $|g_{12}|^2$. The step (b) follows because $\lambda_{p1} = \min\left(\frac{1}{INR_1}, 1\right) \geq \frac{1}{INR_1+1}$. Similarly, we can bound the other δ's and gather the inequalities as:

$$\delta_1, \delta_2 \leq 1; \quad \delta_3, \delta_4 \leq 2 + c_{JG} \tag{4.21}$$

$$\delta_5 \leq 2 + 2c_{JG}; \quad \delta_6, \delta_7 \leq 3 + 2c_{JG}. \tag{4.22}$$

For δ_5, δ_6, and δ_7 we have to use the logarithmic Jensen's gap property twice and hence $2c_{JG}$ appears. We note that δ_1 is associated with bounding R_1, δ_2 with R_2, $(\delta_3, \delta_4, \delta_5)$ with $R_1 + R_2$, δ_6 with $2R_1 + R_2$ and δ_7 with $R_1 + 2R_2$. Hence, it follows that the capacity gap is at most $\max\left(\delta_1, \delta_2, \frac{\delta_3}{2}, \frac{\delta_4}{2}, \frac{\delta_5}{2}, \frac{\delta_6}{3}, \frac{\delta_7}{3}\right) \leq c_{JG} + 1$ bits per channel use. □

4.4.3 Fast Fading Interference MAC channel

We now consider the interference MAC channel [PDT09] with fading links (Figure 4.5), where we can obtain an approximate capacity result similar to the FF-IC. This setup has

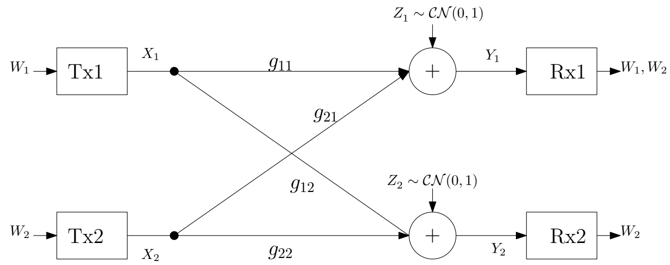


Figure 4.5: Fast Fading Interference MAC channel

similar network structure as FF-IC. However Rx1 needs to decode the messages from both the two transmitters, while Rx2 needs to decode only the message from Tx2.

Theorem 4.10. *A rate splitting scheme based on average INR can achieve the approximate capacity region of fast fading interference MAC channel with a finite logarithmic Jensen's gap c_{JG}, within $1 + \frac{1}{2}c_{JG}$ bits per channel use.*

Proof. The proof is by extending the techniques used in [PDT09] and using similar calculations as for the FF-IC. Details are in Appendix C.5. □

4.5 Approximate Capacity Region of FF-IC with feedback

In this section we make use of the logarithmic Jensen's gap characterization to obtain the approximate capacity region of FF-IC with output and channel state feedback, but transmitters having no prior knowledge of channel states. Under the feedback model, after each reception, each receiver reliably feeds back the received symbol and the channel states to its corresponding transmitter. For example, at time l, Tx1 receives $(Y_1\left(l-1\right), g_{11}(l-1), g_{21}(l-1))$ from Rx1. Thus $X_1(l)$ is allowed to be a function of $\left(W_1, \left\{Y_1\left(k\right), g_{11}(k), g_{21}(k)\right\}_{k<l}\right)$. The model is described in section 5.2 and is illustrated with Figure 4.2 in the same section.

Theorem 4.11. *For a feedback FF-IC with a finite logarithmic Jensen's gap c_{JG} , the rate region $\mathcal{R}_{FB}$ described by* (4.23) *is achievable for* $0 \leq |\rho|^2 \leq 1$, $0 \leq \theta < 2\pi$ *with* $\lambda_{pk} = \min\left(\frac{1}{INR_k}, 1-|\rho|^2\right)$*:*

$$R_1 \leq \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2\right.\right.$$

$$+2|\rho|^2 Re\left(e^{i\theta}g_{11}g_{21}^*\right)+1\Big)\Big]-1 \tag{4.23a}$$

$$R_1 \le \mathbb{E}\left[\log\left(1+\left(1-|\rho|^2\right)|g_{12}|^2\right)\right]$$

$$+\mathbb{E}\left[\log\left(1+\lambda_{p1}|g_{11}|^2+\lambda_{p2}|g_{21}|^2\right)\right]-2 \tag{4.23b}$$

$$R_2 \le \mathbb{E}\Big[\log\Big(|g_{22}|^2+|g_{12}|^2$$

$$+2|\rho|^2 Re\left(e^{i\theta}g_{22}^*g_{12}\right)+1\Big)\Big]-1 \tag{4.23c}$$

$$R_2 \le \mathbb{E}\left[\log\left(1+\left(1-|\rho|^2\right)|g_{21}|^2\right)\right]$$

$$+\mathbb{E}\left[\log\left(1+\lambda_{p2}|g_{22}|^2+\lambda_{p1}|g_{12}|^2\right)\right]-2 \tag{4.23d}$$

$$R_1+R_2 \le \mathbb{E}\Big[\log\Big(|g_{22}|^2+|g_{12}|^2$$

$$+2|\rho|^2 Re\left(e^{i\theta}g_{22}^*g_{12}\right)+1\Big)\Big]$$

$$+\mathbb{E}\left[\log\left(1+\lambda_{p1}|g_{11}|^2+\lambda_{p2}|g_{21}|^2\right)\right]-2 \tag{4.23e}$$

$$R_1+R_2 \le \mathbb{E}\Big[\log\Big(|g_{11}|^2+|g_{21}|^2$$

$$+2|\rho|^2 Re\left(e^{i\theta}g_{11}g_{21}^*\right)+1\Big)\Big]$$

$$+\mathbb{E}\left[\log\left(1+\lambda_{p2}|g_{22}|^2+\lambda_{p1}|g_{12}|^2\right)\right]-2 \tag{4.23f}$$

and the region $\mathcal{R}_{FB}$ has a capacity gap of at most $c_{JG}+2$ bits per channel use.

Proof. The proof is in subsection 4.5.1. □

Remark 4.3. *For the case of Rayleigh fading we obtain a capacity gap of* 2.83 *bits per channel use, following Table 4.1.*

Corollary 4.12. *Instantaneous CSIT cannot improve the capacity region of the FF-IC (with finite logarithmic Jensen's gap c_{JG}) with feedback and delayed CSIT except for a constant.*

Proof. Our outer bounds in subsection 4.5.1 for feedback case are valid even when there is instantaneous CSIT. These outer bounds are within constant gap of the rate region $\mathcal{R}_{FB}$ achieved using only feedback and delayed CSIT. □

Corollary 4.13. *If the phases of the links g_{ij} are uniformly distributed in $[0, 2\pi]$, then within a constant gap, the capacity region of the feedback FF-IC (with finite logarithmic Jensen's* *gap c_{JG}) with feedback and delayed CSIT can be proved to be same as the capacity region of a feedback IC (without fading) with equivalent channel strengths $SNR_i := \mathbb{E}\left[|g_{ii}|^2\right]$ for $i = 1, 2$, and $INR_i := \mathbb{E}\left[|g_{ij}|^2\right]$ for $i \neq j$.*

Proof. This is again an application of the logarithmic Jensen's gap result. The proof is given in Appendix C.6. □

4.5.1 Proof of Theorem 4.11

Note that since the receivers know their respective incoming channel states, we can view the effective channel output at Rxi as the pair $\left(Y_i, \underline{g}_i\right)$. Then the block Markov scheme of [ST11, Lemma 1] implies that the rate pairs (R_1, R_2) satisfying

$$R_1 \leq I\left(U, U_2, X_1; Y_1, \underline{g_1}\right) \tag{4.24a}$$

$$\begin{aligned} R_1 \leq\ & I\left(U_1; Y_2, \underline{g_2} | U, X_2\right) \\ & + I\left(X_1; Y_1, \underline{g_1} | U_1, U_2, U\right) \end{aligned} \tag{4.24b}$$

$$R_2 \leq I\left(U, U_1, X_2; Y_2, \underline{g_2}\right) \tag{4.24c}$$

$$\begin{aligned} R_2 \leq\ & I\left(U_2; Y_1, \underline{g_1} | U, X_1\right) \\ & + I\left(X_2; Y_2, \underline{g_2} | U_1, U_2, U\right) \end{aligned} \tag{4.24d}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I\left(X_1; Y_1, \underline{g_1} | U_1, U_2, U\right) \\ & + I\left(U, U_1, X_2; Y_2, \underline{g_2}\right) \end{aligned} \tag{4.24e}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I\left(X_2; Y_2, \underline{g_2} | U_1, U_2, U\right) \\ & + I\left(U, U_2, X_1; Y_1, \underline{g_1}\right) \end{aligned} \tag{4.24f}$$

for all $p(u)\, p(u_1|u)\, p(u_2|u)\, p(x_1|u_1, u)\, p(x_2|u_2, u)$ are achievable. We choose the input distribution according to

$$U \sim \mathcal{CN}\left(0, |\rho|^2\right), U_k \sim \mathcal{CN}\left(0, \lambda_{ck}\right) \tag{4.25}$$

$$X_{pk} \sim \mathcal{CN}\left(0, \lambda_{pk}\right)$$

$$X_1 = e^{i\theta} U + U_1 + X_{p1}, \quad X_2 = U + U_2 + X_{p2} \tag{4.26}$$

$(U, U_k, X_{pk}$ being mutually independent) with $0 \leq |\rho|^2 \leq 1$, $0 \leq \theta < 2\pi$, $\lambda_{ck} + \lambda_{pk} = 1 - |\rho|^2$ and $\lambda_{pk} = \min\left(1/INR_k, 1 - |\rho|^2\right)$.

With this choice of λ_{pk} we perform the rate-splitting according to the average *inr* in place of rate-splitting based on the constant *inr*. Note that we have introduced an extra rotation θ for the first transmitter, which will become helpful in proving the capacity gap by allowing us to choose a point in inner bound for every point in outer bound (see proof of claim 4.3). On evaluating the terms in (5.14) for this choice of input distribution, we get the inner bound described by (4.23); the calculations are deferred to Appendix C.7.

An outer bound for the feedback case is given by (4.27) with $0 \leq |\rho| \leq 1$ (ρ being a complex number):

$$R_1 \leq \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2\mathrm{Re}\left(\rho g_{11} g_{21}^*\right) + 1\right)\right] \tag{4.27a}$$

$$\begin{aligned} R_1 \leq\ & \mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{12}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \frac{\left(1 - |\rho|^2\right)|g_{11}|^2}{1 + \left(1 - |\rho|^2\right)|g_{12}|^2}\right)\right] \end{aligned} \tag{4.27b}$$

$$R_2 \leq \mathbb{E}\left[\log\left(|g_{22}|^2 + |g_{12}|^2 + 2\mathrm{Re}\left(\rho g_{22}^* g_{12}\right) + 1\right)\right] \tag{4.27c}$$

$$\begin{aligned} R_2 \leq\ & \mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{21}|^2\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \frac{\left(1 - |\rho|^2\right)|g_{22}|^2}{1 + \left(1 - |\rho|^2\right)|g_{21}|^2}\right)\right] \end{aligned} \tag{4.27d}$$

$$\begin{aligned} R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(|g_{22}|^2 + |g_{12}|^2 + 2\mathrm{Re}\left(\rho g_{22}^* g_{12}\right) + 1\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \frac{\left(1 - |\rho|^2\right)|g_{11}|^2}{1 + \left(1 - |\rho|^2\right)|g_{12}|^2}\right)\right] \end{aligned} \tag{4.27e}$$

$$\begin{aligned} R_1 + R_2 \leq\ & \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2\mathrm{Re}\left(\rho g_{11} g_{21}^*\right) + 1\right)\right] \\ & + \mathbb{E}\left[\log\left(1 + \frac{\left(1 - |\rho|^2\right)|g_{22}|^2}{1 + \left(1 - |\rho|^2\right)|g_{21}|^2}\right)\right]. \end{aligned} \tag{4.27f}$$

The outer bounds can be easily derived following the proof techniques from [ST11, Theorem 3] using $\mathbb{E}\left[X_1 X_2^*\right] = \rho$, treating $\left(Y_i, \underline{g}_i\right)$ as output, and using the i.i.d property of the channels. The calculations are deferred to Appendix C.8.

Claim 4.3. *The gap between the inner bound* (4.23) *and the outer bound* (4.27) *for the feedback case is at most* $c_{JG} + 2$ *bits per channel use.*

Proof. Denote the gap between the first outer bound (4.27a) and inner bound (4.23a) by δ_1, for the second pair denote the gap by δ_2, and so on. For comparing the gap between regions we choose the inner bound point with same $|\rho|$ as in any given outer bound point. The rotation θ for the first transmitter also becomes important in proving a constant gap capacity result. We choose θ in the inner bound to match $\arg(\rho)$ in the outer bound. With this choice we get

$$\begin{aligned}\delta_1 &= \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|\rho| \operatorname{Re}\left(e^{i\theta} g_{11} g_{21}^*\right) + 1\right)\right] \\ &\quad - \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|\rho|^2 \operatorname{Re}\left(e^{i\theta} g_{11} g_{21}^*\right) + 1\right)\right] + 1 \qquad (4.28)\\ &= \mathbb{E}\left[\log\left(\frac{1 + \frac{1}{|g_{11}|^2+|g_{21}|^2} + |\rho|\left(\frac{2\operatorname{Re}\left(e^{i\theta} g_{11} g_{21}^*\right)}{|g_{11}|^2+|g_{21}|^2}\right)}{1 + \frac{1}{|g_{11}|^2+|g_{21}|^2} + |\rho|^2\left(\frac{2\operatorname{Re}\left(e^{i\theta} g_{11} g_{21}^*\right)}{|g_{11}|^2+|g_{21}|^2}\right)}\right)\right] + 1. \qquad (4.29)\end{aligned}$$

We have $\left|\frac{2\operatorname{Re}\left(e^{i\theta} g_{11} g_{21}^*\right)}{|g_{11}|^2+|g_{21}|^2}\right| = \frac{\left|e^{-i\theta} g_{11}^* g_{21} + e^{i\theta} g_{11} g_{21}^*\right|}{|g_{11}|^2+|g_{21}|^2} \leq 1$, hence we call $\frac{e^{-i\theta} g_{11}^* g_{21} + e^{i\theta} g_{11} g_{21}^*}{|g_{11}|^2+|g_{21}|^2} = \sin\varphi$ and let $|g_{11}|^2 + |g_{21}|^2 = r^2$. Therefore,

$$\delta_1 = \mathbb{E}\left[\log\left(\frac{1 + 1/r^2 + |\rho|\sin\varphi}{1 + 1/r^2 + |\rho|^2 \sin\varphi}\right)\right] + 1. \qquad (4.30)$$

If $\sin\phi < 0$, then $\frac{1+1/r^2+|\rho|\sin\varphi}{1+1/r^2+|\rho|^2\sin\varphi} \leq 1$. Otherwise, if $\sin\phi > 0$, then $\frac{1+1/r^2+|\rho|\sin\varphi}{1+1/r^2+|\rho|^2\sin\varphi} = 1 + \frac{\left(|\rho|-|\rho|^2\right)\sin\varphi}{1+1/r^2+|\rho|^2\sin\varphi} \leq 2$ since $0 \leq \left(|\rho| - |\rho|^2\right)\sin\phi \leq 1$ and $1 + 1/r^2 + |\rho|^2 \sin\phi > 1$. Hence, we always get

$$\delta_1 \leq 2. \qquad (4.31)$$

Now we consider the gap δ_2 between the second inequality (4.27b) of the outer bound and the second inequality (4.23b) of the inner bound.

$$\begin{aligned}\delta_2 &= \mathbb{E}\left[\log\left(1 + \frac{\left(1-|\rho|^2\right)|g_{11}|^2}{1 + \left(1-|\rho|^2\right)|g_{12}|^2}\right)\right] \\ &\quad - \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] + 2 \qquad (4.32)\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \mathbb{E}\left[\log\left(1+\left(1-|\rho|^2\right)INR_1+\left(1-|\rho|^2\right)|g_{11}|^2\right)\right] \\
&\quad -\log\left(1+\left(1-|\rho|^2\right)INR_1\right)+c_{JG} \\
&\quad -\mathbb{E}\left[\log\left(1+\lambda_{p1}|g_{11}|^2+\lambda_{p2}|g_{21}|^2\right)\right]+2 &(4.33)\\
&\leq \mathbb{E}\left[\log\left(1+\frac{\left(1-|\rho|^2\right)|g_{11}|^2}{1+\left(1-|\rho|^2\right)INR_1}\right)\right] &(4.34)\\
&\quad -\mathbb{E}\left[\log\left(1+\lambda_{p1}|g_{11}|^2\right)\right]+2+c_{JG} \\
&\overset{(b)}{\leq} 2+c_{JG}, &(4.35)
\end{aligned}
$$

where (a) follows by using logarithmic Jensen's gap property on $|g_{12}|^2$ and Jensen's inequality. The step (b) follows because

$$
\begin{aligned}
\frac{\left(1-|\rho|^2\right)}{1+\left(1-|\rho|^2\right)INR_1} &= \frac{1}{\frac{1}{1-|\rho|^2}+INR_1} &(4.36)\\
&\leq \min\left(1/INR_1, 1-|\rho|^2\right)=\lambda_{p1} &(4.37)
\end{aligned}
$$

Similarly, by inspection of the other bounding inequalities we can gather the inequalities on the δ's as:

$$
\delta_1,\delta_3\leq 2;\quad \delta_2,\delta_4\leq c_{JG}+2;\quad \delta_5,\delta_6\leq c_{JG}+3. \tag{4.38}
$$

We note that (δ_1,δ_2) is associated with bounding R_1, (δ_3,δ_4) with R_2, (δ_5,δ_6) with R_1+R_2. Hence it follows that the capacity gap is at most $\max\left(\delta_1,\delta_2,\delta_3,\delta_4,\frac{\delta_5}{2},\frac{\delta_6}{2}\right)\leq c_{JG}+2$ bits per channel use. □

4.6 Approximate capacity of feedback FF-IC using point-to-point codes

As the third illustration for the usefulness of logarithmic Jensen's gap, we propose a strategy that does not make use of rate-splitting, superposition coding or joint decoding for the feedback case, which achieves the entire capacity region for 2-user symmetric FF-ICs to within a constant gap. This constant gap is dictated by the logarithmic Jensen's gap for the fading model. Our scheme only uses point-to-point codes, and a feedback scheme based on amplify-and-forward relaying, similar to the one proposed in [ST11].

The main idea behind the scheme is to have one of the transmitters initially send a very densely modulated block of data, and then refine this information using feedback and amplify-and-forward relaying for the following blocks, in a fashion similar to the Schalkwijk-Kailath scheme [SK66], while treating the interference as noise. Such refinement effectively induces a 2-tap point-to-point inter-symbol-interference (ISI) channel at the unintended receiver, and a point-to-point feedback channel for the intended receiver. As a result, both receivers can decode their intended information using only point-to-point codes.

Consider the symmetric FF-IC, where the channel statistics are symmetric and independent, *i.e.*, $g_{11}(l) \sim g_{22}(l) \sim g_d$ and $g_{12}(l) \sim g_{21}(l) \sim g_c$ and all the random variables independent of each other. We consider n transmission phases, each phase having a block length of N. For Tx1, generate 2^{nNR_1} codewords $\left(X_1^{(1)N}, \ldots, X_1^{(n)N}\right)$ i.i.d according to $\mathcal{CN}(0,1)$ and encode its message $W_1 \in \left\{1, \ldots, 2^{nNR_1}\right\}$ onto $\left(X_1^{(1)N}, \ldots, X_1^{(n)N}\right)$. For Tx2, generate 2^{nNR_2} codewords $X_2^{(1)N} = X_2^N$ i.i.d according to $\mathcal{CN}(0,1)$ and encode its message $W_2 \in \left\{1, \ldots, 2^{nNR_2}\right\}$ onto $X_2^{(1)N} = X_2^N$. Note that for Tx2 the message is encoded into N length sequence to be transmitted at first phase, whereas for Tx1 the message is encoded into nN length sequence to be transmitted through n phases.

Tx1 sends $X_1^{(i)N}$ in phase i. Tx2 sends $X_2^{(1)N} = X_2^N$ in phase 1. At the beginning of phase $i > 1$, Tx2 receives

$$Y_2^{(i-1)N} = g_{22}^{(i-1)N} X_2^{(i-1)N} + g_{12}^{(i-1)N} X_1^{(i-1)N} + Z_2^{(i-1)N} \tag{4.39}$$

from feedback. It can remove $g_{22}^{(i-1)N} X_2^{(i-1)N}$ from $Y_2^{(i-1)N}$ to obtain $g_{12}^{(i-1)N} X_1^{(i-1)N} + Z_2^{(i-1)N}$. Tx2 then transmits the resulting interference-plus-noise after power scaling as $X_2^{(i)N}$, i.e.

$$X_2^{(i)N} = \frac{g_{12}^{(i-1)N} X_1^{(i-1)N} + Z_2^{(i-1)N}}{\sqrt{1 + INR}}. \tag{4.42}$$

Thus, in phase $i > 1$, Rx2 receives

$$Y_2^{(i)N} = g_{22}^{(i)N} X_2^{(i)N} + g_{12}^{(i)N} X_1^{(i)N} + Z_2^{(i)N} \tag{4.43}$$

$$K_{\mathbf{Y_1}}(1) = \left[1 + |g_{11}(1)|^2 + |g_{21}(1)|^2\right], K_{\mathbf{Y_1}}(2) = \begin{bmatrix} |g_{11}(2)|^2 + \frac{|g_{21}(2)|^2\left(|g_{12}(1)|^2+1\right)}{1+INR} + 1 & \frac{g_{11}^*(1)g_{21}(2)g_{12}(1)}{\sqrt{1+INR}} \\ \frac{g_{11}(1)g_{21}^*(2)g_{12}^*(1)}{\sqrt{1+INR}} & |g_{11}(1)|^2 + |g_{21}(1)|^2 + 1 \end{bmatrix}, \tag{4.40}$$

$$K_{\mathbf{Y_1}}(l) = \begin{bmatrix} |g_{11}(l)|^2 + \frac{|g_{21}(l)|^2\left(|g_{12}(l-1)|^2+1\right)}{1+INR} + 1 & \left[\frac{g_{11}^*(l-1)g_{21}(l)g_{12}(l-1)}{\sqrt{1+INR}}, 0_{l-2}\right] \\ \left[\frac{g_{11}^*(l-1)g_{21}(l)g_{12}(l-1)}{\sqrt{1+INR}}, 0_{l-2}\right]^\dagger & K_{\mathbf{Y_1}}(l-1) \end{bmatrix}. \tag{4.41}$$

$$= g_{22}^{(i)N}\left(\frac{g_{12}^{(i-1)N}X_1^{(i-1)N} + Z_2^{(i-1)N}}{\sqrt{1+INR}}\right)$$
$$+ g_{12}^{(i)N}X_1^{(i)N} + Z_2^{(i)N} \tag{4.44}$$

and feeds it back to Tx2 for phase $i+1$. The transmission scheme is summarized in Table 4.2. Note that for phase $i = 1$ Tx1 receives

$$Y_1^{(1)N} = g_{11}^{(1)N}X_1^{(1)N} + g_{21}^{(1)N}X_2^{(1)N} + Z_1^{(1)N} \tag{4.45}$$

and for phase $i > 1$ Tx1 observes a block ISI channel since it receives

$$Y_1^{(i)N} = g_{11}^{(i)N}X_1^{(i)N} + g_{21}^{(i)N}\left(\frac{g_{12}^{(i-1)N}X_1^{(i-1)N} + Z_2^{(i-1)N}}{\sqrt{1+INR}}\right)$$
$$+ Z_1^{(i)N} \tag{4.46}$$
$$= g_{11}^{(i)N}X_1^{(i)N} + \left(\frac{g_{21}^{(i)N}g_{12}^{(i-1)N}}{\sqrt{1+INR}}\right)X_1^{(i-1)N}$$
$$+ \tilde{Z}_1^{(i)N}, \tag{4.47}$$

where $\tilde{Z}_1^{(i)N} = Z_1^{(i)N} + g_{21}^{(i)N}Z_2^{(i-1)N}(1+INR)^{-1/2}$.

Table 4.2: Transmitted symbols in n-phase scheme for symmetric FF-IC with feedback

User	Phase 1	Phase 2	.	.	Phase n
1	$X_1^{(1)N}$	$X_1^{(2)N}$	.	.	$X_1^{(n)N}$
2	$X_2^{(1)N}$	$\frac{g_{12}^{(1)N}X_1^{(1)N}+Z_2^{(1)N}}{\sqrt{1+INR}}$	.	.	$\frac{g_{12}^{(n-1)N}X_1^{(n-1)N}+Z_2^{(n-1)N}}{\sqrt{1+INR}}$

At the end of n blocks, Rx1 collects $\mathbf{Y_1}^N = \left(Y_1^{(1)N}, \ldots, Y_1^{(n)N}\right)$ and decodes W_1

such that $\left(\mathbf{X_1}^N(W_1), \mathbf{Y_1}^N\right)$ is jointly typical (where $\mathbf{X_1}^N = \left(X_1^{(1)N}, \ldots, X_1^{(n)N}\right)$) treating $X_2^{(1)N} = X_2^N$ as noise. At Rx2, channel outputs over n phases can be combined with an appropriate scaling so that the interference-plus-noise at phases $\{1, \ldots, n-1\}$ are successively canceled, *i.e.*, an effective point-to-point channel can be generated through $\tilde{Y}_2^N = Y_2^{(n)N} + \sum_{i=1}^{n-1}\left(\prod_{j=i+1}^{n}\frac{-g_{22}^{(j)N}}{\sqrt{1+INR}}\right)Y_2^{(i)N}$ (see the analysis in the subsection 4.6.1 for details). Note that this can be viewed as a block version of the Schalkwijk-Kailath scheme [SK66]. Given the effective channel $\tilde{Y}_2^N$, the receiver can simply use point-to-point typicality decoding to recover W_2, treating the interference in phase n as noise.

Theorem 4.14. *For a symmetric FF-IC with a finite logarithmic Jensen's gap c_{JG}, the rate pair*

$$\begin{aligned}(R_1, R_2) = \Big(&\log\left(1 + SNR + INR\right) - 3c_{JG} - 2\,,\\ &\mathbb{E}\left[\log^+\left[|g_d|^2 / \left(1 + INR\right)\right]\right]\Big)\end{aligned}$$

is achievable by the scheme. The scheme together with switching the roles of users and time-sharing, achieves the capacity region of symmetric feedback IC within $3c_{JG} + 2$ bits per channel use.

Proof. The proof follows from the analysis in the following subsection. □

4.6.1 Analysis of Point-to-Point Codes for Symmetric FF-ICs

We now provide the analysis for the scheme, going through the decoding at the two receivers and then looking at the capacity gap for the achievable region.

4.6.1.1 Decoding at Rx1

At the end of n blocks Rx1 collects $\mathbf{Y_1}^N = \left(Y_1^{(1)N}, \ldots Y_1^{(n)N}\right)$ and decodes W_1 such that $\left(\mathbf{X_1}^N(W_1), \mathbf{Y_1}^N\right)$ is jointly typical, where $\mathbf{X_1}^N = \left(X_1^{(1)N}, \ldots X_1^{(n)N}\right)$. The joint typicality is considered according the product distribution $p^N(\mathbf{X_1}, \mathbf{Y_1})$, where

$$p(\mathbf{X_1}, \mathbf{Y_1}) = p\left(\left(X_1^{(1)}, \ldots X_1^{(n)}\right), \left(Y_1^{(1)}, \ldots Y_1^{(n)}\right)\right) \tag{4.48}$$

is a joint Gaussian distribution, dictated by the following equations that arise from our n-phase scheme:

$$Y_1^{(1)} = g_{11}^{(1)} X_1^{(1)} + g_{21}^{(1)} X_2^{(1)} + Z_1^{(1)}. \tag{4.49}$$

And for $i = 2, 3, \ldots, n$:

$$Y_1^{(i)} = g_{11}^{(i)} X_1^{(i)} + g_{21}^{(i)} \left(\frac{g_{12}^{(i-1)} X_1^{(i-1)} + Z_2^{(i-1)}}{\sqrt{1 + INR}} \right) + Z_1^{(i)} \tag{4.50}$$

with $X_1^{(i)}, X_2^{(1)}, Z_1^{(i)}, Z_2^{(i-1)}$ being i.i.d $\mathcal{CN}(0,1)$. Essentially $X_2^{(1)}, Z_1^{(i)}$ are both Gaussian noise for Rx1.

Using standard techniques it follows that for the n-phase scheme, as $N \to \infty$ user 1 can achieve the rate $\frac{1}{n}\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)| \left|K_{\mathbf{Y_1}|\mathbf{X_1}}(n)\right|^{-1}\right)\right]$, where $|K_{\mathbf{Y_1}}(n)|$ denotes the determinant of covariance matrix for the n-phase scheme, as defined in (4.40), and (4.41), where 0_{l-2} is $(l-2)$ length zero vector, $\dagger$ indicates Hermitian conjugate, $g_{11}(i) \sim g_d$ i.i.d and $g_{12}(i), g_{21}(i) \sim g_c$ i.i.d. Letting $n \to \infty$, Rx1 can achieve the rate $R_1 = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)| \left|K_{\mathbf{Y_1}|\mathbf{X_1}}(n)\right|^{-1}\right)\right]$. We need to evaluate $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)| \left|K_{\mathbf{Y_1}|\mathbf{X_1}}(n)\right|^{-1}\right)\right]$. The following lemma gives a lower bound on $\frac{1}{n}\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)|\right)\right]$.

Lemma 4.3.

$$(1/n)\,\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)|\right)\right] \geq (1/n)\log\left(\left|\hat{K}_{\mathbf{Y_1}}(n)\right|\right) - 3c_{JG},$$

where $\hat{K}_{\mathbf{Y_1}}(n)$ is obtained from $K_{\mathbf{Y_1}}(n)$ by replacing $g_{12}(i)$'s,$g_{21}(i)$'s with $\sqrt{INR}$ and $g_{11}(i)$'s with $\sqrt{SNR}$.

Proof. The proof involves expanding the matrix determinant and repeated application of the logarithmic Jensen's gap property. The details are given in Appendix C.9. $\square$

Subsequently, we use the following lemma in bounding $\lim_{n\to\infty} \frac{1}{n}\log\left(\left|\hat{K}_{\mathbf{Y_1}}(n)\right|\right)$.

Lemma 4.4. *If* $A_1 = [|a|]\,, A_2 = \begin{bmatrix} |a| & b \\ b^* & |a| \end{bmatrix}, A_3 = \begin{bmatrix} |a| & b & 0 \\ b^* & |a| & b \\ 0 & b^* & |a| \end{bmatrix}$, *etc. with* $|a|^2 > 4\,|b|^2$*, then*

$$\liminf_{n\to\infty} (1/n) \log(|A_n|) \geq \log(|a|) - 1.$$

Proof. The proof is given in Appendix C.10. $\square$

For the n-phase scheme, the $\left|\hat{K}_{\mathbf{Y_1}}(n)\right|$ matrix has the form A_n, as defined in Lemma 4.4 after identifying $|a| = 1 + INR + SNR$ and $b = \left(\sqrt{SNR} \cdot INR\right) / \left(\sqrt{1 + INR}\right)$. Note that with this choice $|a|^2 > 4\,|b|^2$ holds due to AM-GM (Arithmetic Mean $\geq$ Geometric Mean) inequality. Hence, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log\left(\left|\hat{K}_{\mathbf{Y_1}}(n)\right|\right) \geq \log(1 + INR + SNR) - 1 \tag{4.51}$$

using Lemma 4.4. Also, $K_{\mathbf{Y_1}|\mathbf{X_1}}(n)$ is a diagonal matrix of the form

$$\begin{aligned} K_{\mathbf{Y_1}|\mathbf{X_1}}(n) = \operatorname{diag}\Bigg(& \frac{|g_{21}(n)|^2}{1 + INR} + 1, \frac{|g_{21}(n-1)|^2}{1 + INR} + 1, \ldots \\ & , \frac{|g_{21}(2)|^2}{1 + INR} + 1, |g_{21}(1)|^2 + 1 \Bigg). \end{aligned} \tag{4.52}$$

Hence, using Jensen's inequality

$$\begin{aligned} & \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}\left[\log\left(\left|K_{\mathbf{Y_1}|\mathbf{X_1}}(n)\right|\right)\right] && (4.53) \\ & \leq \limsup_{n\to\infty} \frac{1}{n} \log\left(\left(\frac{INR}{1 + INR} + 1\right)^{n-1} (1 + INR)\right) && (4.54) \\ & = \log\left(INR\,(1 + INR)^{-1} + 1\right) && (4.55) \\ & \leq 1. && (4.56) \end{aligned}$$

Hence, by combining Lemma 4.3, Equation (4.51) and Equation (4.56), we get

$$R_1 \leq \log(1 + INR + SNR) - 3c_{JG} - 2 \tag{4.57}$$

is achievable.

4.6.1.2 Decoding at Rx2

For user 2 we can use a block variant of Schalkwijk-Kailath scheme [SK66] to achieve $R_2 = \mathbb{E}\left[\log^+\left(|g_d|^2 / (1+INR)\right)\right]$. The key idea is that the interference-plus-noise sent in subsequent slots can indeed refine the symbols of the previous slot. The chain of refinement over n phases compensate for the fact that the information symbols are sent only in the first phase. We have

$$Y_2^{(1)N} = g_{22}^{(1)N} X_2^N + g_{12}^{(1)N} X_1^{(1)N} + Z_2^{(1)N} \tag{4.58}$$

and

$$\begin{aligned} Y_2^{(i)N} = {} & g_{22}^{(i)N} \left(\frac{g_{12}^{(i-1)N} X_1^{(i-1)N} + Z_2^{(i-1)N}}{\sqrt{1+INR}} \right) + g_{12}^{(i)N} X_1^{(i)N} \\ & + Z_2^{(i)N} \end{aligned} \tag{4.59}$$

for $i > 1$. Now let

$$\tilde{Y}_2^N = Y_2^{(n)N} + \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^{n} \frac{-g_{22}^{(j)N}}{\sqrt{1+INR}} \right) Y_2^{(i)N}.$$

We have

$$\tilde{Y}_2^N = Y_2^{(n)N} + \sum_{i=1}^{n-1} \left(\prod_{j=i+1}^{n} \frac{-g_{22}^{(j)N}}{\sqrt{1+INR}} \right) Y_2^{(i)N} \tag{4.60}$$

$$\begin{aligned} = {} & g_{22}^{(n)N} \left(\frac{g_{12}^{(n-1)N} X_1^{(n-1)N} + Z_2^{(n-1)N}}{\sqrt{1+INR}} \right) \\ & + g_{12}^{(n)N} X_1^{(n)N} + Z_2^{(n)N} \\ & + \left(-g_{22}^{(n)N} (1+INR)^{-1/2} \right) \\ & \quad \times \left(g_{22}^{(n-1)N} \left(\frac{g_{12}^{(n-2)N} X_1^{(n-2)N} + Z_2^{(n-2)N}}{\sqrt{1+INR}} \right) \right. \\ & \quad \left. + g_{12}^{(n-1)N} X_1^{(n-1)N} + Z_2^{(n-1)N} \right) \\ & + \left(g_{22}^{(n)N} g_{22}^{(n-1)N} (1+INR)^{-1} \right) \end{aligned} \tag{4.61}$$

$$
\begin{aligned}
&\times \left(g_{22}^{(n-2)N} \left(\frac{g_{12}^{(n-3)N} X_1^{(n-3)N} + Z_2^{(n-3)N}}{\sqrt{1+INR}} \right) \right. \\
&\left. + g_{12}^{(n-2)N} X_1^{(n-2)N} + Z_2^{(n-2)N} \right) \\
&+ \cdots \\
&+ \left(\prod_{j=2}^{n} \left(-g_{22}^{(j)N} (1+INR)^{-1/2} \right) \right) \\
&\times \left(g_{22}^{(1)N} X_2^N + g_{21}^{(1)N} X_2^{(1)N} + Z_1^{(1)N} \right) \qquad (4.62) \\
= g_{22}^{(1)N} &\left(\prod_{j=2}^{n} \frac{-g_{22}^{(j)N}}{\sqrt{1+INR}} \right) X_2^N + g_{12}^{(n)N} X_1^{(n)N} \\
&+ Z_2^{(n)N}. \qquad (4.63)
\end{aligned}
$$

due to cross-cancellation. Now Rx2 decodes for its message from $\tilde{Y}_2^N$. Hence, Rx2 can achieve the rate

$$
\begin{aligned}
R_2 \le \liminf_{n\to\infty} \frac{1}{n} \mathbb{E} \Bigg[\log \Bigg(1 + &\left(\prod_{j=2}^{n} \left(|g_{22}(j)|^2 (1+INR)^{-1} \right) \right) \\
&\times \left(\frac{|g_{22}(1)|^2}{1+|g_{12}(n)|^2} \right) \Bigg) \Bigg], \qquad (4.64)
\end{aligned}
$$

where $g_{22}(1), \ldots, g_{22}(n) \sim g_d$ being i.i.d and $g_{12}(n) \sim g_c$. Hence, it follows that

$$
R_2 \le \mathbb{E}\left[\log^+ \left(|g_d|^2 (1+INR)^{-1} \right) \right] \qquad (4.65)
$$

is achievable.

4.6.1.3 Capacity gap

We can obtain the following outer bounds from Theorem 4.11 for the special case of symmetric fading statistics.

$$
\begin{aligned}
R_1, R_2 &\le \mathbb{E}\left[\log \left(|g_d|^2 + |g_c|^2 + 1 \right) \right] \qquad (4.66) \\
R_1 + R_2 &\le \mathbb{E}\left[\log \left(1 + |g_d|^2 \left(1 + |g_c|^2 \right)^{-1} \right) \right]
\end{aligned}
$$

$$+\mathbb{E}\left[\log\left(|g_d|^2+|g_c|^2+2|g_d||g_c|+1\right)\right], \tag{4.67}$$

where Equation (4.66) is obtained from Equation (4.27b) and Equation (4.27d) by setting $\rho = 0$ (note that $\rho = 0$ yields the loosest version of outer bounds in Equation (4.27b) and Equation (4.27d)). Similarly, Equation (4.67) is a looser version of outer bound Equation (4.27e) independent of ρ. The outer bounds reduce to a pentagonal region with two non-trivial corner points (see Figure 4.6). Our n-phase scheme can achieve the two corner points within $2 + 3c_{JG}$ bits per channel use for each user. The proof is using logarithmic Jensen's gap property and is deferred to Appendix C.11.

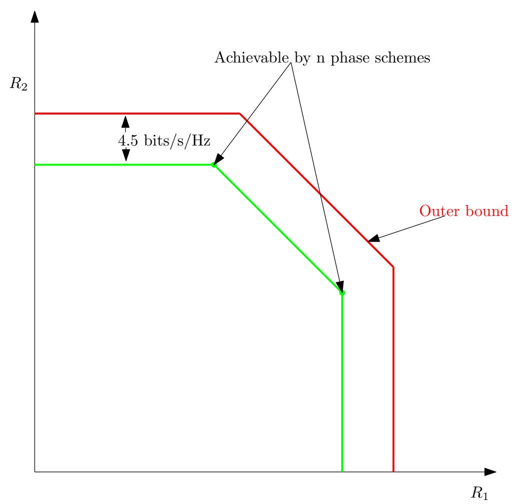


Figure 4.6: Illustration of bounds for capacity region for symmetric FF-IC. The corner points of the outer bound can be approximately achieved by our n-phase schemes. The gap is approximately 4.5 bits per channel use for the Rayleigh fading case.

4.6.2 An auxiliary result: Approximate capacity of 2-tap fast Fading ISI channel

Consider the 2-tap fast fading ISI channel described by

$$Y(l) = g_d(l)X(l) + g_c(l)X(l-1) + Z(l), \tag{4.68}$$

where $g_d \sim \mathcal{CN}(0, SNR)$ and $g_d \sim \mathcal{CN}(0, INR)$ are independent fading known only to the receiver and $Z \sim \mathcal{CN}(0,1)$. Also we assume a power constraint of $\mathbb{E}\left[|X|^2\right] \leq 1$ on the transmit symbols. Our analysis for R_1 can be easily modified to obtain a closed form approximate expression for this channel. This gives rise to the following corollary on the capacity of fading ISI channels.

Corollary 4.15. *The capacity C_{F-ISI} of the 2-tap fast fading ISI channel is bounded by*

$$C_{F-ISI} \leq \log(1 + SNR + INR) + 1$$
$$C_{F-ISI} \geq \log(1 + SNR + INR) - 1 - 3c_{JG},$$

where the channel fading strengths is assumed to have a logarithmic Jensen's gap of c_{JG}.

Proof. The proof is given in Appendix C.12. □

CHAPTER 5

Noncoherent Interference Channel

5.1 Introduction

There has been considerable amount of study on noncoherent wireless channels [MH99, ATS01, ZT02, KK13]. However, most of the progress has been on unicast networks, except the recent work [NYG17] on noncoherent broadcast channel. In this chapter, we consider the noncoherent interference channel with symmetric statistics and demonstrate an achievable gDoF region.

5.1.1 Related work

The noncoherent wireless model for multiple input multiple output (MIMO) channel, was studied by Marzetta and Hochwald [MH99]. In their model, where neither the receiver nor the transmitter knows the fading coefficients and the fading gains remain constant within a block of length T symbol periods. Across the blocks, the fading gains are identically independent distributed (i.i.d.) according to Rayleigh distribution. The capacity behavior at high signal-to-noise ratio (SNR) was studied for the noncoherent MIMO in [ZT02]. Some works have specifically studied the case with $T = 1$ [TE97, ATS01, LM03]. In [ATS01], it was demonstrated that for $T = 1$, the capacity is achieved by a distribution with finite number of mass points, but the number of mass points grew with SNR. The capacity for the $T = 1$ case was shown to behave double logarithimically in [LM03].

There have been other works that studied noncoherent relay channels. The noncoherent single relay network was studied in [KK13], they considered identical link strengths and unit coherence time. They showed that under certain conditions on the fading statistics, the relay does not increase the capacity at high-SNR. In [GY14], similar observations were made for the noncoherent MIMO full-duplex single relay channel with block-fading. They showed that Grassmanian signaling can achieve the degrees of freedom (DoF) without using the relay. Also for certain regimes, decode-and-forward with Grassmanian signaling was shown to approximately achieve the capacity at high-SNR.

The above works considered a DoF framework for the noncoherent model, in the sense that for high-SNR, the link strengths are not significantly different, *i.e.,* the links scale with the same SNR-exponent. The generalized degrees of freedom (gDoF) framework for noncoherent MIMO was considered in Chapter A and it was shown that several insights from the DoF framework may not carry on to the gDoF framework. It was shown that a conventional training scheme is not gDoF optimal and that all antennas may have to be used for achieving the gDoF, even when the coherence time is low, in contrast to the results for the MIMO with i.i.d. links. In Chapter B, the gDoF of the 2-relay diamond network was studied. The training-based schemes were proven to be sub-optimal and a new scheme was proposed, which partially trains the network, performs a scaling and quantize-map-forward operation [OD10, OD13, ADT11] at the relays. These above works focused on unicast networks. Recently an achievability scheme for noncoherent broadcast channel was considered in [NYG17]. They derived an achievable DoF region for noncoherent broadcast channel using statistical channel state information.

In this chapter, we study noncoherent interference channel (IC) with symmetric statistics. This, we believe, is the first work on noncoherent channels in multicast networks. The (coherent) IC is well understood in terms of its capacity [HK81, CMG08, ETW08, ST11] when the channels are known at the receivers and transmitters. The capacity region of the 2-user IC without feedback was characterized in [ETW08], to within 1 bit/s/Hz. In [ST11], a similar result was derived for the IC with feedback, obtaining the capacity region within 2 bits/s/Hz. When the channels are time varying, most of the existing techniques for IC cannot be used without channel state information at transmitter (CSIT). The idea of interference alignment from [CJ08], has been extended to fast fading interference channels (FF-IC) for certain cases, to obtain the DoF region.

The degrees of freedom region for the MIMO FF-IC was studied in [VV12] and their results showed that when *all users have single antenna*, the DoF region is same for the cases of no CSIT, delayed CSIT and instantaneous CSIT. The results from [TMP13] showed that the DoF region for the FF-IC with instantaneous CSIT and no feedback contains the DoF region with output feedback and delayed CSIT. This result changes when one considers more than DoF, for example as in Chapter C, where the approximate capacity region (within a constant additive gap) for FF-IC (with no instantaneous CSIT) was derived. There, we used a rate-splitting scheme based on average interference-to-noise ratio (INR), extending the existing rate-splitting schemes for IC [ETW08, ST11], and proved that this was approximately optimal for FF-IC. This approximate capacity region was derived for FF-IC without feedback and also for the case with feedback; the feedback improves the capacity region for FF-IC, similar to the case for the static IC [ST11].

5.1.2 Contributions

In this chapter we extend the results from Chapter C for FF-IC (where the receivers know the channel, but not the transmitters) to the case when both transmitters and receivers do not know the channel, *i.e.,* the noncoherent IC. We consider the IC with symmetric statistics. We use a noncoherent version of the Han-Kobayashi scheme [HK81], where the transmitters use superposition coding, splitting their messages into common and private parts, and the receivers use joint decoding. We use Gaussian codebooks and use rate-splitting based on average interference-to-noise ratio (INR). We evaluate the achievable gDoF region with this scheme and compare it to a training based scheme. For a 2-user IC, a training based scheme uses at least 2 symbols in every coherence period T, to train the channels. We consider the gDoF of the IC with the rest of the $T-2$ symbols available for communication. We show that our noncoherent scheme outperforms the training-based scheme in gDoF. We also consider the scheme which treats interference as noise (TIN) and observe that TIN has higher gDoF than the noncoherent scheme and the training-based schemes when the INR is low compared to the INR. But in the other regimes, the noncoherent scheme achieves the best gDoF.

We also consider the noncoherent FF-IC with channel state and output feedback. Again we propose a noncoherent scheme that uses Gaussian codebooks and rate-splitting based on average interference-to-noise ratio (INR) similar to Chapter C. We evaluate the gDoF region and compare it with a training based scheme and prove that the noncoherent scheme outperforms the training-based scheme. The noncoherent scheme with feedback increases the gDoF compared to the noncoherent scheme without feedback. Also we observe that with feedback, the performance of our noncoherent scheme is better than TIN schemes for $T \geq 3$, even when the INR is low compared to the SNR.

The chapter is organized as follows. In Section 5.2, we discuss our system model. In Section 5.3, we discuss our results on the FF-IC without feedback and in Section 5.4, we discuss the FF-IC with feedback. Some of the proofs for the analysis is deferred to the appendixes.

5.2 System model

We use the same notations as defined in Section 2.2.1 on page 13. We consider the 2-user noncoherent Gaussian fading IC (Figure 5.1) with coherence time T. We have

$$Y_1 = g_{11}X_1 + g_{21}X_2 + W_1 \tag{5.1}$$

$$Y_2 = g_{12}X_1 + g_{22}X_2 + W_2 \tag{5.2}$$

where the X_i,Y_i, W_i are $1 \times T$ vectors and the links g_{ij} are fading. The realizations of g_{ij} for any fixed (i, j) are i.i.d. across time, and the realizations for different (i, j) are independent. We consider the case with symmetric fading statistics $g_{11} \sim g_{22} \sim \mathcal{CN}(0, \mathsf{SNR})$ and $g_{12} \sim g_{21} \sim \mathcal{CN}(0, \mathsf{INR})$. Neither the receivers nor the transmitters have knowledge of any of the realizations of g_{ij}, but the channel statistics are known to both the receivers and the transmitters.

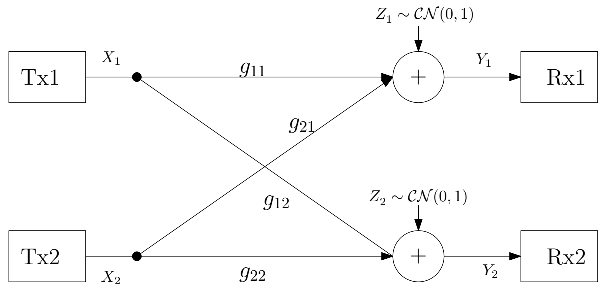


Figure 5.1: The channel model without feedback.

Under the feedback model (Figure 5.2), after each reception, each receiver reliably feeds back the received symbol and the channel states to its corresponding transmitter[1]. We consider the delayed feedback of symbols in blocks of T, however the results that we derive still hold even if the feedback is performed during every symbol period.

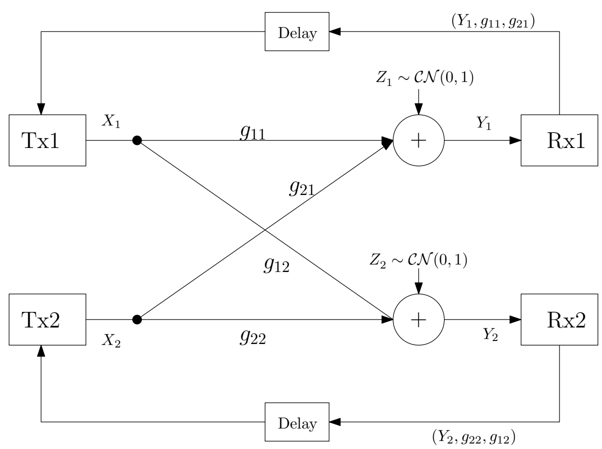


Figure 5.2: The channel model with feedback.

The interference level α is defined as, $\alpha = \log(\mathsf{INR}) / \log(\mathsf{SNR})$. Let $\mathcal{C}(\mathsf{SNR}, \mathsf{INR})$ denote the capacity region. Let $\tilde{\mathcal{D}}$ be a scaled version of $\mathcal{C}(\mathsf{SNR}, \mathsf{INR})$ given by $\tilde{\mathcal{D}}(\mathsf{SNR}, \mathsf{INR}) = \{(R_1/\log(\mathsf{SNR}), R_2/\log(\mathsf{SNR})) : (R_1, R_2) \in \mathcal{C}(\mathsf{SNR}, \mathsf{INR})\}$. Following [ETW08], we define the generalized degrees of freedom region as

$$\mathcal{D}(\alpha) = \lim_{\substack{\mathsf{SNR}, \mathsf{INR} \to \infty \\ \alpha \text{fixed}}} \tilde{\mathcal{D}}(\mathsf{SNR}, \mathsf{INR}).$$

[1]IC with rate limited feedback is considered in [VSA12] where outputs are quantized and fed back. Our schemes can also be extended for such cases.

We also assume $T \geq 2$, since if $T = 1$ the gDoF region of the IC is null following the result for noncoherent MIMO from Chapter 2.

5.3 Noncoherent IC without feedback

Theorem 5.1. *Using a noncoherent rate splitting scheme the gDoF regions given in Table 5.1 are achievable.*

Table 5.1: Achievable gDoF regions for different regimes of α.

$\alpha < 1/2$	$1/2 \leq \alpha \leq 1$	$\alpha \geq 1$
$d_1 \leq \left(1 - \frac{1}{T}\right) - \frac{\alpha}{T}$ $d_2 \leq \left(1 - \frac{1}{T}\right) - \frac{\alpha}{T}$ $d_1 + d_2 \leq 2\left(1 - \frac{1}{T}\right) - 2\alpha$	$d_1 + d_2 \leq \left(2 - \frac{3}{T}\right) - \alpha\left(1 - \frac{1}{T}\right)$ $d_1 + d_2 \leq 2\left(1 - \frac{2}{T}\right)\alpha$ $2d_1 + d_2 \leq \left(2 - \frac{3}{T}\right) - \frac{\alpha}{T}$ $d_1 + 2d_2 \leq \left(2 - \frac{3}{T}\right) - \frac{\alpha}{T}$	$d_1 \leq \left(1 - \frac{2}{T}\right)$ $d_2 \leq \left(1 - \frac{2}{T}\right)$ $d_1 + d_2 \leq \left(1 - \frac{1}{T}\right)\alpha - \frac{1}{T}$

Proof. The proof follows by analyzing a Han-Kobayashi scheme similar to that in Chapter 4 with rate-splitting based on average interference to noise ratio. The details are in Section 5.3.2. $\square$

5.3.1 Discussion

We now compare our achievable gDoF with that of a training-based scheme. The approximate capacity region of coherent fast fading IC is given in Chapter 4. The gDoF for the case which uses 2 symbols for training can be easily obtained from the gDoF region for coherent case with a multiplication factor of $(1 - 2/T)$. Hence, the gDoF regime for a scheme that uses 2 symbols for training is given by

$$d_1, d_2 \leq \left(1 - \frac{2}{T}\right) \tag{5.3a}$$

$$d_1 + d_2 \leq \left(1 - \frac{2}{T}\right)\left(\max\left(1, \alpha\right) + \max\left(1 - \alpha, 0\right)\right) \tag{5.3b}$$

$$d_1 + d_2 \leq 2\left(1 - \frac{2}{T}\right)\max\left(1-\alpha, \alpha\right) \tag{5.3c}$$

$$2d_1 + d_2,\ d_1 + 2d_2 \leq \left(1 - \frac{2}{T}\right)\left(\max\left(1, \alpha\right) + \max\left(1-\alpha, \alpha\right) + \max\left(1-\alpha, 0\right)\right). \tag{5.3d}$$

In Figures 5.3, 5.4, 5.5, the gDoF achievable with our noncoherent scheme is compared with gDoF achievable using the training-based scheme. It can be observed that our noncoherent scheme outperforms the training-based scheme. We also consider the strategy of treating interference-as-noise (TIN) with Gaussian codebooks. Using standard analysis and using Gaussian codebooks, it can be easily shown that the gDoF:

$$d_1, d_2 \leq \left(1 - \frac{1}{T}\right)\left(1-\alpha\right)$$

can be achieved by treating interference as noise. Now, we give the achievable symmetric gDoF for the three strategies that we discussed, with coherence time $T = 5$, in Figure 5.6. It can be calculated from our gDoF regions that treating interference as noise outperforms other strategies when $\alpha < \left(1 - 1/T\right)/\left(2 - 3/T\right)$. Note that for the coherent case, rate-splitting based on INR is only as good as TIN for low INR ($\alpha < 1/2$). For noncoherent case, rate-splitting scheme have lower performance than TIN for low INR, because the uncertainty in the interfering channel together with the uncertainty in the interfering message to be decoded, reduces the amount of the direct message that can be decoded. This reduction is more significant in the noncoherent case (compared to the coherent case) because the uncertainty in the channels does not appear in the coherent case.

Difficulty with Outer Bounds: One trivial outer bound is the coherent outer bound *i.e.,* assuming that the receivers have perfect channel state information. We could also try to derive noncoherent outer bounds following existing techniques. For example, following [ETW08, Theorem 1] and using a genie-aided technique with $S_1 = g_{12}X_1 + Z_2$, and $S_2 = g_{21}X_2 + Z_1$, we could derive an outer bound

$$T\left(R_1 + R_2\right) \leq h\left(Y_1|S_1, \Lambda\right) + h\left(Y_2|S_2, \Lambda\right) - h\left(S_1|X_1, \Lambda\right) - h\left(S_2|X_2, \Lambda\right) \tag{5.4}$$

with input distributions $p\left(\Lambda\right)p\left(X_1|\Lambda\right)p\left(X_2|\Lambda\right)$ with a time-sharing random variable Λ. However, this bound is not better than the coherent outer bound. To understand this, we

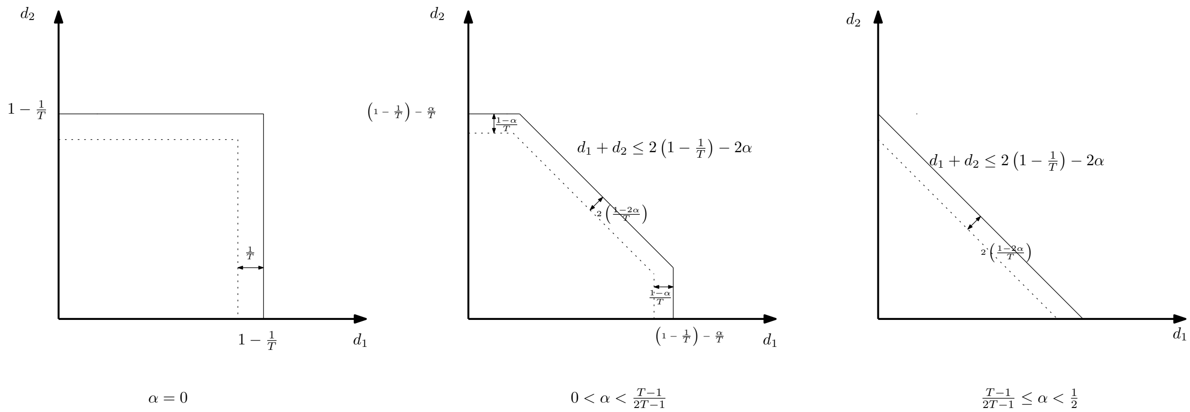


Figure 5.3: gDoF for $\alpha < 1/2, T \geq 2$. The solid line is achievable for a noncoherent scheme and the dotted line is is an outer bound gDoF for a scheme that uses 2 symbols for training.

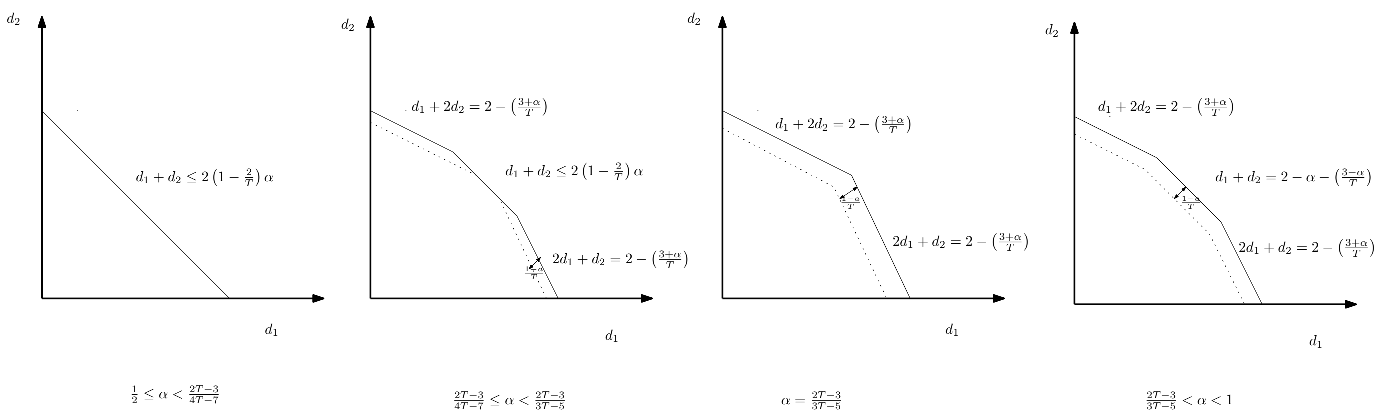


Figure 5.4: gDoF for $1/2 < \alpha < 1$, $T \geq 3$. For $T = 2$ no gDoF is achievable using known schemes. The solid line is achievable for a noncoherent scheme and the dotted line is an outer bound gDoF for a scheme that uses 2 symbols for training.

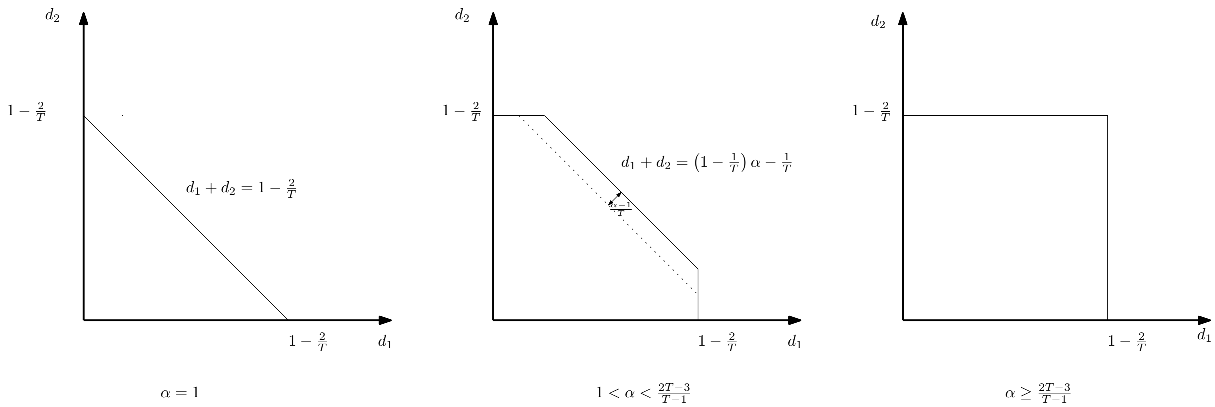


Figure 5.5: gDoF for $1 \leq \alpha, T \geq 3$. For $T = 2$ no gDoF is achievable using known schemes. The solid line is achievable for a noncoherent scheme and the dotted line is is an outer bound gDoF for a scheme that uses 2 symbols for training.

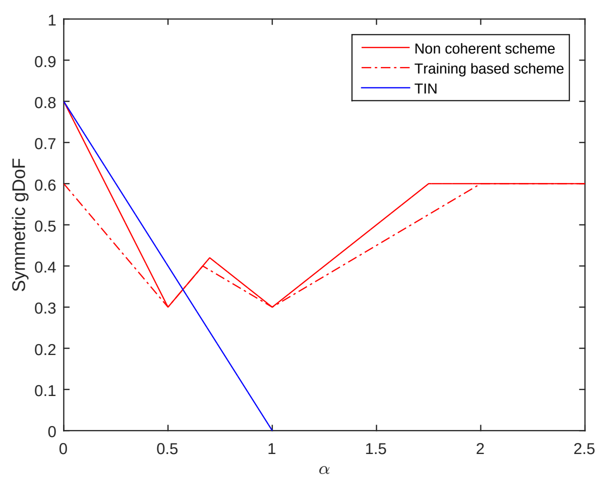


Figure 5.6: Symmetric achievable gDoF for coherence time $T = 5$. Training based scheme uses 2 symbols for training. Treating interference as noise dominates others when $\alpha < (1 - 1/T) / (2 - 3/T)$.

try evaluating (5.4) with X_1, X_2 taken as independent vectors with i.i.d. $\mathcal{CN}(0,1)$ elements. In this case, it can be shown that

$$h(Y_1|S_1) \dot{\geq} \log(1+\mathsf{INR}+\mathsf{SNR}) + (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right),$$

$$h(S_1|X_1) \doteq \log(1+\mathsf{INR}).$$

$$h(Y_1|S_1) - h(S_1|X_1) \dot{\geq} T\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)$$

This means that for $\alpha < 1/2$ for gDoF, the bound (5.4) is looser than the bound $R_1 + R_2 \dot{\leq} 2\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)$, which is same as the coherent outer bound for $\alpha < 1/2$.

5.3.2 Proof of Theorem 5.1

From [CMG08], we obtain that a Han-Kobayashi scheme [HK81] for IC can achieve the following rate region for all $p(U_1)\,p(U_2)\,p(X_1|U_1)\,p(X_2|U_2)$:

$$TR_1 \leq I(X_1;Y_1|U_2) \tag{5.5a}$$

$$TR_2 \leq I(X_2;Y_2|U_1) \tag{5.5b}$$

$$T(R_1+R_2) \leq I(X_2,U_1;Y_2) + I(X_1;Y_1|U_1,U_2) \tag{5.5c}$$

$$T(R_1+R_2) \leq I(X_1,U_2;Y_1) + I(X_2;Y_2|U_1,U_2) \tag{5.5d}$$

$$T(R_1+R_2) \leq I(X_1,U_2;Y_1|U_1) + I(X_2,U_1;Y_2|U_2) \tag{5.5e}$$

$$T(2R_1+R_2) \leq I(X_1,U_2;Y_1) + I(X_1;Y_1|U_1,U_2) + I(X_2,U_1;Y_2|U_2) \tag{5.5f}$$

$$T(R_1+2R_2) \leq I(X_2,U_1;Y_2) + I(X_2;Y_2|U_1,U_2) + I(X_1,U_2;Y_1|U_1). \tag{5.5g}$$

Now similar to that in [ETW08, SKD18, SKD17], we choose U_k as a vector of length T with i.i.d. $\mathcal{CN}(0,\lambda_c)$ elements and X_{pk} as a vector of length T with i.i.d. $\mathcal{CN}(0,\lambda_p)$ elements for $k \in \{1,2\}$ and $X_1 = U_1 + X_{p1}, \quad X_2 = U_2 + X_{p2}$, where $\lambda_c + \lambda_p = 1$ and $\lambda_p = \min(1/\mathsf{INR}, 1)$. For gDoF characterization, we can assume $\mathsf{INR} \geq 1$. If $\mathsf{INR} < 1$, it is equivalent to the case with $\mathsf{INR} = 1$ for gDoF, since both of these cases obtain $\alpha = 0$. Hence, we can have $\lambda_p = 1/\mathsf{INR}$. Here we used the rate splitting using the average interference to noise ratio.

Fact 5.1. *For an exponentially distributed random variable* ξ *and* $a \geq 0$, $b > 0$, $\log(a + b\mu_\xi) - \gamma \log(e) \leq \mathbb{E}[\log(a + b\xi)] \leq \log(a + b\mu_\xi)$.

Proof. This follows due to the results given in Section 4.3 (on page 99). □

We now simplify the region (5.5) for low interference ($\alpha < 1$) regime. We consider the terms in (5.5), one by one.

Claim 5.1. *The term* $I(X_1; Y_1 | U_2)$ *is lower bounded in gDoF by* $(T-1)\log(1 + \mathsf{SNR}) - \log(1 + \mathsf{INR})$.

Proof. See Appendix D.1. □

Claim 5.2. *The term* $I(X_2, U_1; Y_2)$ *is lower bounded in gDoF by* $(T-1)\log(1 + \mathsf{SNR} + \mathsf{INR}) - \log(1 + \mathsf{INR})$.

Proof. We have

$$I(X_2, U_1; Y_2) = h(Y_2) - h(Y_2 | X_2, U_1) \tag{5.6}$$

$$h(Y_2) \dot{\geq} T \log(1 + \mathsf{SNR} + \mathsf{INR}) \tag{5.7}$$

Also from (D.3) for $h(Y_1 | X_1, U_2)$ in Appendix D.1 and using symmetry we can get,

$$h(Y_2 | X_2, U_1) \dot{\leq} \log(1 + \mathsf{SNR} + \mathsf{INR}) + \log(1 + \mathsf{INR}) \tag{5.8}$$

Hence $I(X_2, U_1; Y_2) \dot{\leq} (T-1)\log(1 + \mathsf{SNR} + \mathsf{INR}) - \log(1 + \mathsf{INR})$. □

Claim 5.3. *The term* $I(X_1; Y_1 | U_1, U_2)$ *is lower bounded in gDoF by*

$$(T-2)\log(1 + \mathsf{SNR}/\mathsf{INR}) + \log(1 + \mathsf{SNR}/\mathsf{INR} + \mathsf{INR}) - \log(\mathsf{INR}).$$

Proof. See Appendix D.2. □

Claim 5.4. *The term* $I(X_1, U_2; Y_1 | U_1)$ *is lower bounded in gDoF by* $(T-1)\log(1 + \mathsf{SNR}/\mathsf{INR} + \mathsf{INR}) - \log(\mathsf{INR})$.

Proof. We have

$$\begin{aligned}
I\left(X_1, U_2; Y_1|U_1\right) &= h\left(Y_1|U_1\right) - h\left(Y_1|X_1, U_2, U_1\right) \\
&\geq h\left(Y_1|U_1\right) - h\left(Y_1|X_1, U_2\right) \\
&\dot{\geq} h\left(Y_1|U_1\right) - \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) - \log\left(1 + \mathsf{INR}\right)
\end{aligned}$$

Where the last step is using (D.3) for $h\left(Y_1 \middle| X_1, U_2\right)$ in Appendix D.1. Now

$$\begin{aligned}
h\left(Y_1|U_1\right) &= h\left(g_{11}X_1 + g_{21}X_2 + Z_1 \middle| U_1\right) \\
&= \sum_i h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i} \middle| \left\{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\right\}_{j=1}^{i-1}, U_1\right) \\
&\overset{(i)}{\dot{\geq}} h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11} \middle| U_1, X_{11}, X_{21}\right) + \\
&\qquad + \sum_{i=2}^{T} h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i} \middle| U_{1i}, g_{21}, g_{11}\right) \\
&\dot{\geq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) \\
&\qquad + (T-1)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) \qquad (5.9)
\end{aligned}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\right) - \left(U_{1i}, g_{21}, g_{11}\right) - \left(\left\{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\right\}_{j=1}^{i-1}, U_1\right)$. Hence

$$I\left(X_1, U_2; Y_1|U_1\right) \dot{\geq} (T-1)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) - \log\left(\mathsf{INR}\right).$$

□

We collect the results from the previous four claims in Table 5.2.

Table 5.2: gDoF inner bounds for the terms in achievability region

Term	Inner bound in gDoF
$I\left(X_1, U_2; Y_1\|U_1\right)$	$(T-1)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) - \log\left(\mathsf{INR}\right)$
$I\left(X_1; Y_1\|U_1, U_2\right)$	$(T-2)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) + \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) - \log\left(\mathsf{INR}\right)$
$I\left(X_2, U_1; Y_2\right)$	$(T-1)\log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) - \log\left(1 + \mathsf{INR}\right)$
$I\left(X_1; Y_1\|U_2\right)$	$(T-1)\log\left(1 + \mathsf{SNR}\right) - \log\left(1 + \mathsf{INR}\right)$

Substituting the inner bounds into the achievability region (5.5), we get the following achievability region in gDoF:

$$TR_1 \dot{\leq} (T-1)\log(1+\mathsf{SNR}) - \log(1+\mathsf{INR}) \tag{5.10a}$$

$$TR_2 \dot{\leq} (T-1)\log(1+\mathsf{SNR}) - \log(1+\mathsf{INR}) \tag{5.10b}$$

$$\begin{aligned} T(R_1+R_2) \dot{\leq}\ & (T-1)\log(1+\mathsf{SNR}+\mathsf{INR}) - \log(1+\mathsf{INR}) \\ & + (T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) + \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \end{aligned} \tag{5.10c}$$

$$\begin{aligned} T(R_1+R_2) \dot{\leq}\ & (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \\ & + (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \end{aligned} \tag{5.10d}$$

$$\begin{aligned} T(2R_1+R_2) \dot{\leq}\ & (T-1)\log(1+\mathsf{SNR}+\mathsf{INR}) - \log(1+\mathsf{INR}) \\ & + (T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) + \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \\ & + (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \end{aligned} \tag{5.10e}$$

$$\begin{aligned} T(R_1+2R_2) \dot{\leq}\ & (T-1)\log(1+\mathsf{SNR}+\mathsf{INR}) - \log(1+\mathsf{INR}) \\ & + (T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right) + \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}) \\ & + (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right) - \log(\mathsf{INR}). \end{aligned} \tag{5.10f}$$

Hence we get the following gDoF region:

$$d_1 \leq \left(1-\frac{1}{T}\right) - \frac{\alpha}{T} \tag{5.11a}$$

$$d_2 \leq \left(1-\frac{1}{T}\right) - \frac{\alpha}{T} \tag{5.11b}$$

$$d_1+d_2 \leq \left(1-\frac{1}{T}\right) - \frac{\alpha}{T} + \left(1-\frac{2}{T}\right)(1-\alpha) + \frac{1}{T}\max(1-\alpha,\alpha) - \frac{\alpha}{T} \tag{5.11c}$$

$$d_1+d_2 \leq 2\left(1-\frac{1}{T}\right)\max(1-\alpha,\alpha) - \frac{2\alpha}{T} \tag{5.11d}$$

$$2d_1+d_2 \leq \left(1-\frac{1}{T}\right) - \frac{\alpha}{T} + \left(1-\frac{2}{T}\right)(1-\alpha) + \max(1-\alpha,\alpha) - \frac{2\alpha}{T} \tag{5.11e}$$

$$d_1+2d_2 \leq \left(1-\frac{1}{T}\right) - \frac{\alpha}{T} + \left(1-\frac{2}{T}\right)(1-\alpha) + \max(1-\alpha,\alpha) - \frac{2\alpha}{T}. \tag{5.11f}$$

It can be verified that this gDoF region can be simplified for different regimes of $\alpha < 1$ as given in Table 5.1. Now we consider the regime $\alpha > 1$ and and evaluate the gDoF region.

5.3.2.1 High Interference Case ($\alpha > 1$)

Claim 5.5. *The term $I\left(X_1; Y_1 | U_2\right)$ is lower bounded in gDoF by $(T-2)\log\left(1+\mathsf{SNR}\right)$.*

Proof. See Appendix D.3. □

Claim 5.6. *The term $I\left(X_2, U_1; Y_2\right)$ is lower bounded in gDoF by $(T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\mathsf{SNR}\right)$.*

Proof. We have

$$\begin{aligned} I\left(X_2, U_1; Y_2\right) &= h\left(Y_2\right) - h\left(Y_2 | X_2, U_1\right), \\ h\left(Y_2\right) &\dot{\geq} T\log\left(1+\mathsf{SNR}+\mathsf{INR}\right), \\ h\left(Y_2 | X_2, U_1\right) &\overset{(i)}{\dot{\geq}} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\mathsf{SNR}\right), \end{aligned}$$

where (i) is using (D.45) from Appendix D.3 in the proof of Claim 5.5. Hence $I\left(X_2, U_1; Y_2\right) \dot{\leq} (T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\mathsf{SNR}\right)$ follows. □

Claim 5.7. *The term $I\left(X_1, U_2; Y_1 | U_1\right)$ is lower bounded in gDoF by $(T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right)-\log\left(\mathsf{SNR}\right)$.*

Proof. We have

$$\begin{aligned} I\left(X_1, U_2; Y_1 | U_1\right) &= h\left(Y_1 | U_1\right) - h\left(Y_1 | X_1, U_2\right) \\ &\overset{(i)}{\dot{\geq}} h\left(Y_1 | U_1\right) - \log\left(1+\mathsf{SNR}+\mathsf{INR}\right) - \log\left(\mathsf{SNR}\right), \\ h\left(Y_1 | U_1\right) &\overset{(ii)}{\dot{\geq}} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right) + (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right), \end{aligned}$$

where (i) was using (D.45) for $h\left(Y_1 | X_1, U_2\right)$ and (ii) was from (5.9). Hence the desired result follows. □

We collect the results from the previous three claims and a trivial bound for $I\left(X_1;Y_1|U_1,U_2\right)$ in Table 5.3.

Table 5.3: gDoF inner bounds for the terms in achievability region

Term	Lower bound in gDoF
$I\left(X_1,U_2;Y_1\|U_1\right)$	$(T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right)-\log\left(\mathsf{SNR}\right)$
$I\left(X_1;Y_1\|U_1,U_2\right)$	0
$I\left(X_2,U_1;Y_2\right)$	$(T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\mathsf{SNR}\right)$
$I\left(X_1;Y_1\|U_2\right)$	$(T-2)\log\left(1+\mathsf{SNR}\right)$

Substituting the inner bounds into the achievability region (5.5), we get the following achievability region in gDoF:

$$TR_1, TR_2 \dot{\leq} (T-2)\log\left(1+\mathsf{SNR}\right) \tag{5.12a}$$

$$T\left(R_1+R_2\right) \dot{\leq} (T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\mathsf{SNR}\right) \tag{5.12b}$$

$$\begin{aligned} T\left(R_1+R_2\right) \dot{\leq}\ & (T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right)-\log\left(\mathsf{SNR}\right) \\ & +(T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right)-\log\left(\mathsf{SNR}\right) \end{aligned} \tag{5.12c}$$

$$\begin{aligned} T\left(2R_1+R_2\right), T\left(R_1+2R_2\right) \dot{\leq}\ & (T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\mathsf{SNR}\right) \\ & +(T-1)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\mathsf{INR}\right)-\log\left(\mathsf{SNR}\right). \end{aligned} \tag{5.12d}$$

It can be verified that the above region can be reduced to the gDoF region in third column of Table 5.1 for $\alpha > 1$.

5.4 Noncoherent scheme for feedback case

Theorem 5.2. *For a noncoherent IC with feedback, the gDoF region given in Table 5.4 is achievable:*

Table 5.4: Achievable gDoF regions for noncoherent IC with feedback.

$\alpha < 1/2$	$1/2 \leq \alpha \leq 1$	$\alpha \geq 1$
$d_2, d_2 \leq \left(1 - \frac{1}{T}\right) - \frac{2\alpha}{T}$ $d_1 + d_2 \leq 2\left(1 - \frac{1}{T}\right) - \alpha\left(1 + \frac{1}{T}\right)$	$d_2, d_2 \leq \left(1 - \frac{2}{T}\right)$ $d_1 + d_2 \leq \left(2 - \frac{3}{T}\right) - \alpha\left(1 - \frac{1}{T}\right)$	$d_1 + d_2 \leq \left(1 - \frac{1}{T}\right)\alpha - \frac{1}{T}$

Proof. This is obtained using the block Markov scheme of [ST11, Lemma 1] for the noncoherent case. We again use a rate-splitting based on average interference to noise ratio in this case. The proof is given in Section 5.4.2. □

5.4.1 Discussion

We now compare our achievable gDoF with that of a training-based scheme. The approximate capacity region of coherent fast fading IC with feedback is given in Chapter 4. The gDoF for the case which uses 2 symbols for training can be easily obtained from the gDoF region for the coherent case with a multiplication factor of $(1 - 2/T)$. Hence, the gDoF regime for a scheme that uses 2 symbols for training is given by:

$$d_1, d_2 \leq \left(1 - \frac{2}{T}\right) \max(1, \alpha) \tag{5.13a}$$

$$d_1 + d_2 \leq \left(1 - \frac{2}{T}\right) \left(\max(1, \alpha) + \max(1 - \alpha, 0)\right). \tag{5.13b}$$

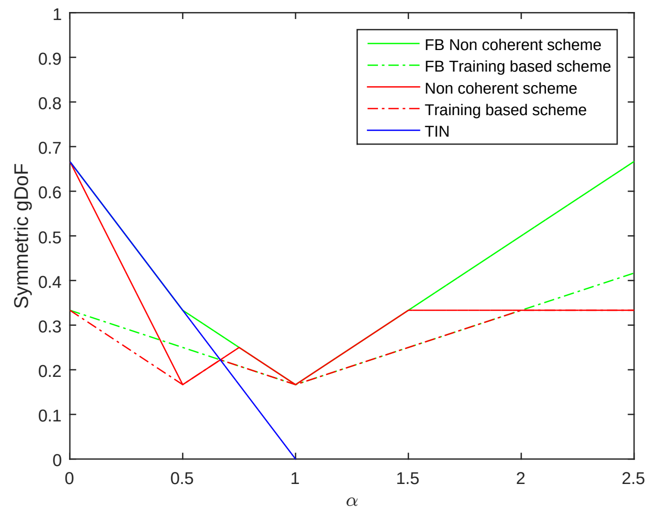


Figure 5.7: Symmetric achievable gDoF for coherence time $T = 3$: feedback and non feedback cases

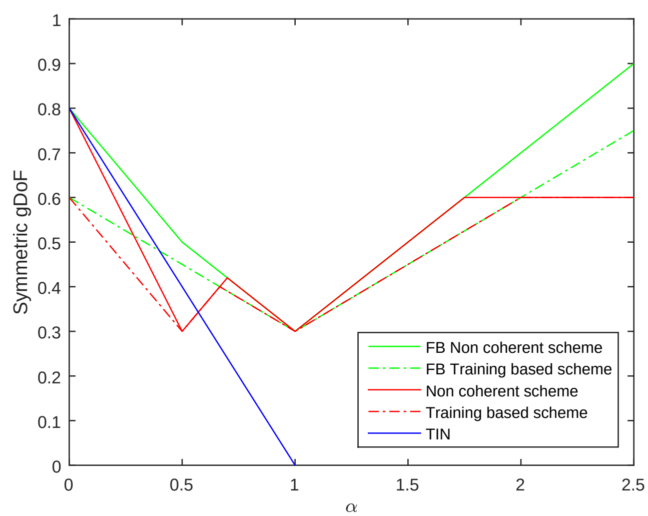


Figure 5.8: Symmetric achievable gDoF for coherence time $T = 5$: feedback and non feedback cases

We give the achievable symmetric gDoF for the noncoherent rate-splitting scheme, training based scheme and treating interference as noise (TIN) scheme, with coherence time $T = 3$, in Figure 5.7 and with coherence time $T = 5$, in Figure 5.8. It can be calculated from Table 5.4 and (5.13) that treating interference as noise (TIN) outperforms our noncoherent strategy with feedback, when $T = 2$ and $\alpha < 1$. Our strategy in the presence of feedback is as good as TIN for or outperforms TIN when $T \geq 3$. The noncoherent rate-splitting scheme attempts to decode part of the interfering message at the transmitter, and use it in subsequent transmissions. The amount of rate that can be decoded increases with T, when $T = 2$ the advantage gained by decoding at the transmitter is low. For low INR, the uncertainty in the interfering channel together with the uncertainty of the interfering message to be decoded at the receiver reduces the amount of direct message that can be decoded in the rate splitting scheme. The advantage gained by decoding at the transmitter outweighs this loss when $T \geq 3$.

5.4.2 Proof of Theorem 5.2

Using the block Markov scheme of [ST11, Lemma 1], we obtain the achievability of the rate pairs (R_1, R_2) satisfying

$$TR_1 \leq I\left(U, U_2, X_1; Y_1\right) \tag{5.14a}$$

$$TR_1 \leq I\left(U_1; Y_2 | U, X_2\right) + I\left(X_1; Y_1 | U_1, U_2, U\right) \tag{5.14b}$$

$$TR_2 \leq I\left(U, U_1, X_2; Y_2\right) \tag{5.14c}$$

$$TR_2 \leq I\left(U_2; Y_1 | U, X_1\right) + I\left(X_2; Y_2 | U_1, U_2, U\right) \tag{5.14d}$$

$$T\left(R_1 + R_2\right) \leq I\left(X_1; Y_1 | U_1, U_2, U\right) + I\left(U, U_1, X_2; Y_2\right) \tag{5.14e}$$

$$T\left(R_1 + R_2\right) \leq I\left(X_2; Y_2 | U_1, U_2, U\right) + I\left(U, U_2, X_1; Y_1\right) \tag{5.14f}$$

for all $p\left(U\right) p\left(U_1|U\right) p\left(U_2|U\right) p\left(X_1|U_1, U\right) p\left(X_2|U_2, U\right)$. We choose $U = 0$, U_k as a vector of length T with i.i.d. $\mathcal{CN}\left(0, \lambda_c\right)$ elements, X_{pk} as a vector of length T with i.i.d. $\mathcal{CN}\left(0, \lambda_p\right)$ elements for $k \in \{1, 2\}$, $X_1 = U_1 + X_{p1}$, $X_2 = U_2 + X_{p2}$ where $\lambda_c + \lambda_p = 1$ and $\lambda_p = \min\left(1/\mathsf{INR}, 1\right)$ similar to [ST11, SKD18]. The region (5.14) following [ST11, Lemma 1] is still valid with $U = 0$. For gDoF characterization, we can assume $\mathsf{INR} \geq 1$. Hence we have $\lambda_p = 1/\mathsf{INR}$. Now we analyze the terms in (5.14) for obtaining an achievable gDoF region.

Claim 5.8. *The term $I\left(U, U_2, X_1; Y_1\right)$ is lower bounded in gDoF by $(T-1)\log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) - \log\left(1 + \min\left(\mathsf{SNR}, \mathsf{INR}\right)\right)$.*

Proof. We have

$$\begin{aligned}
h\left(Y_1\right) &\geq T\mathbb{E}\left[\log\left(\left|g_{11}\right|^2+\left|g_{21}\right|^2+1\right)\right] \\
&\doteq T\log\left(1+\mathsf{SNR}+\mathsf{INR}\right), && (5.15) \\
h\left(Y_1\middle|\, U,U_2,X_1\right) &= h\left(g_{11}X_1+g_{21}X_2+Z_1\middle|\, U,U_2,X_1\right) \\
&\overset{(i)}{\dot{\leq}} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\mathsf{INR}\right), && (5.16) \\
h\left(Y_1\middle|\, U,U_2,X_1\right) &\overset{(ii)}{\dot{\leq}} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\mathsf{SNR}\right), && (5.17)
\end{aligned}$$

where (i) is using $U=0$ and (D.3) on page 225 in the proof of Claim 5.1. The step (ii) is using $U=0$ and (D.45) on page 230 in the proof of Claim 5.5. Hence using the above two equations, we get

$$h\left(Y_1\middle|\, U,U_2,X_1\right) \dot{\leq} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right). \qquad (5.18)$$

Hence the desired result follows.

□

Claim 5.9. *The term* $I\left(U_2;Y_1|U,X_1\right)$ *is lower bounded in gDoF by* $(T-1)\log\left(1+\mathsf{INR}\right)-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$.

Proof. We have

$$\begin{aligned}
h\left(Y_1\middle|\, U,X_1\right) &= h\left(g_{11}X_1+g_{21}X_2+Z_1\middle|\, U,X_1\right) && (5.19) \\
&= \sum_i h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\,\left\{g_{11}X_{1j}+g_{21}X_{2j}+Z_{1j}\right\}_{j=1}^{i-1},U,X_1\right) && (5.20) \\
&\overset{(i)}{\dot{\geq}} h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle|\, X_{21},U,X_1\right)+ \\
&\qquad \sum_{i=2}^{T} h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\, U,X_1,g_{21},g_{11}\right) && (5.21) \\
&\overset{(ii)}{\dot{\geq}} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+(T-1)\log\left(1+\mathsf{INR}\right), && (5.22)
\end{aligned}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\right)$ – $\left(U,X_1,g_{21},g_{11}\right)$ – $\left(\left\{g_{11}X_{1j}+g_{21}X_{2j}+Z_{1j}\right\}_{j=1}^{i-1},U,X_1\right)$

and (ii) is using Gaussianity for terms $h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle| X_{21},U,X_1\right)$ and $h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle| U,X_1,g_{21},g_{11}\right)$. Also

$$h\left(Y_1|\,U,U_2,X_1\right)\dot{\leq}\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$$

from the proof of previous claim. Hence the desired result follows. □

Claim 5.10. *The term $I\left(X_1;Y_1|U_1,U_2,U\right)$ is lower bounded in gDoF by* $\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$.

Proof. The proof is given in Appendix D.4. □

We collect the inner bounds for terms in the achievability region in Table 5.5.

Table 5.5: gDoF inner bounds for the terms in achievability region

Term	Lower bound in gDoF
$I\left(U,U_2,X_1;Y_1\right)$	$(T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$
$I\left(U_2;Y_1|U,X_1\right)$	$(T-1)\log\left(1+\mathsf{INR}\right)-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$
$I\left(X_1;Y_1|U_1,U_2,U\right)$	$\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)$ $-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)$

Using the above results in (5.14) we have the gDoF region:

$$TR_1,TR_2\dot{\leq}(T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right) \tag{5.23a}$$

$$\begin{aligned}TR_1,TR_2\dot{\leq}\,&(T-1)\log\left(1+\mathsf{INR}\right)-2\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)\\&+\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right)\end{aligned} \tag{5.23b}$$

$$\begin{aligned}T\left(R_1+R_2\right)\dot{\leq}\,&(T-1)\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)-2\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)\\&+\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right).\end{aligned} \tag{5.23c}$$

It can be verified that the above region can be reduced to the gDoF region in Table 5.4.

CHAPTER 6

Conclusions and Future work

6.1 Noncoherent MIMO

In Chapter 2, we considered the noncoherent MIMO channel with link strengths scaled with different exponents of SNR. Under this model, we derived a structure for the capacity achieving input distribution. We showed that for $T = 1$, the gDoF is zero for MIMO of any size. Also for the SIMO and the MISO channels, we proved that selecting the best antenna can achieve the gDoF. We derived the gDoF for 2×2 symmetric MIMO with two different exponents in the direct and cross links and showed that both the antennas are always needed to achieve the gDoF. Also training-based schemes were shown to be suboptimal for the 2×2 symmetric MIMO. We extended this observation to $M \times M$ symmetric MIMO with two different exponents in the direct and cross links; we demonstrated a strategy that could achieve larger gDoF than training-based schemes. A possible direction for future work would be to try to derive the gDoF of $M \times M$ symmetric MIMO. Then to look into the case of arbitrary size MIMO with different exponents. The outer bounds for larger MIMO seems to be a challenge at the moment, our outer bounds for 2×2 MIMO illustrates some of the difficulties: we used a Gram Schmidt process for LQ decomposition of matrices and developed new lemmas to bound the terms in the mutual information expression. The same methods do not seem to be directly applicable to larger MIMO.

6.2 Noncoherent Diamond Network

In Chapter 3, we characterized the gDoF of the diamond network with 2 relays, when the links are scaling differently with SNR. For some regimes, a simple decode-and-forward scheme was sufficient to achieve the gDoF and a conventional form of the cut-set outer bound could be used. There were other regimes, where relay selection or training-based schemes would not achieve the conventional cut-set bound gDoF. For these cases, we derived a new outer bound beginning with a modification of the conventional cut-set outer bound. In order to analyze this optimization problem in the outer bound, we then loosened the terms in this outer bound and discretized the terms without losing gDoF. Then we proved that the optimal distribution for the outer bound can have just two mass points without losing gDoF. This distribution could be explicitly obtained. To obtain the inner bound, we used the structure of the solution of the outer bound. The inner bound used a time-sharing random variable with two mass points. This design mimics the gDoF-optimal distribution for the outer bound which had two mass points. In our scheme, the channels from the source to the relays were trained using a single symbol in every block of length T. The relays scale the received data symbols using the channel estimate, and then perform a quantize-map-forward (QMF) operation on the scaled symbols: this we called train-scale QMF (TS-QMF) scheme. We do not use training from the relays to the destination, as seen in the TS-QMF scheme, which is shown to be gDoF-optimal. We show that if training is to be done on all links of the channel, then the gDoF cannot always be achieved.

One of the future directions of study include noncoherent n-relay diamond networks. Our achievability scheme for the noncoherent 2-relay diamond network can be extended to the n-relay case. However, the outer bounds are still an open problem. The more general open problem is the capacity of general noncoherent networks.

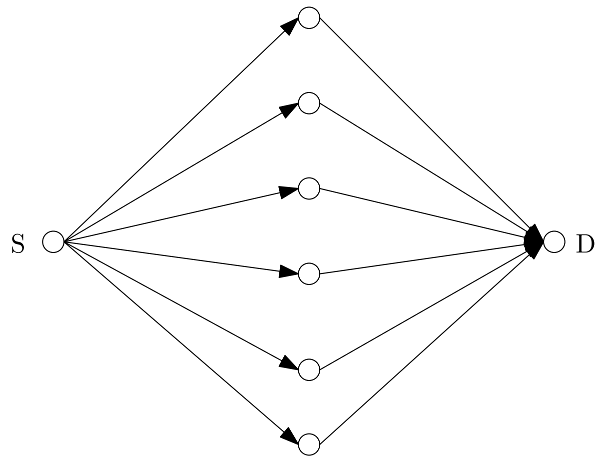


Figure 6.1: n-relay diamond network.

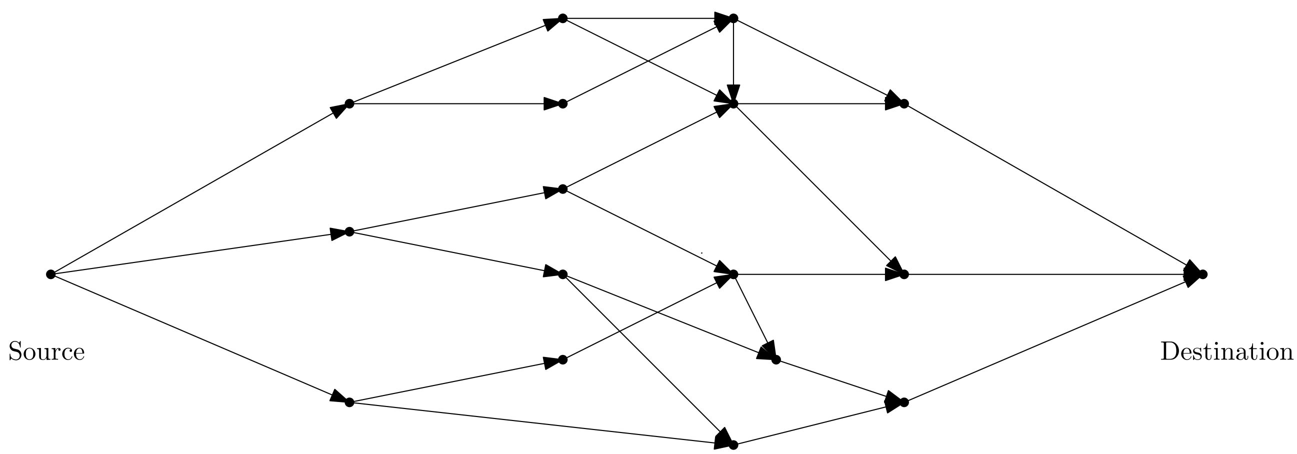


Figure 6.2: A general wireless network with single source and destination.

6.3 Fast Fading ICs

In Chapter 4, we introduced the notion of logarithmic Jensen's gap and demonstrated that it can be used to obtain approximate capacity region for FF-ICs. We proved that the rate-splitting schemes for ICs [ETW08, CMG08, ST11], when extended to the fast fading case give capacity gap as a function of the logarithmic Jensen's gap. Our analysis of logarithmic Jensen's gap for fading models like Rayleigh fading show that rate-splitting is approximately optimal for such cases. We then developed a scheme for symmetric FF-ICs, which can be implemented using point-to-point codes and can approximately achieve the capacity region. An important direction to study will be to see if similar schemes with point-to-point codes can be extended to general FF-ICs. Also our schemes are not approximately optimal for bursty IC [WSD13] since it does not have finite logarithmic Jensen's gap, it would be interesting to study if the schemes can be extended to bursty IC and then to any arbitrary fading distribution. Extension to FF-ICs with more than 2 users seems difficult, since there are no approximate (within constant additive gap) capacity results known even for 3-user IC with fixed channels.

6.4 Noncoherent IC

In Chapter 5, we studied the noncoherent IC with symmetric channel statistics. We proposed an achievability scheme based on noncoherent rate-splitting using the channel statistics. We derived the achievable gDoF using this scheme. We demonstrated that our scheme achieves higher gDoF than a scheme which trains the links of the IC. We also studied a noncoherent rate-splitting scheme for IC with feedback and proved that our scheme achieves higher gDoF than a training-based scheme. For low INR and when there is no feedback, treating interference as noise is better than noncoherent rate-splitting. A simple outer bound is the coherent outer bound (assuming channel state information at receiver). The noncoherent outer bounds that we derived using existing techniques were not better than the coherent outer bound. Hence a possible direction for further studies is to explore new techniques to derive better outer bounds than the coherent outer bound.

6.5 Backscatter communication systems

Another direction of study is to study backscatter communication systems in a noncoherent setting. Backscatter communication systems typically use a Reader and a radio frequency identification (RFID) tag [Dob12]. Reader transmits a radio frequency (RF) signal; the RFID tag adapts the level of its antenna impedance to vary the reflection coefficient and transmits data via reflecting and modulating the incident signal back to Reader [XYV14, BR14]. We have some preliminary results on backscatter systems with intersymbol-interference. We

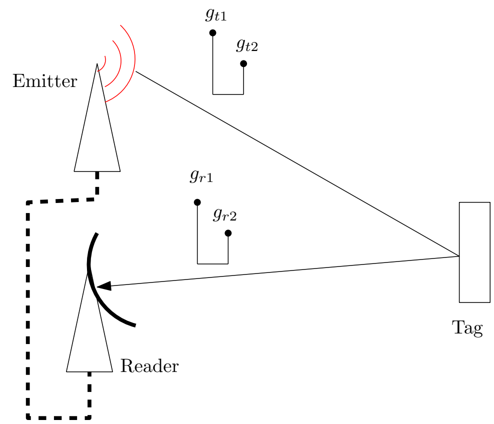


Figure 6.3: Backscatter system with 2-tap ISI and collaboration between Emitter and Reader

demonstrate that instead of using a constant carrier sequence, we can optimize the sequence to obtain larger rates or smaller bit error probability. We optimize the sequence based on the effective channel G in the system. The original optimization problem is numerically hard, hence we have two approximate problems based on $\det\left(G^{\dagger}G+I\right)$ and $\det\left(G^{\dagger}G\right)$. Figure 6.4 illustrates our results for optimizing mutual information rates. Figure 6.5 illustrates our results with a given channel code and using our optimization technique to reduce the bit error rate. Figure 6.6 is similar, but includes estimation errors in obtaining the channel G, which arises due to the noise involved during training. The details of our scheme are given in Appendix E.1. The current results involve training the channel states. The noncoherent version of the problem will be to consider whether a noncoherent scheme can be designed to outperform the training-based schemes.

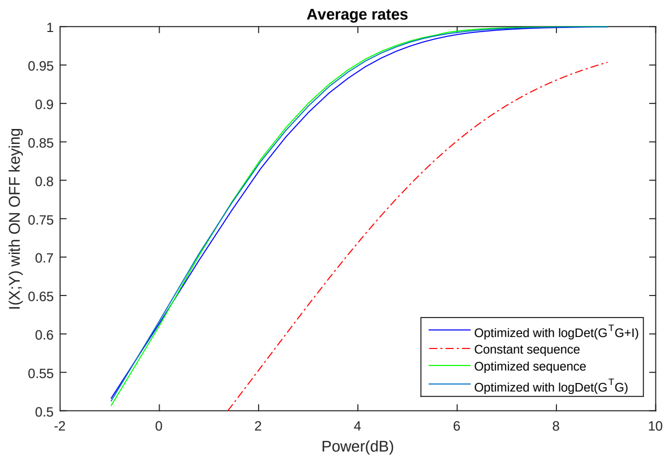


Figure 6.4: Rate achieved with different optimization techniques. We have about 28% gain in the mutual information rate at power 4 dB by optimizing the carrier.

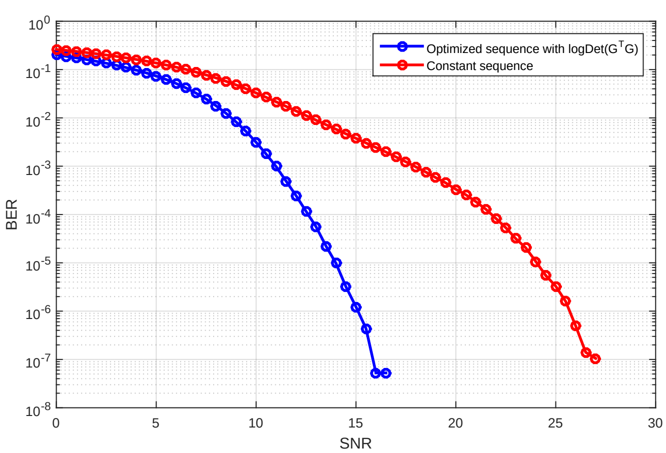


Figure 6.5: BER comparison with ON-OFF keying

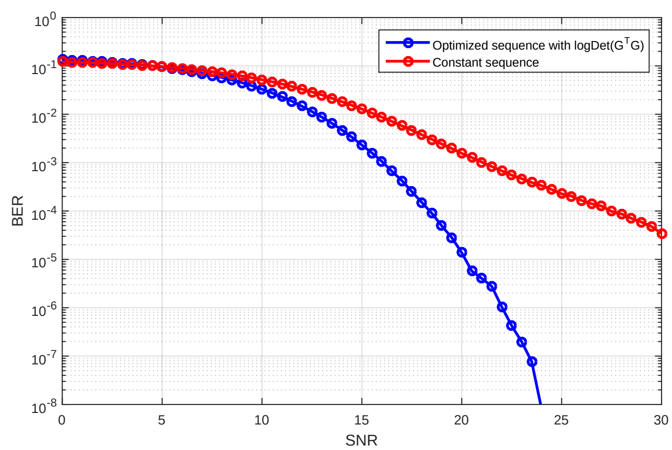


Figure 6.6: BER comparison with ON-OFF keying and noise from channel training.

APPENDIX A

Proofs for Chapter 2

A.1 Proof of Lemma 3.2

Here we derive the formula for calculating $h([\xi_1, \xi_2, \ldots, \xi_n] Q)$ with $[\xi_1, \xi_2, \ldots, \xi_n]$ being an arbitrary complex random vector and Q being an $n \times n$ isotropically distributed unitary matrix independent of ξ_i. We do this by noting that in radial coordinates, the distribution of $[\xi_1, \xi_2, \ldots, \xi_n] Q$ will be dependent only on the radius. Let

$$V = [\xi_1, \xi_2, \ldots, \xi_n] Q.$$

Now for any fixed unitary Q', V and VQ' have the same distribution due to the property of isotropic distribution. Hence for any $v_1, v_2 \in \mathbb{C}^n$ if $\|v_1\| = \|v_2\|$ then

$$p_v(v_1) = p_v(v_2), \tag{A.1}$$

since there exists a unitary matrix Q'' such that $v_1 Q'' = v_2$. One such Q'' can be obtained using Householder transformation. Now the probability distribution can be viewed in $\mathbb{R}^{2n}$ and we use $2n$ dimensional vector U. Let

$$\Upsilon = \sum |\xi_i|^2. \tag{A.2}$$

Let $\left(r, \overline{\theta}\right)$ be the radial coordinates, $\left(t, \overline{\theta}\right)$ be similar coordinates but with $t = r^2$. Let $p_{u,t}\left(t, \overline{\theta}\right) = p_u\left(u\left(t, \overline{\theta}\right)\right)$ be obtained from $p_u(u)$ by expressing u in $\left(t, \overline{\theta}\right)$ coordinates. Similarly $p_{u,r}\left(r, \overline{\theta}\right) = p_u\left(u\left(r, \overline{\theta}\right)\right)$.

The $2n-1$ dimensional surface area (embedded in $2n$ dimensional Euclidean) is $\left(\frac{2\pi^n}{\Gamma(n)}\right) r^{2n-1}$. Hence

$$\left(\frac{2\pi^n}{\Gamma(n)}\right) p_{u,r}\left(r, \overline{\theta}\right) r^{2n-1} dr$$

is the probability that $|U| \in [r, r+dr]$. Hence $\left(\frac{\pi^n}{\Gamma(n)}\right) p_{u,t}\left(t,\overline{\theta}\right) t^{n-1} dt$ is the probability that $\Upsilon = \|U\|^2 \in [t, t+dt]$. Hence

$$\left(\frac{\pi^n}{\Gamma(n)}\right) p_{u,t}\left(t,\overline{\theta}\right) t^{n-1} = p_{\Upsilon}(t) \tag{A.3}$$

$$p_{u,t}\left(t,\overline{\theta}\right) = p_{\Upsilon}(t) \frac{1}{t^{n-1}\left(\frac{\pi^n}{\Gamma(n)}\right)}. \tag{A.4}$$

Now

$$\begin{aligned}
h(U) &= -\int p_u(u) \log\left(p_u(u)\right) du && \text{(A.5)}\\
&\overset{(i)}{=} -\int p_u\left(u\left(r,\overline{\theta}\right)\right) \log\left(p_u\left(u\left(r,\overline{\theta}\right)\right)\right) \left(\frac{2\pi^n}{\Gamma(n)}\right) r^{2n-1} dr && \text{(A.6)}\\
&\overset{(ii)}{=} -\int p_u\left(u\left(t,\overline{\theta}\right)\right) \log\left(p_u\left(u\left(t,\overline{\theta}\right)\right)\right) \left(\frac{\pi^n}{\Gamma(n)}\right) t^{n-1} dt && \text{(A.7)}\\
&\overset{(iii)}{=} -\int p_{\Upsilon}(t) \log\left(p_{\Upsilon}(t) \frac{1}{t^{n-1}\left(\frac{\pi^n}{\Gamma(n)}\right)}\right) dt && \text{(A.8)}\\
&= -\int p_{\Upsilon}(t) \log\left(p_{\Upsilon}(t)\right) dt + \log\left(\frac{\pi^n}{\Gamma(n)}\right) + (n-1)\int p_{\Upsilon}(t) \log(t)\, dt && \text{(A.9)}\\
&= h(\Upsilon) + (n-1)\,\mathbb{E}\left[\log(\Upsilon)\right] + \log\left(\frac{\pi^n}{\Gamma(n)}\right) && \text{(A.10)}\\
&= h\left(\sum |\xi_i|^2\right) + (n-1)\,\mathbb{E}\left[\log\left(\sum |\xi_i|^2\right)\right] + \log\left(\frac{\pi^n}{\Gamma(n)}\right), && \text{(A.11)}
\end{aligned}$$

where (i) is by change of variables to $\left(r,\overline{\theta}\right)$ and integrating over $\overline{\theta}$ and noting that $p_u\left(u\left(r,\overline{\theta}\right)\right)$ is independent of $\overline{\theta}$, (ii) is by change of variables to $\left(t,\overline{\theta}\right)$, (iii) is using (A.4).

A.2 Proof of Lemma 3.1

Here we consider the optimization problem $\mathcal{P}_1$ from (2.73) on page 37 and show that its objective function $\mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right]$ can be optimized for gDoF by a point mass distribution. We have the form for $f\left(|a|^2, |b|^2, |c|^2\right)$ as

$$\begin{aligned}
&f\left(|a|^2, |b|^2, |c|^2\right)\\
&= \log\left(\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + 1\right)\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + 1\right) + \left(|c|^2 \rho_{12}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right)\right)
\end{aligned}$$

$$+ (T-1)\log\left(\left(|a|^2\rho_{11}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)+|b|^2\left(\rho_{12}^2+\rho_{22}^2\right)+\left(|a|^2\rho_{21}^2+1\right)\left(|c|^2\rho_{12}^2+1\right)\right)$$
$$-\log\left(\left(1+|a|^2\rho_{11}^2\right)\left(1+|c|^2\rho_{12}^2\right)+|b|^2\rho_{12}^2\right)$$
$$-\log\left(\left(1+|a|^2\rho_{21}^2\right)\left(1+|c|^2\rho_{22}^2\right)+|b|^2\rho_{22}^2\right). \tag{A.12}$$

Now

$$\begin{aligned}
&\frac{\partial}{\partial |a|^2} f\left(|a|^2,|b|^2,|c|^2\right)\\
&= \frac{\rho_{11}^2\left(|a|^2\rho_{21}^2+|b|^2\rho_{22}^2+1\right)+\left(|a|^2\rho_{11}^2+|b|^2\rho_{12}^2+1\right)\rho_{21}^2}{\left(|a|^2\rho_{11}^2+|b|^2\rho_{12}^2+1\right)\left(|a|^2\rho_{21}^2+|b|^2\rho_{22}^2+1\right)+\left(|c|^2\rho_{12}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)}\\
&\quad + (T-1)\frac{\rho_{11}^2\left(|c|^2\rho_{22}^2+1\right)+\left(|c|^2\rho_{12}^2+1\right)\rho_{21}^2}{\left(|a|^2\rho_{11}^2+1\right)\left(|c|^2\rho_{22}^2+1\right)+|b|^2\left(\rho_{12}^2+\rho_{22}^2\right)+\left(|a|^2\rho_{21}^2+1\right)\left(|c|^2\rho_{12}^2+1\right)}\\
&\quad - \frac{\left(\rho_{11}^2\right)\left(1+|c|^2\rho_{12}^2\right)}{\left(1+|a|^2\rho_{11}^2\right)\left(1+|c|^2\rho_{12}^2\right)+|b|^2\rho_{12}^2}\\
&\quad - \frac{\left(\rho_{21}^2\right)\left(1+|c|^2\rho_{22}^2\right)}{\left(1+|a|^2\rho_{21}^2\right)\left(1+|c|^2\rho_{22}^2\right)+|b|^2\rho_{22}^2}.
\end{aligned} \tag{A.13}$$

Hence

$$\left|\frac{\partial}{\partial |a|^2} f\left(|a|^2,|b|^2,|c|^2\right)\right| \le \rho_{11}^2+\rho_{21}^2+(T-1)\left(\rho_{11}^2+\rho_{21}^2\right)+\left(\rho_{11}^2\right)+\left(\rho_{21}^2\right) \tag{A.14}$$
$$\le 2(T+1)\max_{i,j}\rho_{ij}^2. \tag{A.15}$$

Similarly

$$\left|\frac{\partial}{\partial |b|^2} f\left(|a|^2,|b|^2,|c|^2\right)\right| \le 2(T+1)\max_{i,j}\rho_{ij}^2, \tag{A.16}$$
$$\left|\frac{\partial}{\partial |c|^2} f\left(|a|^2,|b|^2,|c|^2\right)\right| \le 2(T+1)\max_{i,j}\rho_{ij}^2 \tag{A.17}$$

holds. Let $\rho_*^2 = \max_{i,j}\rho_{ij}^2$. Now with $\Delta = 1/\left(2(T+1)\rho_*^2\right)$, if $\left\|\left(|a|^2,|b|^2,|c|^2\right)-\left(|a'|^2,|b'|^2,|c'|^2\right)\right\| \le \sqrt{3}\Delta$, then

$$\left|f\left(|a|^2,|b|^2,|c|^2\right)-f\left(|a'|^2,|b'|^2,|c'|^2\right)\right| \tag{A.18}$$
$$\le \left\|\left[2(T+1)\rho_*^2, 2(T+1)\rho_*^2, 2(T+1)\rho_*^2\right]\right\|\sqrt{3}\Delta \tag{A.19}$$
$$\le 3. \tag{A.20}$$

Hence by considering a discrete version of the problem as

$$\mathcal{P}_2 : \begin{cases} \underset{\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T}{\text{maximize}} \ \mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right] \\ \text{Support}\left(|a|^2, |b|^2, |c|^2\right) = \{0, \Delta, 2\Delta, \ldots, \infty\}^3 \end{cases} \tag{A.21}$$

the optimum value achieved is within 3 of the optimum value of $\mathcal{P}_1$. Hence for outer bound on gDoF, it is sufficient to solve $\mathcal{P}_2$.

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right). \tag{A.22}$$

We will now show that it is sufficient to restrict $\text{Support}\left(|a|^2, |b|^2, |c|^2\right) = \{0, \Delta, 2\Delta, \ldots, \lfloor\rho_*^4\rfloor\Delta\}^3$ for outer bound on gDoF.

Let the optimum value of $\mathcal{P}_2$ be achieved by a probability distribution $\{p_i^*\}$ at the points $\{(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta)\}$ with $l_{ji}^* \in \mathbb{Z}$. Let

$$S_1 = \left\{i : \max\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right) \leq \rho_*^4\Delta\right\}, \tag{A.23}$$

$$S_2 = \left\{i : \max\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right) > \rho_*^4\Delta\right\} \tag{A.24}$$

and let $\max\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right) = l_{Mi}^*\Delta$ for labeling. The optimum value $(\mathcal{P}_2)$ is given by

$$(\mathcal{P}_2) = \sum_{i\in S_1} p_i^* f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right) + \sum_{i\in S_2} p_i^* f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right). \tag{A.25}$$

We will now show that $\sum_{i\in S_2} p_i^* f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right)$ does not contribute to gDoF; the points in S_2 have large power and hence they have low probability due to power constraints; this ends up limiting the contribution to gDoF. We prove this precisely in the following steps. Using the structure of $f\left(|a|^2, |b|^2, |c|^2\right)$ and $\Delta = 1/\left(2\left(T+1\right)\rho_*^2\right)$, we can bound

$$\begin{aligned} \left|f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right)\right| \leq & \log\left(\left(2l_{Mi}^*+1\right)\left(2l_{Mi}^*+1\right) + \left(l_{Mi}^*+1\right)\left(l_{Mi}^*+1\right)\right) \\ & + \left(T-1\right)\log\left(\left(l_{Mi}^*+1\right)\left(l_{Mi}^*+1\right) + 2l_{Mi}^* + \left(l_{Mi}^*+1\right)\left(l_{Mi}^*+1\right)\right) \\ & + 2\log\left(\left(1+l_{Mi}^*\right)\left(1+l_{Mi}^*\right) + l_{Mi}^*\right) && \text{(A.26)} \\ \leq & \left(T+2\right)\log\left(\left(2l_{Mi}^*+1\right)\left(2l_{Mi}^*+1\right)3\right) && \text{(A.27)} \\ = & \, 2\left(T+2\right)\log\left(2l_{Mi}^*+1\right) + \left(T+2\right)\log\left(3\right). && \text{(A.28)} \end{aligned}$$

Hence

$$\left|\sum_{i\in S_2} p_i^* f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right)\right|$$

$$\leq \sum_{i\in S_2} p_i^* 2\left(T+2\right)\log\left(2l_{Mi}^*+1\right) + \left(T+2\right)\log\left(3\right) \tag{A.29}$$

$$\overset{(i)}{\leq} 2\left(T+2\right)\left(\sum_{i\in S_2} p_i^*\right)\log\left(2\frac{\sum_{i\in S_2} p_i^* l_{Mi}^*}{\sum_{j\in S_2} p_j^*}+1\right) + \left(T+2\right)\log\left(3\right) \tag{A.30}$$

$$\overset{(ii)}{\leq} 2\left(T+2\right)\left(\sum_{i\in S_2} p_i^*\right)\log\left(2\frac{T}{\Delta\sum_{j\in S_2} p_j^*}+1\right) + \left(T+2\right)\log\left(3\right) \tag{A.31}$$

$$= 2\left(T+2\right)\left(\sum_{i\in S_2} p_i^*\right)\log\left(2\frac{T}{\Delta}+\sum_{j\in S_2} p_j^*\right)$$

$$- 2\left(T+2\right)\left(\sum_{i\in S_2} p_i^*\right)\log\left(\sum_{j\in S_2} p_j^*\right) + \left(T+2\right)\log\left(3\right) \tag{A.32}$$

$$\overset{(iii)}{\leq} 2\left(T+2\right)\left(\sum_{i\in S_2} p_i^*\right)\log\left(2\frac{T}{\Delta}+1\right) + 2\left(T+2\right)\frac{\log\left(e\right)}{e} + \left(T+2\right)\log\left(3\right) \tag{A.33}$$

$$\overset{(iv)}{\leq} 2\left(T+2\right)\left(\frac{T}{\rho_*^4\Delta}\right)\log\left(2\frac{T}{\Delta}+1\right) + 2\left(T+2\right)\frac{\log\left(e\right)}{e} + \left(T+2\right)\log\left(3\right) \tag{A.34}$$

$$\overset{(v)}{=} 2\left(T+2\right)\left(\frac{2T\left(T+1\right)}{\rho_*^2}\right)\log\left(4T\left(T+1\right)\rho_*^2+1\right)$$

$$+ 2\left(T+2\right)\frac{\log\left(e\right)}{e} + \left(T+2\right)\log\left(3\right) \tag{A.35}$$

$$\overset{(vi)}{\leq} r_1\left(T\right) \text{ independent of } \mathsf{SNR}, \tag{A.36}$$

where (i) is due to Jensen's inequality, (ii) is due to the power constraint $\sum_{i\in S_2} p_i^* l_{Mi}^*\Delta \leq T \Rightarrow \sum_{i\in S_2} p_i^* l_{Mi}^* \leq \frac{T}{\Delta}$, (iii) is due to the fact $0 \leq \left(\sum_{i\in S_2} p_i^*\right) \leq 1$ and $-x\log\left(x\right) \leq \frac{\log(e)}{e}$ for $x \in [0,1]$, $(iv)$is due to the fact $\sum_{i\in S_2} p_i^* l_{Mi}^*\Delta \leq T$ (power constraint) and $\rho_*^4\Delta < l_{Mi}^*\Delta$ in S_2 and hence $\sum_{i\in S_2} p_i^*\rho_*^4\Delta \leq T$ and hence $\sum_{i\in S_2} p_i^* \leq \frac{T}{\rho_*^4\Delta}$, (v) is using $\Delta = \frac{1}{2(T+1)\rho_*^2}$ and (vi) is due to the fact $\frac{1}{x}\log\left(x\right)$ is bounded for $x \in [1,+\infty)$ and assuming $\rho_*^2 > 1$. Hence it follows that

$$\left(\mathcal{P}_2\right) = \sum_{i\in S_1} p_i^* f\left(l_{1i}^*\Delta, l_{2i}^*\Delta, l_{3i}^*\Delta\right) + r_1\left(T\right).$$

Hence it follows that

$$\mathcal{P}_3 : \begin{cases} \underset{\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T}{\text{maximize}} \ \mathbb{E}\left[f\left(|a|^2, |b|^2, |c|^2\right)\right] \\ \text{Support}\left(|a|^2, |b|^2, |c|^2\right) = S_1 \end{cases} \tag{A.37}$$

achieves the same gDoF as $\mathcal{P}_2$, because any non-zero probability outside S_1 in $\mathcal{P}_2$ can be assigned to $(0,0,0)$ in $\mathcal{P}_3$ by changing the value of objective function by a constant independent of SNR. Hence

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right) = \text{gDoF}\left(\mathcal{P}_3\right). \tag{A.38}$$

Now $\mathcal{P}_3$ is a linear program with finite number of variables and constraints (with a finite optimum value because of Jensen's inequality). The variables are $\{p_i\}$ and the maximum number of non trivial active constraints on $\{p_i\}$ is 2, derived from

$$\mathbb{E}\left[|a|^2 + |b|^2 + |c|^2\right] = T, \tag{A.39}$$

$$\sum p_i = 1. \tag{A.40}$$

Trivial constraints are $p_i \geq 0$. Hence by the theory of linear optimization, there exists an optimal $\{p_i^*\}$ with at most 2 non zero values. Hence it follows that

$$\mathcal{P}_4 : \begin{cases} \text{maximize} \sum_{i=1}^{2} p_i f_1\left(|a_i|^2, |b_i|^2, |c_i|^2\right) \\ \sum_{i=1}^{2} p_i \left(|a_i|^2 + |b_i|^2 + |c_i|^2\right) \leq T, \\ \sum p_i = 1, \\ |a_i|^2, |b_i|^2, |c_i|^2 \geq 0 \end{cases} \tag{A.41}$$

has $(\mathcal{P}_4) \geq (\mathcal{P}_3)$. Note that we have allowed $\left(|a_i|^2, |b_i|^2, |c_i|^2\right)_{i=1}^{2}$ to be real positive variables to be optimized. But it is also clear that $(\mathcal{P}_4) \leq (\mathcal{P}_1)$. Hence

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right) = \text{gDoF}\left(\mathcal{P}_3\right) = \text{gDoF}\left(\mathcal{P}_4\right). \tag{A.42}$$

Now consider

$$\mathcal{P}_5 : \begin{cases} \text{maximize} \sum_{i=1}^{2} p_i f_1\left(|a_i|^2, |b_i|^2, |c_i|^2\right) \\ p_i |a_i|^2 \le T, p_i |b_i|^2 \le T, p_i |c_i|^2 \le T, \\ \sum p_i = 1, \\ |a_i|^2, |b_i|^2, |c_i|^2 \ge 0. \end{cases} \tag{A.43}$$

It can be easily shown that $\text{gDoF}(\mathcal{P}_4) = \text{gDoF}(\mathcal{P}_5)$, we omit the proof.

Claim A.1. *Adding the constraints $|a_i|^2 \le T, |b_i|^2 \le T, |c_i|^2 \le T$ preserves gDoF$(\mathcal{P}_5)$.*

Proof. We have

$$\begin{aligned} & f\left(|a|^2, |b|^2, |c|^2\right) \\ & = \log\left(\left(|a|^2 \rho_{11}^2 + |b|^2 \rho_{12}^2 + 1\right)\left(|a|^2 \rho_{21}^2 + |b|^2 \rho_{22}^2 + 1\right) + \left(|c|^2 \rho_{12}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right)\right) \\ & \quad + (T-1)\log\left(\left(|a|^2 \rho_{11}^2 + 1\right)\left(|c|^2 \rho_{22}^2 + 1\right) + |b|^2\left(\rho_{12}^2 + \rho_{22}^2\right) + \left(|a|^2 \rho_{21}^2 + 1\right)\left(|c|^2 \rho_{12}^2 + 1\right)\right) \\ & \quad - \log\left(\left(1 + |a|^2 \rho_{11}^2\right)\left(1 + |c|^2 \rho_{12}^2\right) + |b|^2 \rho_{12}^2\right) \\ & \quad - \log\left(\left(1 + |a|^2 \rho_{21}^2\right)\left(1 + |c|^2 \rho_{22}^2\right) + |b|^2 \rho_{22}^2\right). \end{aligned} \tag{A.44}$$

Suppose $|a_i|^2 > T$ and consider

$$t_1 = p_i \log\left(\left(|a_i|^2 \rho_{11}^2 + |b_i|^2 \rho_{12}^2 + 1\right)\left(|a_i|^2 \rho_{21}^2 + |b_i|^2 \rho_{22}^2 + 1\right) + \left(|c_i|^2 \rho_{12}^2 + 1\right)\left(|c_i|^2 \rho_{22}^2 + 1\right)\right).$$

We will show that setting $|a_i|^2 = T$ would change the value of t_1 only by a constant independent of SNR. The other terms have a similar structure and can be handled in a similar way. If $\left(|c_i|^2 \rho_{12}^2 + 1\right)\left(|c_i|^2 \rho_{22}^2 + 1\right) > \left(|a_i|^2 \rho_{11}^2 + |b_i|^2 \rho_{12}^2 + 1\right)\left(|a_i|^2 \rho_{21}^2 + |b_i|^2 \rho_{22}^2 + 1\right)$ the claim is trivially true that we can replace $|a_i|^2 > T$ with $|a_i|^2 = T$ while changing the value of t_1 by only a constant. Otherwise

$$t_1 \doteq p_i \log\left(\left(|a_i|^2 \rho_{11}^2 + |b_i|^2 \rho_{12}^2 + 1\right)\left(|a_i|^2 \rho_{21}^2 + |b_i|^2 \rho_{22}^2 + 1\right)\right) \tag{A.45}$$

$$= \underbrace{p_i \log\left(|a_i|^2 \rho_{11}^2 + |b_i|^2 \rho_{12}^2 + 1\right)}_{t_{11}} + \underbrace{p_i \log\left(|a_i|^2 \rho_{21}^2 + |b_i|^2 \rho_{22}^2 + 1\right)}_{t_{12}}. \tag{A.46}$$

Now consider $t_{11} = p_i \log\left(|a_i|^2 \rho_{11}^2 + |b_i|^2 \rho_{12}^2 + 1\right)$. If $|a_i|^2 \rho_{11}^2 < |b_i|^2 \rho_{12}^2 + 1$ we can replace $|a_i|^2 > T$ with $|a_i|^2 = T$.

If $|a_i|^2 \rho_{11}^2 > |b_i|^2 \rho_{12}^2 + 1$ then $t_2 \doteq p_i \log\left(|a_i|^2 \rho_{11}^2 + 1\right)$ where the approximation is tight within a constant (constant less than 1). Now if we replace $|a_i|^2 > T$ with $|a_i|^2 = T$ the difference arising is bounded independent of SNR:

$$p_i \log\left(|a_i|^2 \rho_{11}^2 + 1\right) \overset{(i)}{\leq} p_i \log\left(\frac{T}{p_i}\rho_{11}^2 + 1\right) \tag{A.47}$$

$$= p_i \log\left(T\rho_{11}^2 + p_i\right) - p_i \log\left(p_i\right) \tag{A.48}$$

$$\leq p_i \log\left(T\rho_{11}^2 + 1\right) - p_i \log\left(p_i\right), \tag{A.49}$$

where (i) is because $p_i |a_i|^2 \leq T$ due to the power constraint. Also $|a_i|^2 > T$, hence it follows that

$$\left|p_i \log\left(|a_i|^2 \rho_{11}^2 + 1\right) - p_i \log\left(T\rho_{11}^2 + 1\right)\right| \leq |p_i \log\left(p_i\right)| \tag{A.50}$$

$$\leq \frac{\log\left(e\right)}{e}. \tag{A.51}$$

Following the same logic for other terms, it can be shown that adding the constraints $|a_i|^2 \leq T,\ |b_i|^2 \leq T,\ |c_i|^2 \leq T$ preserves $\text{gDoF}\left(\mathcal{P}_5\right)$. $\square$

With the additional constraints $|a_i|^2 \leq T,\ |b_i|^2 \leq T,\ |c_i|^2 \leq T$ the existing constraints $p_i |a_i|^2 \leq T,\ p_i |b_i|^2 \leq T,\ p_i |c_i|^2 \leq T$ become redundant. Hence we have $\text{gDoF}\left(\mathcal{P}_5\right) = \text{gDoF}\left(\mathcal{P}_6\right)$ for $\mathcal{P}_6$ defined as

$$\mathcal{P}_6 : \begin{cases} \text{maximize} \sum_{i=1}^{2} p_i f\left(|a_i|^2, |b_i|^2, |c_i|^2\right) \\ |a_i|^2 \leq T,\ |b_i|^2 \leq T,\ |c_i|^2 \leq T, \\ \sum p_i = 1. \end{cases} \tag{A.52}$$

It is clear from the structure of $\mathcal{P}_6$ that the solution has $\left(|a_1|^2, |b_1|^2, |c_1|^2\right) = \left(|a_2|^2, |b_2|^2, |c_2|^2\right)$ hence it suffices to solve $\mathcal{P}_7$ defined as

$$\mathcal{P}_7 : \begin{cases} \text{maximize} f\left(|a|^2, |b|^2, |c|^2\right) \\ |a|^2 \leq T,\ |b|^2 \leq T,\ |c|^2 \leq T, \end{cases} \tag{A.53}$$

that is it suffices to consider one point mass distribution.

A.3 Proof of Theorem 2.2: Decomposing into disjoint parts of MIMO graph

Here we prove that for a MIMO whose graph can be decomposed into disjoint parts, the capacity can be achieved by allocating power to the disjoint parts separately. Let the channel matrix G of the system be block diagonal as $G = \operatorname{diag}(G_1, \ldots, G_K)$, where G_i are the diagonal blocks corresponding to the disjoint parts of the graph, then the capacity $C(P, \operatorname{diag}(G_1, \ldots, G_K))$ of the channel for a power P can be achieved by splitting power across the blocks, *i.e.,* $C(P, \operatorname{diag}(G_1, \ldots, G_K)) = \max_{P_1+\cdots+P_K \leq P} (C(P_1, G_1) + \cdots + C(P_K, G_K))$. We just need to show that for

$$G = \operatorname{diag}(G_1, G_2)$$

the capacity of the channel can be achieved by a power splitting across the two blocks of channels G_1, G_2 i.e

$$C(P, \operatorname{diag}(G_1, G_2)) = \max_{P_1+P_2 \leq P} (C(P_1, G_1) + C(P_2, G_2)) \tag{A.54}$$

and the general result for multiple disjoint parts in MIMO graph will follow due to induction.

$$h(Y) \overset{(ii)}{\leq} h(Y_{G1}) + h(Y_{G2}) \tag{A.55}$$

$$h(Y|\, X) = h(Y_{G1} Y_{G2}|\, X_{G1} X_{G2}) \tag{A.56}$$

$$= h(Y_{G1}|\, X_{G1} X_{G2}) + h(Y_{G2}|\, Y_{G1} X_{G1} X_{G2}) \tag{A.57}$$

$$\overset{(ii)}{=} h(Y_{G1}|\, X_{G2}) + h(Y_{G2}|\, X_{G2}), \tag{A.58}$$

where (i) is because conditioning reduces entropy and (ii) is because $X_{G2} - X_{G1} - Y_{G1}$ and $(X_{G1}, Y_{G1}) - X_{G2} - Y_{G2}$ are Markov chains.

Hence

$$I(X; Y) \leq I(X_{G1}; Y_{G1}) + I(X_{G2}; Y_{G2}) \tag{A.59}$$

subject to $\mathbb{E}\left[\|X_{G1}\|^2+\|X_{G2}\|^2\right]\le P$. The RHS can be achieved by treating the two blocks of channels G_1, G_2 separately with a power allocation, hence

$$C\left(P,\operatorname{diag}\left(G_1,G_2\right)\right)=\max_{P_1+P_2\le P}\left(C\left(P_1,G_1\right)+C\left(P_2,G_2\right)\right). \tag{A.60}$$

A.4 Inner bound for 2×2 symmetric MIMO

Here we prove the achievability result from Theorem 2.7 for the 2×2 MIMO with exponents γ_D in the direct links and γ_{CL} in the crosslinks. We use the input distribution $X=\begin{bmatrix} a & 0 & 0 & . & . & 0\\ \eta & c & 0 & . & . & 0\end{bmatrix}Q$ with constant a, c and $\eta\sim\mathcal{CN}\left(0,|b|^2\right)$ with

$$|a|^2=\mathsf{SNR}^{\gamma_a}, |b|^2=\mathsf{SNR}^{\gamma_b}, |c|^2=\mathsf{SNR}^{\gamma_c}, \gamma_a\le 0, \gamma_b\le 0, \gamma_c\le 0.$$

With this choice, we proceed to lower bound $I\left(X;Y\right)$.

$$I\left(X;Y\right)=h\left(Y\right)-h\left(Y|\,X\right) \tag{A.61}$$

$$h\left(Y\right)=h\left(GX+W\right) \tag{A.62}$$

$$\ge h\left(GX\right) \tag{A.63}$$

$$=h\left(\begin{bmatrix} g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix}\begin{bmatrix} a & 0 & 0 & . & . & 0\\ \eta & c & 0 & . & . & 0\end{bmatrix}Q\right). \tag{A.64}$$

Now

$$\begin{aligned}&\begin{bmatrix} g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix}\begin{bmatrix} a & 0 & 0 & . & . & 0\\ \eta & c & 0 & . & . & 0\end{bmatrix}\\ &=\begin{bmatrix} ag_{11}+\eta g_{12} & cg_{12} & 0 & . & . & 0\\ ag_{21}+\eta g_{22} & cg_{22} & 0 & . & . & 0\end{bmatrix} &\text{(A.65)}\\ &=\begin{bmatrix} \sqrt{|ag_{11}+\eta g_{12}|^2+|cg_{12}|^2} & 0 & 0 & . & . & 0\\ \frac{(ag_{21}+\eta g_{22})(ag_{11}+\eta g_{12})^*+|c|^2g_{22}g_{12}^*}{\sqrt{|ag_{11}+\eta g_{12}|^2+|cg_{12}|^2}} & \frac{(ag_{21}+\eta g_{22})cg_{12}-cg_{22}(ag_{11}+\eta g_{12})}{\sqrt{|ag_{11}+\eta g_{12}|^2+|cg_{12}|^2}} & 0 & . & . & 0\end{bmatrix}\Phi, &\text{(A.66)}\end{aligned}$$

where the last step performed an LQ transformation and Φ is unitary. Hence due to the property of isotropic unitary matrices and steps similar to (2.48) to (2.51) in Section 2.4.4, we get:

$$
\begin{aligned}
& h\left(GX\right) \\
& = h\left(\sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}\overline{q_1}^{(T)}\right) \\
& \quad + h\left(\left[\frac{\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^*+\left|c\right|^2 g_{22}g_{12}^*}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}}, \frac{\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}}\overline{q}_2^{(T-1)}\right]\middle|\,\xi_{11}\right) \\
& = h\left(\sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}\overline{q_1}^{(T)}\right) - T\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2\right)\right] \\
& \quad + \underbrace{h\left(\left[\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^*+\left|c\right|^2 g_{22}g_{12}^*, \left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right)\overline{q}_2^{(T-1)}\right]\middle|\,\xi_{11}\right)}_{\alpha},
\end{aligned}
\tag{A.67}
$$

where $\overline{q_1}^{(i)}$ is i dimensional unitary isotropically distributed vector and $\xi_{11} = \sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}$. Now

$$
\begin{aligned}
& \alpha \\
& = h\left(\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right)\overline{q}_2^{(T-1)}\middle|\,\xi_{11}\right) \\
& \quad + h\left(\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^*+\left|c\right|^2 g_{22}g_{12}^*\middle|\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right)\overline{q}_2^{(T-1)}, \xi_{11}\right)
\end{aligned}
\tag{A.68}
$$

$$
\begin{aligned}
& \overset{(i)}{\geq} \left(T-2\right)\mathbb{E}\left[\log\left(\left|\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right|^2\right)\right] + \log\left(\frac{\pi^{T-1}}{\Gamma\left(T-1\right)}\right) \\
& \quad + h\left(\left|\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right|^2\middle|\,\xi_{11}\right) \\
& \quad + h\left(\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^*+\left|c\right|^2 g_{22}g_{12}^*\middle|\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right), \overline{q}_2^{(T-1)}, \xi_{11}\right)
\end{aligned}
\tag{A.69}
$$

$$
\begin{aligned}
& \overset{(ii)}{=} \left(T-2\right)\mathbb{E}\left[\log\left(\left|\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right|^2\right)\right] + \log\left(\frac{\pi^{T-1}}{\Gamma\left(T-1\right)}\right) \\
& \quad + h\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\middle|\,\xi_{11}\right) - \log\left(\pi\right)
\end{aligned}
$$

$$+h\left(\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^{*}+\left|c\right|^{2}g_{22}g_{12}^{*}\middle|\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right),\xi_{11}\right) \tag{A.70}$$

$$\overset{(iii)}{=}\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^{2}\right)\right]+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)+2\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}\right)\right]$$
$$h\left(\left[\frac{\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^{*}+\left|c\right|^{2}g_{22}g_{12}^{*}}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}},\frac{\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}}\right]\middle|\xi_{11}\right) \tag{A.71}$$

$$\overset{(iv)}{=}\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^{2}\right)\right]+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)+2\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}\right)\right]$$
$$h\left(\left[\frac{\left(ag_{21}+\eta g_{22}\right)\left(ag_{11}+\eta g_{12}\right)^{*}+\left|c\right|^{2}g_{22}g_{12}^{*}}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}},\frac{\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)}{\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}}\right]\middle|ag_{11}+\eta g_{12},g_{12}\right) \tag{A.72}$$

$$\overset{(v)}{=}\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^{2}\right)\right]+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)+2\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}\right)\right]$$
$$+h\left(\left[\begin{array}{cc} ag_{21}+\eta g_{22} & cg_{22}\end{array}\right]\middle|ag_{11}+\eta g_{12},g_{12}\right), \tag{A.73}$$

where (i) is using Lemma 3.2 in the first term and conditioning reduces entropy in the second term, (ii) is using Lemma 3.2 on $h\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\middle|\,\xi_{11}\right)$ (note that with $\theta\sim\mathrm{Unif}\left[0,2\pi\right]$ independent, $\left(\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\right)e^{i\theta}\middle|\,\xi_{11}$ and $\left(ag_{21}+\eta g_{22}\right)cg_{12}-cg_{22}\left(ag_{11}+\eta g_{12}\right)\middle|\,\xi_{11}$ have the same distribution; $e^{i\theta}$ is the unitary distribution in one dimension; hence Lemma 3.2 can be applied), (iii) is simply using $\xi_{11}=\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}$ and rearranging terms, (iv) is because conditioning reduces entropy and (v) is by a unitary transformation in the last term. Hence by substituting (A.73) in (A.67) we have

$$\begin{aligned}&h\left(GX\right)\\&\geq h\left(\sqrt{\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}}\,\overline{q_{1}}^{(T)}\right)-\left(T-2\right)\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^{2}+\left|cg_{12}\right|^{2}\right)\right]\\&\quad+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)+\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^{2}\right)\right]\\&\quad+h\left(\left[\begin{array}{cc} ag_{21}+\eta g_{22} & cg_{22}\end{array}\right]\middle|ag_{11}+\eta g_{12},g_{12}\right)\end{aligned} \tag{A.74}$$

$$\stackrel{(i)}{=} h\left(\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2\right)+\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2\right)\right]+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)$$
$$+\log\left(\frac{\pi^T}{\Gamma\left(T\right)}\right)+\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^2\right)\right]$$
$$+h\left(\left[\begin{array}{cc} ag_{21}+\eta g_{22} & cg_{22}\end{array}\right]\middle| ag_{11}+\eta g_{12}, g_{12}\right), \tag{A.75}$$

where (i) is using Lemma 3.2 on $h\left(\sqrt{\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2}\overline{q_1}^{(T)}\right)$. Also

$$h\left(\left[\begin{array}{cc} ag_{21}+\eta g_{22} & cg_{22}\end{array}\right]\middle| ag_{11}+\eta g_{12}, g_{12}\right)=h\left(cg_{22}\right)+h\left(ag_{21}+\eta g_{22}\middle| ag_{11}+\eta g_{12}, g_{12}, g_{22}\right). \tag{A.76}$$

Using $\eta\sim\mathcal{CN}\left(0,|b|^2\right)$

$$h\left(ag_{21}+\eta g_{22}\middle| ag_{11}+\eta g_{12}, g_{12}, g_{22}\right)$$
$$=h\left(ag_{21}+\eta g_{22}, ag_{11}+\eta g_{12}\middle| g_{12}, g_{22}\right)-h\left(ag_{11}+\eta g_{12}\middle| g_{12}, g_{22}\right) \tag{A.77}$$
$$=\mathbb{E}\left[\log\left(\pi e\left|\begin{array}{cc} |a|^2\,\mathsf{SNR}^{\gamma_{21}}+|b|^2\left|g_{22}\right|^2 & |b|^2\,g_{22}g_{12}^* \\ |b|^2\,g_{22}^*g_{12} & |a|^2\,\mathsf{SNR}^{\gamma_{11}}+|b|^2\left|g_{12}\right|^2\end{array}\right|\right)\right]$$
$$-\mathbb{E}\left[\log\left(|a|^2\,\mathsf{SNR}^{\gamma_d}+|b|^2\left|g_{12}\right|^2\right)\right] \tag{A.78}$$
$$\stackrel{(i)}{\geq}\log\left(|a|^4\,\mathsf{SNR}^{\gamma_{11}+\gamma_{21}}+|a|^2|b|^2\,\mathsf{SNR}^{\gamma_{12}+\gamma_{21}}+|a|^2|b|^2\,\mathsf{SNR}^{\gamma_{11}+\gamma_{22}}\right)$$
$$-\log\left(|a|^2\,\mathsf{SNR}^{\gamma_{11}}+|b|^2\,\mathsf{SNR}^{\gamma_{12}}\right)-2\gamma_E\log\left(e\right), \tag{A.79}$$

where (i) is using Fact 5.1 from page 135 on $\left|g_{22}\right|^2$ and $\left|g_{12}\right|^2$. Now substituting (A.79) and (A.76) in (A.75) we get

$$h\left(GX\right)\geq h\left(\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2\right)+\mathbb{E}\left[\log\left(\left|ag_{11}+\eta g_{12}\right|^2+\left|cg_{12}\right|^2\right)\right]+\log\left(\frac{\pi^{T-2}}{\Gamma\left(T-1\right)}\right)$$
$$+\log\left(\frac{\pi^T}{\Gamma\left(T\right)}\right)+\left(T-2\right)\mathbb{E}\left[\log\left(\left|acg_{12}g_{21}-acg_{11}g_{22}\right|^2\right)\right]+h\left(cg_{22}\right)-2\gamma_E\log\left(e\right)$$
$$+\log\left(|a|^4\,\mathsf{SNR}^{\gamma_{11}+\gamma_{21}}+|a|^2|b|^2\,\mathsf{SNR}^{\gamma_{12}+\gamma_{21}}+|a|^2|b|^2\,\mathsf{SNR}^{\gamma_{11}+\gamma_{22}}\right)$$
$$-\log\left(|a|^2\,\mathsf{SNR}^{\gamma_{11}}+|b|^2\,\mathsf{SNR}^{\gamma_{12}}\right). \tag{A.80}$$

Now we use our choice $\eta \sim \mathcal{CN}\left(0, |b|^2\right)$ $|a|^2 = \mathsf{SNR}^{\gamma_a}, |b|^2 = \mathsf{SNR}^{\gamma_b}, |c|^2 = \mathsf{SNR}^{\gamma_c}, \gamma_a \leq 0, \gamma_b \leq 0, \gamma_c \leq 0$. We have

$$
\begin{aligned}
& h\left(|ag_{11} + \eta g_{12}|^2 + |cg_{12}|^2\right) \\
& \overset{(i)}{\geq} \max\left(h\left(|ag_{11} + \eta g_{12}|^2\right), h\left(|cg_{12}|^2\right)\right) && \text{(A.81)} \\
& \geq \max\left(h\left(|ag_{11} + \eta g_{12}|^2 \middle| g_{12}\right), h\left(|cg_{12}|^2\right)\right) && \text{(A.82)} \\
& \overset{(ii)}{\doteq} \max\left(\mathbb{E}\left[\log\left(\mathsf{SNR}^{\gamma_a+\gamma_{11}} + \mathsf{SNR}^{\gamma_b}|g_{12}|^2\right)\right], \log\left(\mathsf{SNR}^{\gamma_c+\gamma_{12}}\right)\right) && \text{(A.83)} \\
& \doteq \max\left(\log\left(\mathsf{SNR}^{\gamma_a+\gamma_{11}} + \mathsf{SNR}^{\gamma_b+\gamma_{12}}\right), \log\left(\mathsf{SNR}^{\gamma_c+\gamma_{12}}\right)\right), && \text{(A.84)}
\end{aligned}
$$

where (i) was using conditioning reduces entropy, (ii) was using property of exponential distributions and (iii) was using Fact 5.1. Hence

$$
\lim_{\mathsf{SNR}\to\infty} h\left(|ag_{11} + \eta g_{12}|^2 + |cg_{12}|^2\right) / \log(\mathsf{SNR}) \geq \max\left(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}, \gamma_c + \gamma_{12}\right).
$$

Now

$$
\begin{aligned}
& \mathbb{E}\left[\log\left(|ag_{11} + \eta g_{12}|^2 + |cg_{12}|^2\right)\right] \\
& \geq \max\left(\mathbb{E}\left[\log\left(|ag_{11} + \eta g_{12}|^2\right)\right], \mathbb{E}\left[\log\left(|cg_{12}|^2\right)\right]\right) && \text{(A.85)} \\
& \overset{(i)}{\doteq} \max\left(\log\left(\mathsf{SNR}^{\gamma_a+\gamma_{11}} + \mathsf{SNR}^{\gamma_b+\gamma_{12}}\right), \log\left(\mathsf{SNR}^{\gamma_c+\gamma_{12}}\right)\right), && \text{(A.86)}
\end{aligned}
$$

where (i) was using Fact 5.1. Also

$$
\mathbb{E}\left[\log\left(|acg_{12}g_{21} - acg_{11}g_{22}|^2\right)\right] \doteq \log\left(\mathsf{SNR}^{\gamma_a+\gamma_c+\gamma_{12}+\gamma_{21}} + \mathsf{SNR}^{\gamma_a+\gamma_c+\gamma_{11}+\gamma_{22}}\right) \qquad \text{(A.87)}
$$

using Fact 5.1 repeatedly. Similarly evaluating other terms in A.80, we get

$$
\begin{aligned}
& \lim_{\mathsf{SNR}\to\infty} h(GX) / \log(\mathsf{SNR}) \\
& \geq 2\max\left(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}, \gamma_c + \gamma_{12}\right) \\
& \quad + (T-2)\left(\gamma_a + \gamma_c + \max\left(\gamma_{12} + \gamma_{21}, \gamma_{11} + \gamma_{22}\right)\right) + +\gamma_c + \gamma_{22} \\
& \quad + \max\left(2\gamma_a + \gamma_{11} + \gamma_{21}, \gamma_a + \gamma_b + \gamma_{12} + \gamma_{21}, \gamma_a + \gamma_b + \gamma_{11} + \gamma_{22}\right) \\
& \quad - \max\left(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}\right). && \text{(A.88)}
\end{aligned}
$$

Also using (3.50), (3.52) we have

$$\begin{aligned}
&h(Y|X) \\
&= \mathbb{E}\left[\log\left(|a|^2\rho_{11}^2 + |\eta|^2\rho_{12}^2 + |c|^2\rho_{12}^2 + |a|^2|c|^2\rho_{11}^2\rho_{12}^2 + 1\right)\right] \\
&\quad + \mathbb{E}\left[\log\left(|a|^2\rho_{21}^2 + |\eta|^2\rho_{22}^2 + |c|^2\rho_{22}^2 + |a|^2|c|^2\rho_{21}^2\rho_{22}^2 + 1\right)\right] \\
&\quad + 2T\log(\pi e). \qquad \text{(A.89)}
\end{aligned}$$

Now since $\eta \sim \mathcal{CN}\left(0, |b|^2\right)$ and $|a|^2 = \mathsf{SNR}^{\gamma_a}, |b|^2 = \mathsf{SNR}^{\gamma_b}, |c|^2 = \mathsf{SNR}^{\gamma_c}$ and Fact 5.1, we get

$$\begin{aligned}
&h(Y|X) \\
&\doteq \log\left(|a|^2\rho_{11}^2 + |b|^2\rho_{12}^2 + |c|^2\rho_{12}^2 + |a|^2|c|^2\rho_{11}^2\rho_{12}^2 + 1\right) \\
&\quad + \log\left(|a|^2\rho_{21}^2 + |b|^2\rho_{22}^2 + |c|^2\rho_{22}^2 + |a|^2|c|^2\rho_{21}^2\rho_{22}^2 + 1\right) \qquad \text{(A.90)} \\
&\doteq \max(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}, \gamma_c + \gamma_{12}, \gamma_a + \gamma_c + \gamma_{11} + \gamma_{12}, 0) \\
&\quad + \max(\gamma_a + \gamma_{21}, \gamma_b + \gamma_{22}, \gamma_c + \gamma_{22}, \gamma_a + \gamma_c + \gamma_{21} + \gamma_{22}, 0), \qquad \text{(A.91)}
\end{aligned}$$

and hence

$$\begin{aligned}
&\lim_{\mathsf{SNR}\to\infty} h(Y|X)/\log(\mathsf{SNR}) \\
&= \max(\gamma_a + \gamma_{11}, \gamma_b + \gamma_{12}, \gamma_c + \gamma_{12}, \gamma_a + \gamma_c + \gamma_{11} + \gamma_{12}, 0) \\
&\quad + \max(\gamma_a + \gamma_{21}, \gamma_b + \gamma_{22}, \gamma_c + \gamma_{22}, \gamma_a + \gamma_c + \gamma_{21} + \gamma_{22}, 0). \qquad \text{(A.92)}
\end{aligned}$$

Using (A.88),(A.92) with $\gamma_a = 0, \gamma_c = 0, \gamma_b = 0\gamma_{11} = \gamma_{22} = \gamma_D > \gamma_{CL} = \gamma_{12} = \gamma_{21}$ we get

$$\lim_{\mathsf{SNR}\to\infty} h(GX)/\log(\mathsf{SNR}) \geq 2T\gamma_D, \qquad \text{(A.93)}$$

and

$$\lim_{\mathsf{SNR}\to\infty} h(Y|X)/\log(\mathsf{SNR}) = 2(\gamma_D + \gamma_{CL}). \qquad \text{(A.94)}$$

Hence we have

$$\lim_{\mathsf{SNR}\to\infty} (1/T)\, I(X;Y)/\log(\mathsf{SNR}) \geq 2\left((1 - 1/T)\gamma_D - (1/T)\gamma_{CL}\right) \qquad \text{(A.95)}$$

achievable. Also with $\gamma_a = 0, \gamma_c = -\gamma_{CL}, \gamma_b = 0 \gamma_{11} = \gamma_{22} = \gamma_D > \gamma_{CL} = \gamma_{12} = \gamma_{21}$ in (A.88),(A.92) we get

$$\lim_{\mathsf{SNR}\to\infty} h\left(GX\right)/\log\left(\mathsf{SNR}\right) \geq 2\gamma_D + \left(T-1\right)\left(2\gamma_D - \gamma_{CL}\right) \tag{A.96}$$

and

$$\lim_{\mathsf{SNR}\to\infty} h\left(Y|X\right)/\log\left(\mathsf{SNR}\right) = 2\gamma_D \tag{A.97}$$

Hence for $T = 2$

$$\lim_{\mathsf{SNR}\to\infty} \left(1/2\right) I\left(X;Y\right)/\log\left(\mathsf{SNR}\right) \geq \left(\gamma_D - \left(1/2\right)\gamma_{CL}\right) \tag{A.98}$$

is achievable. Hence the outer bounds for all regimes of T from Table 2.3 are achievable.

A.5 Gaussian codebooks for asymmetric MIMO

Here we prove Theorem 2.8 for an $M \times M$ MIMO (Figure 2.5) with coherence time $T > M$ and with exponents γ_D in direct links and γ_{CL} in crosslinks ($\gamma_D > \gamma_{CL}$). We consider i.i.d. Gaussian codebooks across antennas and time periods and prove that a gDoF of $M\left(\left(1-\frac{1}{T}\right)\gamma_D - \frac{M-1}{T}\gamma_{CL}\right)$ is achievable. Using Gaussian codebooks, the rate $R \geq I\left(GX + W; X\right)$ is achievable, where

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1T} \\ \vdots & & \vdots \\ X_{M1} & \cdots & X_{MT} \end{bmatrix} = \begin{bmatrix} \overline{X_1} & \ldots & \overline{X_T} \end{bmatrix}, \tag{A.99}$$

$$\overline{X_i} = \mathsf{Tran}\begin{bmatrix} X_{1i} & \ldots & X_{Mi} \end{bmatrix} \tag{A.100}$$

with all of the elements of the $M \times T$ matrix X being i.i.d. $\mathcal{CN}\left(0,1\right)$ and W being an $M \times T$ matrix with i.i.d. $\mathcal{CN}\left(0,1\right)$ noise elements. The channel matrix

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdot & g_{1M} \\ g_{21} & g_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ g_{M1} & \cdot & \cdot & g_{MM} \end{bmatrix} \tag{A.101}$$

has $g_{ii} \sim \mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_D}\right)$ and rest of the elements distributed according to $\mathcal{CN}\left(0, \mathsf{SNR}^{\gamma_{CL}}\right)$. We will show that the mutual information satisfies

$$I\left(GX+W;X\right) \dot{\geq} M\left(\left(T-1\right)\gamma_D - \left(M-1\right)\gamma_{CL}\right)\log\left(\mathsf{SNR}\right). \tag{A.102}$$

We have

$$\begin{aligned} I\left(GX+W;X\right) &= h\left(GX+W\right) - h\left(GX+W\middle| X\right) && \text{(A.103)}\\ &\geq h\left(GX+W\middle| G\right) \\ &\quad - h\left(GX+W\middle| X\right). && \text{(A.104)} \end{aligned}$$

Now

$$\begin{aligned} & h\left(GX+W\middle| G\right) \\ & \overset{(i)}{\geq} h\left(GX\middle| G\right) && \text{(A.105)}\\ & \overset{(ii)}{=} T \times h\left(G\overline{X_1}\middle| G\right) && \text{(A.106)}\\ & \overset{(iii)}{=} T\mathbb{E}\left[\log\left(\left|\det\left(\pi e G\right)\right|^2\right)\right] && \text{(A.107)}\\ & \overset{(iv)}{\doteq} TM\gamma_D \log\left(\mathsf{SNR}\right), && \text{(A.108)} \end{aligned}$$

where (i) is using conditioning reduces entropy after conditioning on W, (ii) is using the structure of X from A.99and the fact that elements X_{ij} are i.i.d. Gaussian, (iii) is again using the fact that X_{ij} are i.i.d. Gaussian and (iv) is by repeated application of Fact 5.1, Tower property of expectation on Gaussian distributed g_{ij} and the structure of the determinant involved. Now we will show that

$$h\left(GX+W\middle| X\right) \dot{\leq} M\left(\gamma_D \log\left(\mathsf{SNR}\right) + \left(M-1\right)\gamma_{CL}\log\left(\mathsf{SNR}\right)\right) \tag{A.109}$$

and will complete the proof.

$$\begin{aligned} & h\left(GX+W\middle| X\right) \\ & \overset{(i)}{\dot{\leq}} \sum_i h\left(\left[\begin{array}{cccc} g_{i1} & g_{i2} & . & g_{iM} \end{array}\right] X + W_i \middle| X\right) && \text{(A.110)} \end{aligned}$$

$$\overset{(ii)}{=} Mh\left(\left[\begin{array}{cccc} g_{11} & g_{12} & . & g_{1M} \end{array}\right] X + W_1 \Big| X\right), \tag{A.111}$$

where (i) is using conditioning reduces entropy and W_i is $1\times T$ with i.i.d. $\mathcal{CN}(0,1)$ elements, (ii) is by symmetry of the channels ($g_{ii} \sim \mathcal{CN}(0, \mathsf{SNR}^{\gamma_D})$ and rest of the elements distributed according to $\mathcal{CN}(0, \mathsf{SNR}^{\gamma_{CL}})$) and i.i.d. nature of X_{ij}. Now we will show that

$$h\left(\left[\begin{array}{cccc} g_{11} & g_{12} & . & g_{1M} \end{array}\right] X + W_1 \Big| X\right) \dot{\leq} \gamma_D \log(\mathsf{SNR}) + (M-1)\gamma_{CL}\log(\mathsf{SNR}) \tag{A.112}$$

and will complete the proof. Let us denote $W_1 = \left[\begin{array}{cccc} w_{11} & w_{12} & . & w_{1T} \end{array}\right]$, $\underline{g_1} = \left[\begin{array}{cccc} g_{11} & g_{12} & . & g_{1M} \end{array}\right]$. We have

$$\begin{aligned}
& h\left(\underline{g_1} X + W_1 \big| X\right) \\
& \leq h\left(\underline{g_1}\overline{X_1} + w_{11} \big| X\right) \\
& \quad + \sum_{i=2}^{M} h\left(\underline{g_1}\overline{X_i} + w_{1i} \big| X, \underline{g_1}\overline{X_1} + w_{11}\right) \\
& \quad + \sum_{i=M+1}^{T} h\left(\underline{g_1}\overline{X_i} + w_{1i} \big| X \left\{\underline{g_1}\overline{X_k} + w_{1k}\right\}_{k=1}^{M}\right).
\end{aligned} \tag{A.113}$$

Now the first term in (A.113)

$$\begin{aligned}
h\left(\underline{g_1}\overline{X_1} + w_{11} \big| X\right) &= h\left(\sum_{j=1}^{M} g_{j1}X_{j1} + w_{11} \Bigg| X_{j1}\right) \\
&\dot{\leq} \gamma_D \log(\mathsf{SNR})
\end{aligned} \tag{A.114}$$

using maximum entropy results and since $\gamma_D \geq \gamma_{CL}$. Now consider the second term in (A.113), $h\left(\underline{g_1}\overline{X_i} + w_{1i} \big| X, \underline{g_1}\overline{X_1} + w_{11}\right)$. In $\underline{g_1} = \left[\begin{array}{cccc} g_{11} & g_{12} & . & g_{1M} \end{array}\right]$, only g_{11} has SNR exponent γ_D and it can be removed due to the conditioning as follows:

$$\begin{aligned}
& h\left(\underline{g_1}\overline{X_i} + w_{1i} \big| X, \underline{g_1}\overline{X_1} + w_{11}\right) \\
& \overset{(i)}{\leq} h\left(\underline{g_1}\left[\begin{array}{c} 0 \\ X_{11}X_{2i} - X_{1i}X_{21} \\ . \\ X_{11}X_{Mi} - X_{1i}X_{M1} \end{array}\right] + X_{11}w_{1i} - X_{1i}w_{11} \Bigg| X\right) - \mathbb{E}\left[\log\left(|X_{11}|\right)\right]
\end{aligned} \tag{A.115}$$

$$= \mathbb{E}\left[\log\left(\pi e\left(\rho_{CL}^2\sum_{j=2}^{M}|X_{11}X_{ji}-X_{1i}X_{j1}|^2+|X_{11}|^2+|X_{1i}|^2\right)\right)\right]-\mathbb{E}\left[\log\left(|X_{11}|\right)\right] \quad \text{(A.116)}$$

$$\overset{(ii)}{\doteq} \gamma_{CL}\log\left(\mathsf{SNR}\right), \quad \text{(A.117)}$$

where (i) is by multiplying $\underline{g_1}\overline{X_i}+w_{1i}$ with X_{11} and subtracting $X_{1i}\left(\underline{g_1}\overline{X_1}+w_{11}\right)$ from it and using conditioning reduces entropy, and (ii) is by repeated application of Fact 5.1 and Tower property of expectation on Gaussian distributed X_{ij}.

Now consider the last term in (A.113)

$$h\left(\underline{g_1}\overline{X_i}+w_{1i}\middle| X,\left\{\underline{g_1}\overline{X_k}+w_{1k}\right\}_{k=1}^{M}\right).$$

This term would not have any gDoF since all the SNR exponents from $\underline{g_1} = \begin{bmatrix} g_{11} & g_{12} & . & g_{1M}\end{bmatrix}$ can be canceled due to availability of M linear equations in the conditioning. Let

$$X_{M\times M}=\begin{bmatrix} X_{11} & \dots & X_{1M} \\ \vdots & & \vdots \\ X_{M1} & \dots & X_{MM}\end{bmatrix},\ \underline{w_1}=\begin{bmatrix} w_{11} & \dots & w_{1M}\end{bmatrix}.$$

In the conditioning $\underline{g_1}X_{M\times M}+\underline{w_1}$ and $X_{M\times M}$ are available. Let $\mathrm{Adj}\left(X_{M\times M}\right)$ be the adjoint of $X_{M\times M}$ and $\det\left(X_{M\times M}\right)$ be the determinant of $X_{M\times M}$. Hence the term $\underline{g_1}\det\left(X_{M\times M}\right)\overline{X_i}+\underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}$ is available in the conditioning. The M linear equations in the conditioning can cancel off the gDoF contribution from $\underline{g_1}=\begin{bmatrix} g_{11} & g_{12} & . & g_{1M}\end{bmatrix}$ only if $\det\left(X_{M\times M}\right)$ is non-zero. Since X is Gaussian i.i.d., this is true almost surely. We handle this more precisely in the following steps:

$$h\left(\underline{g_1}\overline{X_i}+w_{1i}\middle| X,\left\{\underline{g_1}\overline{X_k}+w_{1k}\right\}_{k=1}^{M}\right) \quad \text{(A.118)}$$

$$\overset{(i)}{\leq} h\left(\underline{g_1}\overline{X_i}+w_{1i}\middle| X,\underline{g_1}\det\left(X_{M\times M}\right)\overline{X_i}+\underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}\right) \quad \text{(A.119)}$$

$$\overset{(ii)}{=} h\left(\underline{g_1}\det\left(X_{M\times M}\right)\overline{X_i}+\det\left(X_{M\times M}\right)w_{1i}\middle| X,\underline{g_1}\det\left(X_{M\times M}\right)\overline{X_i}+\underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}\right)$$
$$-\mathbb{E}\left[\log\left(\left|\det\left(X_{M\times M}\right)\right|\right)\right] \quad \text{(A.120)}$$

$$\stackrel{(iii)}{\leq} h\left(w_{1i}\det\left(X_{M\times M}\right) - \underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}\middle| X\right)$$

$$-\mathbb{E}\left[\log\left(\det\left(X_{M\times M}\right)\right)\right] \quad \text{(A.121)}$$

$$\stackrel{(iv)}{\dot{\leq}} h\left(w_{1i}\det\left(X_{M\times M}\right) - \underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}\middle| X\right) \quad \text{(A.122)}$$

$$\stackrel{(v)}{=} \log\left(\mathbb{E}\left[\left|\det\left(X_{M\times M}\right)\right|^2 + \left\|\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}\right\|^2\right]\right) + \log\left(\pi e\right) \quad \text{(A.123)}$$

$$\stackrel{(vi)}{\dot{\leq}} 0, \quad \text{(A.124)}$$

where (i) is using the availability of $\underline{g_1}\det\left(X_{M\times M}\right)\overline{X_i} + \underline{w_1}\mathrm{Adj}\left(X_{M\times M}\right)\overline{X_i}$ in conditioning and using the fact conditioning reduces entropy, (ii) is by multiplying with $\det\left(X_{M\times M}\right)$ and compensating with $-\mathbb{E}\left[\log\left(\left|\det\left(X_{M\times M}\right)\right|\right)\right]$ since $\det\left(X_{M\times M}\right)$ is known from the values in conditioning, (iii) is by subtracting the term available from conditioning and using the fact conditioning reduces entropy, (iv) is because $\mathbb{E}\left[\log\left(\left|\det\left(X_{M\times M}\right)\right|\right)\right]$ is finite by repeated application of Fact 5.1 and Tower property of expectation on Gaussian distributed X_{ij}, (v) is because $w_{1k} \sim \mathcal{CN}\left(0,1\right)$ i.i.d. and (vi) is because $X_{ij} \sim \mathcal{CN}\left(0,1\right)$ i.i.d.

Now by substituting (A.124), (A.117) and (A.114) in (A.113) we get the desired result.

A.6 Outer bound for MISO with $T < M$

Here we prove the gDoF outer bound given in Theorem 3.10 for the $M \times 1$ MISO system with $1 < T < M$. The steps follow similar to the case with $T \geq M$, given in (2.4.3). We have the structure of input distribution as $X = LQ$ with

$$L = \begin{bmatrix} x_{11} & 0 & 0 & \\ . & . & 0 & 0 \\ . & . & . & 0 \\ . & . & . & x_{TT} \\ . & . & . & . \\ x_{M1} & . & . & x_{MT} \end{bmatrix}. \quad \text{(A.125)}$$

For the channel we have, $G = \begin{bmatrix} g_{11} & . & . & g_{1M} \end{bmatrix}$, $g_{1i} \sim \mathcal{CN}(0, \rho_{1i}^2)$, $\rho_{1i}^2 = \mathsf{SNR}^{\gamma_{1i}}$, $Y = GX + W$, where W is $1 \times T$ with i.i.d. $\mathcal{CN}(0,1)$ components. We assume $\rho_{11}^2 \geq \rho_{1i}^2$ without loss of generality. Now note that WQ has the same distribution as W and is independent of Q (using the fact that W is isotropically distributed). Hence

$$Y = (GL + W) Q \tag{A.126}$$

$$= \begin{bmatrix} \left(w_{11} + \sum_{i=1}^{M} x_{i1} g_{1i}\right) & \left(w_{12} + \sum_{i=2}^{M} x_{i2} g_{1i}\right) & . & . & \left(w_{1T} + \sum_{i=T}^{M} x_{i2} g_{1i}\right) \end{bmatrix} Q. \tag{A.127}$$

Now using Lemma 3.2 on 66, we get

$$\begin{aligned} h(Y) &= h\left(\sum_{j=1}^{T} \left| w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i} \right|^2\right) \\ &\quad + (T-1)\,\mathbb{E}\left[\log\left(\sum_{j=1}^{T} \left| w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i} \right|^2\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right) && \text{(A.128)} \\ &\overset{(i)}{\leq} h\left(\sum_{j=1}^{T} \left| w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i} \right|^2\right) \\ &\quad + (T-1)\,\mathbb{E}\left[\log\left(\sum_{j=1}^{T} \left(1 + \sum_{i=j}^{M} |x_{ij}|^2 \rho_{ij}^2\right)\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right) && \text{(A.129)} \\ &\overset{(ii)}{\leq} h\left(\sum_{j=1}^{T} \left| w_{1j} + \sum_{i=j}^{M} x_{ij} g_{1i} \right|^2\right) \\ &\quad + (T-1)\,\mathbb{E}\left[\log\left(\sum_{i=1}^{M} \rho_{1i}^2 \left(\sum_{j=1}^{\min(i,T)} |x_{ij}|^2\right) + T\right)\right] + \log\left(\frac{\pi^T}{\Gamma(T)}\right), && \text{(A.130)} \end{aligned}$$

where (i) was using Tower property of expectation and Jensen's inequality and (ii) was using $\sum_{j=1}^{T} \sum_{i=j}^{M} |x_{ij}|^2 \rho_{1i}^2 = \sum_{i=1}^{M} \sum_{j=1}^{\min(i,T)} |x_{ij}|^2 \rho_{1i}^2$. Now using (3.52), we have

$$\begin{aligned} h(Y|X) &= \mathbb{E}\left[\log\left(\det\left(L^{\dagger} \mathrm{diag}\;\left(\rho_{11}^2, \ldots, \rho_{1M}^2\right) L + I_T\right)\right)\right] \\ &\quad + (T) \log(\pi e) && \text{(A.131)} \\ &= \mathbb{E}\left[\log\left(\prod_{i=1}^{M} (1 + \omega_i)\right)\right] + (T) \log(\pi e), && \text{(A.132)} \end{aligned}$$

where ω_i are the eigenvalues of $L^{\dagger}\text{diag}\ \left(\rho_{11}^2,\ldots,\rho_{1M}^2\right)L$. Hence

$$h\left(Y|X\right)=\mathbb{E}\left[\log\left(\prod_{i=1}^{M}\left(1+\omega_i\right)\right)\right]+\left(T\right)\log\left(\pi e\right) \tag{A.133}$$

$$\geq\mathbb{E}\left[\log\left(1+\sum\omega_i\right)\right]+\left(T\right)\log\left(\pi e\right). \tag{A.134}$$

The last step is true because $\omega_i\geq 0$. Now

$$\begin{aligned}\sum\omega_i&=\text{Trace}\left(L^{\dagger}\text{diag}\ \left(\rho_{11}^2,\ldots,\rho_{1M}^2\right)L\right)\\&=\text{Trace}\left(\text{diag}\ \left(\rho_{11}^2,\ldots,\rho_{1M}^2\right)LL^{\dagger}\right)\\&=\sum_{i=1}^{M}\rho_{1i}^2\left(\sum_{j=1}^{\min(i,T)}\left|x_{ij}\right|^2\right).\end{aligned} \tag{A.135}$$

Hence

$$h\left(Y|X\right)\geq\mathbb{E}\left[\log\left(1+\sum_{i=1}^{M}\rho_{1i}^2\left(\sum_{j=1}^{\min(i,T)}\left|x_{ij}\right|^2\right)\right)\right]+\left(T\right)\log\left(\pi e\right). \tag{A.136}$$

Hence

$$\begin{aligned}I\left(X;Y\right)\leq h&\left(\sum_{j=1}^{M}\left|w_{1j}+\sum_{i=j}^{M}x_{ij}g_{1i}\right|^2+\sum_{i=M+1}^{T}\left|w_{1i}\right|^2\right)\\&+\left(T-2\right)\mathbb{E}\left[\log\left(\sum_{i=1}^{M}\rho_{1i}^2\left(\sum_{j=1}^{i}\left|x_{ij}\right|^2\right)+T\right)\right]\\&+\log\left(\frac{\pi^T}{\Gamma\left(T\right)}\right)-T\log\left(\pi e\right)\end{aligned} \tag{A.137}$$

$$\dot{\leq}\left(T-1\right)\log\left(\sum_{i=1}^{M}\rho_{1i}^2MT+T\right), \tag{A.138}$$

where the last step was using maximum entropy result and Jensen's inequality. Hence

$$\underset{\mathsf{SNR}\to\infty}{\text{limsup}}\frac{I\left(X;Y\right)}{\log\left(\mathsf{SNR}\right)}\leq\left(T-1\right)\gamma_{11}. \tag{A.139}$$

A.7 Proof of Lemma 3.3

Here we prove that $h\left(\left|ag_{11}+bg_{12}+w_{11}\right|^2+\left|cg_{12}+w_{12}\right|^2+\sum_{i=3}^{T}\left|w_{1i}\right|^2\right)$ and

$$\mathbb{E}\left[\log\left(\left|a\right|^2\rho_{11}^2+\left(\left|b\right|^2+\left|c\right|^2\right)\rho_{12}^2+1\right)\right]$$

have the same gDoF. For this, consider the point to point channel

$$\mathcal{C}_1 : V = |ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \tag{A.140}$$

with inputs a, b, c and power constraint T. Its capacity is given by

$$\begin{aligned} C_1 = \max_{p(a,b,c);\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T} \Bigg\{ & h\left(\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)\right) \\ & -h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right) \Bigg\} \end{aligned} \tag{A.141}$$

From [LM03, (32)] we have

$$\begin{aligned} I(U; V) \leq \mathbb{E}[\log(V)] - h(V|U) + \log(\Gamma(\alpha)) \\ + \alpha(1 + \log(\mathbb{E}[V]) - \mathbb{E}[\log(V)]) - \alpha\log(\alpha) \end{aligned} \tag{A.142}$$

for any $\alpha > 0$ for channels whose output V takes values in $\mathbb{R}^+$. We will use this result to bound $I(U; V)$ for any input distribution $p(a, b, c); \mathbb{E}\left[|a|^2 + |b|^2 + |c|^2\right] \leq T$ for the channel $\mathcal{C}_1$ with $U = (a, b, c)$ as input. Now

$$\begin{aligned} & h(V|U) \\ & = h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right) && \text{(A.143)} \\ & \overset{(i)}{\leq} \mathbb{E}\left[\log\left(e\mathbb{E}\left[|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right]\right)\right] && \text{(A.144)} \\ & \overset{(ii)}{=} \mathbb{E}\left[\log\left(e\left(\left(\rho_{11}^2|a|^2 + \rho_{12}^2|b|^2 + 1\right) + \left(\rho_{12}^2|c|^2 + 1\right) + (T-2)\right)\right)\right] && \text{(A.145)} \\ & = \mathbb{E}\left[\log\left(\rho_{11}^2|a|^2 + \rho_{12}^2|b|^2 + \rho_{12}^2|c|^2 + T\right)\right] + \log(e), && \text{(A.146)} \end{aligned}$$

where (i) was using the definition of conditional entropy and the fact that exponential distribution has the maximum entropy among positive random variable with a given mean, (ii) is using the fact that given (a, b, c), $ag_{11} + bg_{12} + w_{11}, cg_{12} + w_{12}$ are sums of independent Gaussians. Note

$$\mathbb{E}[\log(V)] = \mathbb{E}\left[\mathbb{E}\left[\log\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right)\right]\right]$$

$$\leq \mathbb{E}\left[\log\left(\rho_{11}^2 |a|^2 + \rho_{12}^2 |b|^2 + \rho_{12}^2 |c|^2 + T\right)\right] \tag{A.147}$$

using Jensen's inequality. Also

$$\begin{aligned}\mathbb{E}[\log(V)] &= \mathbb{E}\left[\mathbb{E}\left[\log\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right)\right]\right] \\ &\geq \mathbb{E}\left[\log\left(\rho_{11}^2 |a|^2 + \rho_{12}^2 |b|^2 + \rho_{12}^2 |c|^2 + T\right)\right] - 3\gamma \log(e) \end{aligned} \tag{A.148}$$

by using using Fact 5.1 on page 135 for exponentially distributed $|ag_{rd2} + bg_{rd1} + w_{d1}|^2$, $|cg_{rd1} + w_{d2}|^2$ (for given a, b, c) and Fact 3.3 for chi-squared distributed $\sum_{i=3}^{T} |w_{di}|^2$.

Claim A.2. $\mathbb{E}[\log(V)] - h(V|U) \leq \log(3+T) + \frac{7}{2}\log(e)$ *independent of* SNR

Proof. It suffices to show that for any constant (a', b', c'), $\mathbb{E}\left[\log(V) \middle| U = (a', b', c')\right] - h\left(V \middle| U = (a', b', c')\right) \leq \log(3+T) + \frac{7}{2}\log(e)$ independent of a', b', c' and SNR

$$\mathbb{E}\left[\log(V) \middle| U = (a', b', c')\right] - h\left(V \middle| U = (a', b', c')\right)$$

$$\begin{aligned} &= \mathbb{E}\left[\log\left(|a'g_{11} + b'g_{12} + w_{11}|^2 + |c'g_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)\right] \\ &\quad - h\left(|a'g_{11} + b'g_{12} + w_{11}|^2 + |c'g_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right) && \text{(A.149)} \\ &\overset{(i)}{\leq} \log\left(\mathbb{E}\left[|a'g_{11} + b'g_{12} + w_{11}|^2 + |c'g_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right]\right) \\ &\quad - h\left(|a'g_{11} + b'g_{12} + w_{11}|^2 + |c'g_{12} + w_{12}|^2\right) && \text{(A.150)} \\ &\overset{(ii)}{=} \log\left(\rho_{11}^2 |a'|^2 + \rho_{12}^2 |b'|^2 + \rho_{12}^2 |c'|^2 + T\right) \\ &\quad - h\left(|a'\rho_{11}\eta_{11} + b'\rho_{12}\eta_{12} + w_{11}|^2 + |c'\rho_{12}\eta_{12} + w_{12}|^2\right), && \text{(A.151)} \end{aligned}$$

where (i) is using Jensen's inequality and conditioning reduces entropy to remove $\sum_{i=3}^{T} |w_{di}|^2$ in the negative term; (ii) is using the fact that $a'g_{11} + b'g_{12} + w_{11}, c'g_{12} + w_{12}$ are sums of independent Gaussians. We introduced $\eta_{ij} \sim \mathcal{CN}(0,1)$ so that $g_{ij} = \rho_{ij}\eta_{ij}$.

Consider the case when $1 \leq \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$. Assume $1 \leq |b\rho_{12}| = \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$

$$\mathbb{E}\left[\log(V)\middle| U = (a', b', c')\right] - h\left(V\middle| U = (a', b', c')\right)$$

$$\begin{aligned}
&\leq \log\left(\rho_{11}^2 |a'|^2 + \rho_{12}^2 |b'|^2 + \rho_{12}^2 |c'|^2 + T\right) - \log\left(\rho_{12}^2 |b'|^2\right) \\
&\quad - h\left(\left|\frac{a'\rho_{11}}{b'\rho_{12}}\eta_{11} + \eta_{12} + \frac{w_{11}}{b'\rho_{12}}\right|^2 + \left|\frac{c'}{b'}\eta_{12} + \frac{w_{12}}{b'\rho_{12}}\right|^2\right). \qquad \text{(A.152)}
\end{aligned}$$

Now now using the result from Appendix A.9 to lower bound entropy of sum of norm-squared of Gaussian vectors, we have

$$\begin{aligned}
&h\left(\left|\frac{a'\rho_{11}}{b'\rho_{12}}\eta_{11} + \eta_{12} + \frac{w_{11}}{b'\rho_{12}}\right|^2 + \left|\frac{c'}{b'}\eta_{12} + \frac{w_{12}}{b'\rho_{12}}\right|^2\right) \\
&\geq h\left(\left|\frac{a'\rho_{11}}{b'\rho_{12}}\eta_{11} + \eta_{12} + \frac{w_{11}}{b'\rho_{12}}\right|^2 + \left|\frac{c'}{b'}\eta_{12} + \frac{w_{12}}{b'\rho_{12}}\right|^2 \middle| \eta_{11}\right) \\
&\geq -\frac{7}{2}\log(e). \qquad \text{(A.153)}
\end{aligned}$$

Hence we get

$$\mathbb{E}\left[\log(V)\middle| U = (a', b', c')\right] - h\left(V\middle| U = (a', b', c')\right)$$

$$\begin{aligned}
&\leq \log\left(\rho_{11}^2 |a'|^2 + \rho_{12}^2 |b'|^2 + \rho_{12}^2 |c'|^2 + T\right) - \log\left(\rho_{12}^2 |b'|^2\right) + \frac{7}{2}\log(e) \qquad \text{(A.154)} \\
&\overset{(i)}{\leq} \log(3 + T) + \frac{7}{2}\log(e), \qquad \text{(A.155)}
\end{aligned}$$

where in step (i) we used $1 \leq |b\rho_{12}| \leq \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$.

Similarly for other cases $1 \leq |a'\rho_{11}| = \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$ and $1 \leq |c'\rho_{12}| = \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$, we can show that $\mathbb{E}\left[\log(V)\middle| U = (a', b', c')\right] - h\left(V\middle| U = (a', b', c')\right)$ is upper bounded by $\log(3 + T) + \frac{7}{2}\log(e)$.

Now if $1 > \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$

$$\mathbb{E}\left[\log(V)\middle| U = (a', b', c')\right] - h\left(V\middle| U = (a', b', c')\right)$$

$$\begin{aligned}
&\leq \log\left(\rho_{11}^2 |a'|^2 + \rho_{12}^2 |b'|^2 + \rho_{12}^2 |c'|^2 + T\right) \\
&\qquad - h\left(|a'\rho_{11}\eta_{11} + b'\rho_{12}\eta_{12} + w_{11}|^2 + |c'\rho_{12}\eta_{12} + w_{12}|^2\right) && \text{(A.156)} \\
&\overset{(i)}{\leq} \log(3+T) - h\left(|a'\rho_{11}\eta_{11} + b'\rho_{12}\eta_{12} + w_{11}|^2 + |c'\rho_{12}\eta_{12} + w_{12}|^2 \middle| \eta_{11}, \eta_{12}, w_{12}\right) && \text{(A.157)} \\
&= \log(3+T) - h\left(|a'\rho_{11}\eta_{11} + b'\rho_{12}\eta_{12} + w_{11}|^2 \middle| \eta_{11}, \eta_{12}\right) && \text{(A.158)} \\
&\overset{(ii)}{\leq} \log(3+T) + \frac{7}{2}\log(e), && \text{(A.159)}
\end{aligned}$$

where in step (i) we used the fact $1 > \max\left(|a'\rho_{11}|, |b'\rho_{12}|, |c'\rho_{12}|\right)$ and conditioning reduces entropy, in step (ii) we used the result from Appendix A.9 to lower bound $h\left(|a'\rho_{11}\eta_{11} + b'\rho_{12}\eta_{12} + w_{11}|^2\middle| \eta_{11}, \eta_{12}\right)$. □

Using (A.146), (A.147), (A.148) and using the Claim A.2 we get

$$\begin{aligned}
\mathbb{E}[\log(V)] &\doteq \mathbb{E}\left[\log\left(\rho_{11}^2 |a|^2 + \rho_{12}^2 |b|^2 + \rho_{12}^2 |c|^2 + T\right)\right] \\
&\doteq h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2 \middle| a, b, c\right) = h\left(V\middle| U\right) && \text{(A.160)}
\end{aligned}$$

and the above approximation is tight within a constant independent of SNR. Hence it follows that

$$\begin{aligned}
C_1 \doteq \max_{p(a,b,c);\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T} &\left\{ h\left(|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2\right)\right. \\
&\left. -\mathbb{E}\left[\log\left(\rho_{11}^2 |a|^2 + \rho_{12}^2 |b|^2 + \rho_{12}^2 |c|^2 + T\right)\right]\right\} && \text{(A.161)}
\end{aligned}$$

and the above inequality is tight within a constant independent of SNR. Now we shall prove that

$$\limsup_{\mathsf{SNR}\to\infty} C_1(\mathsf{SNR}) - \log(\log(\mathsf{SNR})) < \infty \qquad \text{(A.162)}$$

and hence it will prove our claim that $h\left(|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right)$ and $\mathbb{E}\left[\log\left(\rho_{11}^2|a|^2+\rho_{12}^2|b|^2+\rho_{12}^2|c|^2+T\right)\right]$ have the same gDoF.

Now looking at (A.142) again, if the term $\log(\mathbb{E}[V])-\mathbb{E}[\log(V)]$ does not approach infinity with the SNR then the result follows directly by choosing any fixed $\alpha>0$. When $\log(\mathbb{E}[V])-\mathbb{E}[\log(V)]$ does tend to infinity with SNR, we choose

$$\alpha^*=\left(1+\log(\mathbb{E}[V])-\mathbb{E}[\log(V)]\right)^{-1} \tag{A.163}$$

with $\alpha^*\downarrow 0$ with the SNR and we have $\log(\Gamma(\alpha^*))=\log\left(\frac{1}{\alpha^*}\right)+o(1)$ and $\alpha^*\log(\alpha^*)=o(1)$ where $o(1)$ tends to zero as α^* tends to zero, following [LM03, (337)]. Hence we have

$$C_1\leq r(T)+1+\log\left(\frac{1}{\alpha^*}\right)+o(1) \tag{A.164}$$

$$\begin{aligned}
\frac{1}{\alpha^*}&=1+\log(\mathbb{E}[V])-\mathbb{E}[\log(V)] && \text{(A.165)}\\
&=1+\log\left(\mathbb{E}\left[|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right]\right)\\
&\quad-\mathbb{E}\left[\log\left(|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right)\right] && \text{(A.166)}\\
&\overset{(i)}{=}1+\log\left(\mathbb{E}\left[\rho_{11}^2|a|^2+\rho_{12}^2|b|^2+\rho_{12}^2|c|^2+T\right]\right)\\
&\quad-\mathbb{E}\left[\log\left(|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right)\right] && \text{(A.167)}\\
&\overset{(ii)}{\leq}1+\log\left(\rho_{11}^2T+\rho_{12}^2T+T\right) && \text{(A.168)}\\
&\quad-\mathbb{E}\left[\log\left(|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2\right)\right] && \text{(A.169)}\\
&\overset{(iii)}{\leq}1+\log\left(\rho_{11}^2+\rho_{12}^2+1\right)+\log(T)-\mathbb{E}\left[\mathbb{E}\left[\log\left(|ag_{11}+bg_{12}+w_{11}|^2+0\right)\middle|\,a,b\right]\right] && \text{(A.170)}\\
&\overset{(iv)}{\leq}1+\log\left(\rho_{11}^2+\rho_{12}^2+1\right)+\log(T)-\mathbb{E}\left[\log\left(\rho_{11}^2|a|^2+\rho_{12}^2|b|^2+1\right)\right]+\gamma\log(e) && \text{(A.171)}\\
&\overset{(v)}{\leq}1+\log\left(\rho_{11}^2+\rho_{12}^2+1\right)+\log(T)-0+\gamma\log(e), && \text{(A.172)}
\end{aligned}$$

where (i) is using tower property of expectation and that given (a,b,c), $ag_{11}+bg_{12}+w_{11}, cg_{12}+w_{12}$ are sums of independent Gaussians, (ii) is using power constraints on a,b,c, (iii) is because $|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2>0$, (iv) is using Fact 5.1 on page 135 and (v) is because $\log\left(\rho_{11}^2|a|^2+\rho_{12}^2|b|^2+1\right)>0$. Hence

$$C_1 \leq r(T)+1+\log\left(1+\log\left(\rho_{11}^2+\rho_{12}^2+\rho_{12}^2+1\right)+\log(T)+\gamma\log(e)\right)+o(1) \quad \text{(A.173)}$$

and the proof is complete.

A.8 Proof of Lemma 2.5

Here we prove that $h\left(|\xi_{22}|^2\middle||\xi_{11}|^2\right) \doteq h\left(|\xi_{22}|^2\middle||\xi_{11}|^2,a,b,c\right)$ with $|\xi_{11}|^2,|\xi_{22}|^2$ defined in (2.44),2.46. For this proof, we just need to show that $I\left(|\xi_{22}|^2;a,b,c\middle||\xi_{11}|^2\right)$ has zero gDoF. Now

$$I\left(|\xi_{22}|^2;a,b,c\middle||\xi_{11}|^2\right) \leq I\left(|\xi_{22}|^2;a,b,c,|\xi_{11}|^2\right).$$

We will show that $I\left(|\xi_{22}|^2;a,b,c,|\xi_{11}|^2\right)$ has no gDoF. From [LM03, (32)] we have

$$\begin{aligned} I(U;V) \leq\ & \mathbb{E}[\log(V)]-h(V|U)+\log(\Gamma(\alpha)) \\ & +\alpha\left(1+\log(\mathbb{E}[V])-\mathbb{E}[\log(V)]\right)-\alpha\log(\alpha) \end{aligned} \quad \text{(A.174)}$$

for any $\alpha>0$ for channels whose output V takes values in $\mathbb{R}^+$. We will use this result to bound $I\left(|\xi_{22}|^2;a,b,c,|\xi_{11}|^2\right)$ with $U=\left(a,b,c,|\xi_{11}|^2\right), V=|\xi_{22}|^2$ for any distribution of a,b,c with the power constraint $\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq T$. The result from [LM03] can be applied assuming the channel induced by $p\left(|\xi_{22}|^2\middle|a,b,c,|\xi_{11}|^2\right)$ satisfies the Borel measurability conditions in [LM03, (Theorem 5.1)], i.e for any given Borel set $\mathcal{B}\subset\mathbb{R}^+$, $f_{\mathcal{B}}(v)=p\left(\mathcal{B}\middle|v=\left(a,b,c,|\xi_{11}|^2\right)\right)$ is a Borel measurable function.

Recall that from (2.44) and (2.46), we have

$$|\xi_{11}|^2=|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2+\sum_{i=3}^{T}|w_{1i}|^2,$$

$$|\xi_{22}|^2=|ag_{21}+bg_{22}+w_{21}|^2+|cg_{22}+w_{22}|^2+\sum_{i=3}^{T}|w_{2i}|^2$$

$$-\frac{\left|\left(ag_{21}+bg_{22}+w_{21}\right)\left(ag_{11}+bg_{12}+w_{11}\right)^*+\left(cg_{22}+w_{22}\right)\left(cg_{12}+w_{12}\right)^*+\sum_{i=3}^{T}w_{2i}w_{1i}^*\right|^2}{\left|ag_{11}+bg_{12}+w_{11}\right|^2+\left|cg_{12}+w_{12}\right|^2+\sum_{i=3}^{T}\left|w_{1i}\right|^2}.$$

We first consider $\log\left(\mathbb{E}\left[V\right]\right)-\mathbb{E}\left[\log\left(V\right)\right]=\mathbb{E}\left[\log\left(\left|\xi_{22}\right|^2\right)\right]-h\left(\left|\xi_{22}\right|^2\middle|a,b,c,\left|\xi_{11}\right|^2\right)$ and show that it is bounded independent of SNR. Note that we can manipulate $\left|\xi_{22}\right|^2$ as

$$\begin{aligned}
&\left|\xi_{22}\right|^2\\
&=\left(\left|ag_{21}+bg_{22}+w_{21}\right|^2+\left|cg_{22}+w_{22}\right|^2+\sum_{i=3}^{T}\left|w_{2i}\right|^2\right)\\
&\qquad-\left|\left(ag_{21}+bg_{22}+w_{21}\right)u_1^*+\left(cg_{22}+w_{22}\right)u_2^*+\sum_{i=3}^{T}w_{2i}u_i^*\right|^2 &\text{(A.175)}\\
&=\left\|\left[ag_{21}+bg_{22}+w_{21},cg_{22}+w_{22},w_{23},\ldots,w_{2T}\right]\right.\\
&\qquad\left.-\left[ag_{21}+bg_{22}+w_{21},cg_{22}+w_{22},w_{23},\ldots,w_{2T}\right]\begin{bmatrix}u_1^*\\u_2^*\\.\\.\\u_T^*\end{bmatrix}\left[u_1,\ldots,u_T\right]\right\|^2, &\text{(A.176)}
\end{aligned}$$

where $\|\cdot\|$ indicates 2-norm for a vector and (u_i) forms a unit norm complex vector

$$\left[u_1,\ldots,u_T\right]=\frac{\left[ag_{11}+bg_{12}+w_{11},cg_{12}+w_{12},w_{13},\ldots,w_{1T}\right]}{\left|ag_{11}+bg_{12}+w_{11}\right|^2+\left|cg_{12}+w_{12}\right|^2+\sum_{i=3}^{T}\left|w_{1i}\right|^2},\tag{A.177}$$

$$\begin{aligned}
\left|\xi_{22}\right|^2=&\left|ag_{21}+bg_{22}+w_{21}-\left(\left(ag_{21}+bg_{22}+w_{21}\right)u_1^*+\left(cg_{22}+w_{22}\right)u_2^*+\sum_{i=3}^{T}w_{2i}u_i^*\right)u_1\right|^2\\
&+\left|cg_{22}+w_{22}-\left(\left(ag_{21}+bg_{22}+w_{21}\right)u_1^*+\left(cg_{22}+w_{22}\right)u_2^*+\sum_{i=3}^{T}w_{2i}u_i^*\right)u_2\right|\\
&+\left|w_{23}-\left(\left(ag_{21}+bg_{22}+w_{21}\right)u_1^*+\left(cg_{22}+w_{22}\right)u_2^*+\sum_{i=3}^{T}w_{2i}u_i^*\right)u_3\right|^2
\end{aligned}$$

$$+\dots$$

$$+\left|w_{2T}-\left(\left(ag_{21}+bg_{22}+w_{21}\right)u_1^*+\left(cg_{22}+w_{22}\right)u_2^*+\sum_{i=3}^{T}w_{2i}u_i^*\right)u_T\right|^2 \tag{A.178}$$

$$\overset{(i)}{=}\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2, \tag{A.179}$$

where in step (i) η_{ij} are independent $\mathcal{CN}(0,1)$ after the substitution $g_{ij}=\rho_{ij}\eta_{ij}$, (κ_{ij}) are functions of a,b,c,ρ_{ij},u_i obtained after collecting the coefficients of η_{ij},w_{2j}; Note that $\max_{i,j}\left(|\kappa_{ij}|\right)\geq 1$. Now

$$\begin{aligned}
&\mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right]-h\left(|\xi_{22}|^2\,\middle|\,a,b,c,|\xi_{11}|^2\right)\\
&=\mathbb{E}\left[\log\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2\right)\right]\\
&\quad-h\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2\,\middle|\,a,b,c,|\xi_{11}|^2\right) &\text{(A.180)}\\
&\leq\mathbb{E}\left[\log\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2\right)\right]\\
&\quad-h\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2\,\middle|\,\{\kappa_{ij}\}_{i,j\leq T}\right), &\text{(A.181)}
\end{aligned}$$

where the last step uses the fact that conditioning reduces entropy and Markovity $\left(a,b,c,|\xi_{11}|^2\right)$ — $\left(\{\kappa_{ij}\}_{i,j\leq T}\right)$—$\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa_{1i}+\eta_{22}\kappa_{2i}+\sum_{j=1}^{T}w_{2j}\kappa_{(j+2)i}\right|^2\right)$. Note that $\eta_{21},\eta_{22},w_{2j}$ are independent of κ_{ij}. Now it suffices to show that for any given set of *constant* κ'_{ij} the difference

$$\mathbb{E}\left[\log\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa'_{1i}+\eta_{22}\kappa'_{2i}+\sum_{j=1}^{T}w_{2j}\kappa'_{(j+2)i}\right|^2\right)\right]-h\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa'_{1i}+\eta_{22}\kappa'_{2i}+\sum_{j=1}^{T}w_{2j}\kappa'_{(j+2)i}\right|^2\right)$$

is uniformly bounded independent of κ'_{ij}. We will show this by assuming $|\kappa'_{11}|=\max_{i,j}\left(\left|\kappa'_{ij}\right|\right)$. This is without loss of generality since η_{ij},w_{ij} are all i.i.d. $\mathcal{CN}(0,1)$. Now

$$\mathbb{E}\left[\log\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa'_{1i}+\eta_{22}\kappa'_{2i}+\sum_{j=1}^{T}w_{2j}\kappa'_{(j+2)i}\right|^2\right)\right]-h\left(\sum_{i=1}^{T}\left|\eta_{21}\kappa'_{1i}+\eta_{22}\kappa'_{2i}+\sum_{j=1}^{T}w_{2j}\kappa'_{(j+2)i}\right|^2\right)$$

$$= \mathbb{E}\left[\log\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right)\right]$$
$$- h\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right) \tag{A.182}$$
$$\overset{(i)}{\leq} \log\left(\sum_{i=1}^{T}\mathbb{E}\left[\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right]\right)$$
$$- h\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right) \tag{A.183}$$
$$\overset{(ii)}{=} \log\left(\sum_{i=1}^{T}\left(\left|\frac{\kappa'_{1i}}{\kappa'_{11}}\right|^2+\left|\frac{\kappa'_{2i}}{\kappa'_{11}}\right|^2+\sum_{j=1}^{T}\left|\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right)\right)$$
$$- h\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right) \tag{A.184}$$
$$\overset{(iii)}{\leq} \log\left(T\left(T+2\right)\right) - h\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\right) \tag{A.185}$$
$$\overset{(iv)}{\leq} \log\left(T\left(T+2\right)\right) - h\left(\sum_{i=1}^{T}\left|\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}\right|^2\middle|\, \eta_{22}, w_{2j}\right) \tag{A.186}$$
$$\overset{(v)}{\leq} \log\left(T\left(T+2\right)\right) + \frac{7}{2}\log\left(e\right), \tag{A.187}$$

where (i) is using Jensen's inequality, (ii) is using the fact that $\eta_{21}\frac{\kappa'_{1i}}{\kappa'_{11}}+\eta_{22}\frac{\kappa'_{2i}}{\kappa'_{11}}+\sum_{j=1}^{T}w_{2j}\frac{\kappa'_{(j+2)i}}{\kappa'_{11}}$ is Complex Gaussian, (iii) is because $\frac{\left|\kappa'_{ij}\right|}{\left|\kappa'_{11}\right|} \leq 1$ since $\left|\kappa'_{11}\right| = \max_{i,j}\left(\left|\kappa'_{ij}\right|\right)$ (note that $\max_{i,j}\left(\left|\kappa'_{ij}\right|\right) \geq 1$ for a valid set of κ'_{ij}, due to the way κ_{ij} is defined), (iv) is because conditioning reduces entropy and (v) is by invoking the result from Appendix A.9.

Now if the term $\log\left(\mathbb{E}\left[V\right]\right)-\mathbb{E}\left[\log\left(V\right)\right]$ does not approach infinity with the SNR then the desired result follows directly by choosing any fixed $\alpha > 0$. When $\log\left(\mathbb{E}\left[V\right]\right) - \mathbb{E}\left[\log\left(V\right)\right]$ does tend to infinity with SNR, following [LM03, (336)] we choose

$$\alpha^* = \left(1 + \log\left(\mathbb{E}\left[V\right]\right) - \mathbb{E}\left[\log\left(V\right)\right]\right)^{-1} \tag{A.188}$$

with $\alpha^* \downarrow 0$ with the SNR and we have $\log\left(\Gamma\left(\alpha^*\right)\right) = \log\left(\frac{1}{\alpha^*}\right) + o\left(1\right)$ and $\alpha^* \log\left(\alpha^*\right) = o\left(1\right)$

where $o(1)$ tends to zero as α^* tends to zero, following [LM03, (337)]. Hence we have

$$I\left(|\xi_{22}|^2; a, b, c, |\xi_{11}|^2\right) \leq \left(\log\left(T\left(T+2\right)\right) + \frac{7}{2}\log\left(e\right)\right) + 1 + \log\left(\frac{1}{\alpha^*}\right) + o\left(1\right), \quad \text{(A.189)}$$

$$\frac{1}{\alpha^*} = 1 + \log\left(\mathbb{E}\left[|\xi_{22}|^2\right]\right) - \mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right]. \quad \text{(A.190)}$$

Now

$$|\xi_{22}|^2 \leq |ag_{21} + bg_{22} + w_{21}|^2 + |cg_{22} + w_{22}|^2 + \sum_{i=3}^{T} |w_{2i}|^2. \quad \text{(A.191)}$$

Hence

$$\mathbb{E}\left[|\xi_{22}|^2\right] \overset{(i)}{\leq} \mathbb{E}\left[\left(\rho_{21}^2 |a|^2 + \rho_{22}^2\left(|b|^2 + |c|^2\right) + T\right)\right], \quad \text{(A.192)}$$

$$\log\left(\mathbb{E}\left[|\xi_{22}|^2\right]\right) \overset{(ii)}{\leq} \log\left(\rho_{21}^2 + \rho_{22}^2 + 1\right) + \log\left(T\right), \quad \text{(A.193)}$$

where (i) is using the fact that given (a, b, c), $ag_{21} + bg_{22} + w_{21}$,$cg_{22} + w_{22}$ are sums of independent Gaussians and (ii) is using the power constraint on a, b, c. Hence we have

$$\frac{1}{\alpha^*} \leq 1 + \log\left(\rho_{21}^2 + \rho_{22}^2 + 1\right) + \log\left(T\right) - \mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right]. \quad \text{(A.194)}$$

Now we lower bound $\mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right]$. Note that

$$|\xi_{22}|^2 = |ag_{21} + bg_{22} + w_{21}|^2 + |cg_{22} + w_{22}|^2 + \sum_{i=3}^{T} |w_{2i}|^2 \quad \text{(A.195)}$$

$$- \frac{\left|\left(ag_{21} + bg_{22} + w_{21}\right)\left(ag_{11} + bg_{12} + w_{11}\right)^* + \left(cg_{22} + w_{22}\right)\left(cg_{12} + w_{12}\right)^* + \sum_{i=3}^{T} w_{2i} w_{1i}^*\right|^2}{|ag_{11} + bg_{12} + w_{11}|^2 + |cg_{12} + w_{12}|^2 + \sum_{i=3}^{T} |w_{1i}|^2} \quad \text{(A.196)}$$

is the magnitude squared of the projection of the Complex vector

$$\left[\left(ag_{21} + bg_{22} + w_{21}\right), \left(cg_{22} + w_{22}\right), w_{23}, \ldots, w_{2T}\right]$$

onto the subspace orthogonal to the Complex vector

$$\left[\left(ag_{11} + bg_{12} + w_{11}\right), \left(cg_{12} + w_{12}\right), w_{13}, \ldots, w_{1T}\right].$$

Note that $\left[(cg_{12}+w_{12})^*, -(ag_{11}+bg_{12}+w_{11})^*, 0, \ldots, 0\right]$ is orthogonal to

$$\left[(ag_{11}+bg_{12}+w_{11}), (cg_{12}+w_{12}), w_{13}, \ldots, w_{1T}\right].$$

Hence

$$|\xi_{22}|^2 \geq \frac{\left|\begin{bmatrix} ag_{21}+bg_{22}+w_{21}, & cg_{22}+w_{22}, & w_{23}, & \ldots, & w_{2T} \end{bmatrix}\begin{bmatrix} cg_{12}+w_{12} \\ -(ag_{11}+bg_{12}+w_{11}) \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right|^2}{|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2} \tag{A.197}$$

$$= \frac{|(ag_{21}+bg_{22}+w_{21})(cg_{12}+w_{12})-(cg_{22}+w_{22})(ag_{11}+bg_{12}+w_{11})|^2}{|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2} \tag{A.198}$$

and hence

$$\begin{aligned}
&\mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right] && \text{(A.199)}\\
&\geq \mathbb{E}\left[\log\left(\frac{|(ag_{21}+bg_{22}+w_{21})(cg_{12}+w_{12})-(cg_{22}+w_{22})(ag_{11}+bg_{12}+w_{11})|^2}{|ag_{11}+bg_{12}+w_{11}|^2+|cg_{12}+w_{12}|^2}\right)\right]\\
&= \mathbb{E}\left[\log\left(|(ag_{21}+bg_{22}+w_{21})u_1-(cg_{22}+w_{22})u_2|^2\right)\right], && \text{(A.200)}
\end{aligned}$$

where (u_1, u_2) is a unit norm complex vector independent of g_{2i}, w_{2i}. Hence

$$\begin{aligned}
&\mathbb{E}\left[\log\left(|\xi_{22}|^2\right)\right]\\
&\stackrel{(i)}{\geq} \mathbb{E}\left[\log\left(|au_1|^2\rho_{21}^2+|bu_1-cu_2|^2\rho_{22}^2+|u_1|^2+|u_2|^2\right)\right]-\gamma\log(e) && \text{(A.201)}\\
&\stackrel{(ii)}{\geq} \mathbb{E}\left[\log\left(|au_1|^2\rho_{21}^2+|bu_1-cu_2|^2\rho_{22}^2+1\right)\right]-\gamma\log(e) && \text{(A.202)}\\
&\geq -\gamma\log(e), && \text{(A.203)}
\end{aligned}$$

where (i) is using the facts that given (a, b, c, u_1, u_2), $(ag_{21}+bg_{22}+w_{21})u_1-(cg_{22}+w_{22})u_2$ is Complex Gaussian distributed with variance $|au_1|^2\rho_{21}^2+|bu_1-cu_2|^2\rho_{22}^2+|u_1|^2+|u_2|^2$ and applying Fact 5.1 on page 135 together with Tower property of expectation. The step (ii) is because (u_1, u_2) is a unit norm vector.

Substituting (A.203) in (A.194) we get

$$\frac{1}{\alpha^*} \le \log\left(\rho_{21}^2 + \rho_{22}^2 + 1\right) + 1 + \log(T) + \gamma \log(e) \tag{A.204}$$

$$= \log\left(\rho_{21}^2 + \rho_{22}^2 + 1\right) + r_2(T) \tag{A.205}$$

and hence by substituting the above in (A.189), we get

$$I\left(|\xi_{22}|^2; a, b, c, |\xi_{11}|^2\right) \le \left(\log\left(T\left(T+2\right)\right) + \frac{7}{2}\log(e)\right) + 1 + \log\left(\log\left(\rho_{21}^2 + \rho_{22}^2 + 1\right) + r_2(T)\right) + o(1), \tag{A.206}$$

where $r_2(T) = 1 + \log(T) + \gamma \log(e)$ is a function of T alone. Hence $I\left(|\xi_{22}|^2; a, b, c, |\xi_{11}|^2\right)$ has zero gDoF. Now since

$$I\left(|\xi_{22}|^2; a, b, c \middle| |\xi_{11}|^2\right) \le I\left(|\xi_{22}|^2; a, b, c, |\xi_{11}|^2\right)$$

it follows that $h\left(|\xi_{22}|^2 \middle| |\xi_{11}|^2\right) \doteq h\left(|\xi_{22}|^2 \middle| |\xi_{11}|^2, a, b, c\right)$.

Also $h\left(|\xi_{22}|^2 \middle| |\xi_{11}|^2, a, b, c\right) \le \mathbb{E}\left[\log\left(e\mathbb{E}\left[|\xi_{22}|^2 \middle| a, b, c\right]\right)\right]$ using the maximum entropy result for positive random variables with given mean.

A.9 A lower bound on entropy of squared 2-norm of a Gaussian vector

For complex l_i, k_i, l for finite number of i's with $|k_i| \le 1$ and $\eta \sim \mathcal{CN}(0,1)$ we will show that

$$h\left(|\eta + l|^2 + \sum_i |k_i \eta + l_i|^2\right) \ge -\frac{7}{2}\log(e). \tag{A.207}$$

We have

$$h\left(|\eta + l|^2 + \sum_i |k_i \eta + l_i|^2\right)$$

$$= h\left(|l|^2 + 2\mathrm{Re}\left(\left(\left(l^* + \sum_i l_i^* k_i\right)\eta\right) + |\eta|^2\left(1 + \sum_i |k_i|^2\right)\right)\right) \tag{A.208}$$

$$= h\left(\left|\eta\sqrt{1+\sum_i |k_i|^2} + \frac{l+\sum_i l_i k_i^*}{\sqrt{1+\sum_i |k_i|^2}}\right|^2\right). \tag{A.209}$$

Now it suffices to show that $h\left(|\eta k' + l'|^2\right) > -(7/2)\log(e)$ for $|k'| \geq 1$. Now,

$$h\left(|\eta k' + l'|^2\right) = h\left(|\eta k'|^2 + 2|\eta||k'||l'|\cos\theta + |l'|^2\right), \tag{A.210}$$

where θ is uniformly distributed in $[0, 2\pi]$ and is independent of $|\eta|$ since η is circularly symmetric Gaussian.

$$h\left(|\eta k' + l'|^2\right) \geq h\left(|\eta k'|^2 + 2|\eta||k'||l'|\cos\theta + |l'|^2 \,\middle|\, \theta\right) \tag{A.211}$$

$$= h\left(\left(|\eta k'| + |l'|\cos\theta\right)^2 \,\middle|\, \theta\right) \tag{A.212}$$

Consider $S = ||\eta||k'| + |l'|\cos\theta'|$ for a constant θ'. It suffices to show that $h(S^2) \geq -(7/2)\log(e)$ to complete the proof. Now $\eta' = |\eta||k'|$ is Rayleigh distributed with probability density function $p_{\eta'}(x) = \left(x/|k'|^2\right)\exp\left(-x^2/\left(2|k'|^2\right)\right)$ and it easily follows that $p_{\eta'}(x) \leq (1/|k'|)\exp(-1/2) \leq \exp(-1/2)$ since $|k'| \geq 1$. Hence the probability density function of S has $p_s(x) \leq 2\exp(-1/2)$. Hence

$$h(S) = -\mathbb{E}\left[\log\left(p_s(S)\right)\right] \tag{A.213}$$

$$\geq -\log\left(2e^{-\frac{1}{2}}\right) \tag{A.214}$$

Using [LM03, (316)] for rates in bits, we have

$$h\left(S^2\right) = h(S) + \mathbb{E}\left[\log(S)\right] + 1 \tag{A.215}$$

$$\geq -\log\left(2e^{-\frac{1}{2}}\right) + \mathbb{E}\left[\log\left(\left||\eta||k'| + |l'|\cos\theta'\right|\right)\right] + 1 \tag{A.216}$$

$$= \frac{1}{2}\log(e) + \mathbb{E}\left[\log\left(\left||\eta||k'| + |l'|\cos\theta'\right|\right)\right] \tag{A.217}$$

Now it suffices to show that $\mathbb{E}\left[\log\left(\left||\eta||k'| + |l'|\cos\theta'\right|\right)\right]$ is lower bounded by $-4\log(e)$ to complete the proof. For a random variable X we define $h^-(X) = \int_{p(x)>1} p(x)\log(p(x))\,dx$. We have

$$h^-\left(|\eta||k'|\right) = \int_{p_{\eta'}(x)>1} p_{\eta'}(x)\log\left(p_{\eta'}(x)\right)dx \tag{A.218}$$

$$= 0 \tag{A.219}$$

since $p_{\eta'}(x) \leq (1/|k'|)\exp(-1/2) \leq \exp(-1/2)$. Using [LM03, (257)] to bound the expected logarithm ($\mathbb{E}[\log(|X|)] \geq -\frac{1}{(1-\alpha)^2}\log(e) - \frac{1}{\alpha}h^-(X)$ with $h^-(X) = \int_{p(x)>1} p(x)\log(p(x))\,dx$ for any $0 < \alpha < 1$), we have

$$\mathbb{E}[\log(|X|)] \geq -\frac{1}{(1-\alpha)^2}\log(e) - \frac{1}{\alpha}h^-(X),\ \ 0 < \alpha < 1, \tag{A.220}$$

$$\mathbb{E}\left[\log\left(\left||\eta|\,|k'| + |l'|\cos\theta'\right|\right)\right] \overset{(i)}{\geq} -\frac{1}{\left(1-\frac{1}{2}\right)^2}\log(e) - 2h^-(|\eta|\,|k'| + |l'|\cos\theta') \tag{A.221}$$

$$= -2h^-(|\eta|\,|k'|) - 4\log(e) \tag{A.222}$$

$$\overset{(ii)}{=} 0 - 4\log(e), \tag{A.223}$$

where (i) is using [LM03, (257)] with $\alpha = \frac{1}{2}$ and (ii) is using (A.219). Now using (A.223) in (A.217) the proof is complete.

APPENDIX B

Proofs for Chapter 3

B.1 Proof of the modified cut set outer bound for the 2-relay diamond network

Consider the cut in Figure B.1. We consider $1 \times T$ vectors X_{S},Y_{R_i},X_{R_i},Y_{D} as explained in Section 5.2.

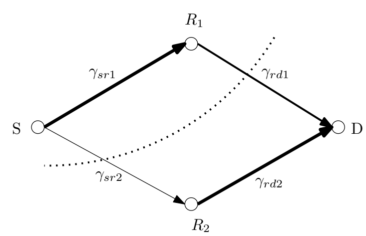


Figure B.1: The cut to be analyzed.

Considering message $M \in \left[1, 2^{nTR}\right]$ drawn uniformly, we have

$$nTR \leq H\left(M\right) \tag{B.1}$$

$$= I\left(Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n; M\right) + H\left(M \middle| Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n\right). \tag{B.2}$$

Now, $H\left(W \middle| Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n\right) \rightarrow n\epsilon_n$ due to Fano's inequality since M can be decoded from $\left(Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n\right)$. Hence

$$nTR - n\epsilon_n \leq h\left(Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n\right) - h\left(Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n \middle| M\right), \tag{B.3}$$

$$h\left(Y_{\mathrm{D}}^n, Y_{\mathrm{R}_2}^n\right) = h\left(Y_{\mathrm{R}_2}^n\right) + h\left(Y_{\mathrm{D}}^n \middle| Y_{\mathrm{R}_2}^n\right) \tag{B.4}$$

$$\overset{(i)}{\leq} \sum \left(h\left(Y_{\mathrm{R}_2k}\right) + h\left(Y_{\mathrm{D}k} \middle| Y_{\mathrm{R}_2}^{n}\right)\right) \tag{B.5}$$

$$\overset{(ii)}{=} \sum \left(h\left(Y_{\mathrm{R}_2k}\right) + h\left(Y_{\mathrm{D}k} \middle| Y_{\mathrm{R}_2}^{n}, X_{\mathrm{R}_2k}\right)\right) \tag{B.6}$$

$$\overset{(iii)}{\leq} \sum \left(h\left(Y_{\mathrm{R}_2k}\right) + h\left(Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_2k}\right)\right), \tag{B.7}$$

where (i) is using the fact that conditioning reduces entropy, (ii) is because X_{R_2k} is a function of $Y_{\mathrm{R}_2}^{k}$ which is within $Y_{\mathrm{R}_2}^{n}$; this step is different from the coherent case, where the transmitted symbols at the relays are dependent only on previously received symbols. Here we are dealing with vector symbols of size T for the noncoherent case, hence X_{R_2k} is a function of $Y_{\mathrm{R}_2}^{k}$, the transmitted block can depend on the current received block (see Figure 3.3 on page 50). The last step (iii) is using the fact that conditioning reduces entropy. Now,

$$h\left(Y_{\mathrm{D}}^{n} Y_{\mathrm{R}_2}^{n} \middle| M\right) = \sum \left(h\left(Y_{\mathrm{D}k}, Y_{\mathrm{R}_2k} \middle| M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k-1}\right)\right) \tag{B.8}$$

$$= \sum \left(h\left(Y_{\mathrm{R}_2k} \middle| M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k-1}\right) + h\left(Y_{\mathrm{D}k} \middle| M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k}\right)\right) \tag{B.9}$$

$$\overset{(i)}{\geq} \sum \left(h\left(Y_{\mathrm{R}_2k} \middle| X_{Sk}, M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k-1}\right)\right)$$

$$+ h\left(Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k}, M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k}\right) \tag{B.10}$$

$$\overset{(ii)}{=} \sum \left(h\left(Y_{\mathrm{R}_2k} \middle| X_{Sk}\right) + h\left(Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k}\right)\right), \tag{B.11}$$

where (i) is using the fact that conditioning reduces entropy and (ii) is due to the Markov chains $Y_{\mathrm{R}_2k} - X_{Sk} - \left(M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k-1}\right)$ and $Y_{\mathrm{D}k} - (X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k}) - \left(M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k}\right)$. Note that $Y_{\mathrm{D}k} - (X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k}) - \left(M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k}\right)$ is a Markov chain because given $(X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k})$, the only randomness in

$$Y_{\mathrm{D}k} = \begin{bmatrix} g_{\mathrm{rd}1k} & g_{\mathrm{rd}2k} \end{bmatrix} \begin{bmatrix} X_{\mathrm{R}_1k} \\ X_{\mathrm{R}_2k} \end{bmatrix} + W_{\mathrm{D}k}$$

is through $(g_{\mathrm{rd}1k}, g_{\mathrm{rd}2k}, W_{\mathrm{D}k})$ which is independent of $\left(M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k}\right)$. Similarly the Markovity $Y_{\mathrm{R}_2k} - X_{\mathrm{S}k} - \left(M, Y_{\mathrm{D}}^{k-1}, Y_{\mathrm{R}_2}^{k-1}\right)$ can be verified. Hence we get

$$nTR - n\epsilon_n \leq \sum \left(h\left(Y_{\mathrm{R}_2k}\right) + h\left(Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_2k}\right) - h\left(Y_{\mathrm{R}_2k} \middle| X_{\mathrm{S}k}\right) - h\left(Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_1k}, X_{\mathrm{R}_2k}\right)\right) \tag{B.12}$$

$$= \sum \left(I\left(X_{\mathrm{S}k}; Y_{\mathrm{R}_2k}\right) + I\left(X_{\mathrm{R}_1k}; Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_2k}\right)\right). \tag{B.13}$$

Due to symmetry, it follows for the second cut (Figure B.2) that

$$nTR - n\epsilon_n \leq \sum \left(I\left(X_{\mathrm{S}k}; Y_{\mathrm{R}_1k}\right) + I\left(X_{\mathrm{R}_2k}; Y_{\mathrm{D}k} \middle| X_{\mathrm{R}_1k}\right)\right). \tag{B.14}$$

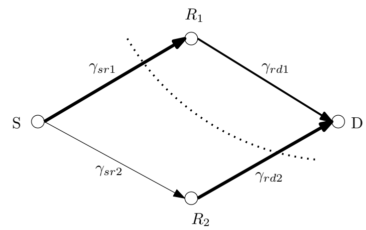


Figure B.2: The second cut.

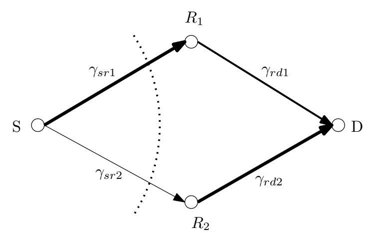


Figure B.3: The SIMO cut.

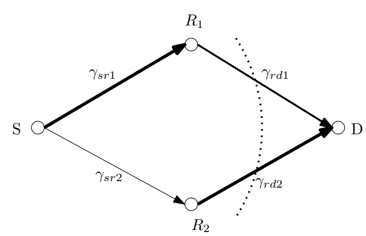


Figure B.4: The MISO cut.

For MISO and SIMO cuts it easily follows that

$$nTR - n\epsilon_n \leq \sum I\left(X_{Sk}; Y_{\mathrm{R}_1 k}, Y_{\mathrm{R}_2 k}\right), \tag{B.15}$$

$$nTR - n\epsilon_n \leq \sum I\left(X_{\mathrm{R}_1 k}, X_{\mathrm{R}_2 k}; Y_{\mathrm{D}}\right). \tag{B.16}$$

Using equations (B.13), (B.14), (B.15) and (B.16) and a time-sharing argument as used for the usual cut-set outer bounds [CT12, (Theorem 15.10.1)], we get the outer bound

$$\begin{aligned} T\bar{C} = \sup_{p\left(X_{\mathrm{S}}, X_{\mathrm{R}_1}, X_{\mathrm{R}_2}\right)} \min \Big\{ & I\left(X_{\mathrm{S}}; Y_{\mathrm{R}}\right), I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_2}\right), \\ & I\left(X_{\mathrm{S}}; Y_{\mathrm{R}_1}\right) + I\left(X_{\mathrm{R}_2}; Y_{\mathrm{D}} \middle| X_{\mathrm{R}_1}\right), I\left(X_{\mathrm{R}}; Y_{\mathrm{D}}\right) \Big\}. \end{aligned} \tag{B.17}$$

B.2 A generalization of the cut set outer bound for acyclic noncoherent networks

Consider an acyclic noncoherent wireless network with coherence time T and independent fading in the links and additive white Gaussian noise. We consider the transmitted vector symbols X_i (transmitted from node i) and received vector symbols Y_i (received at node i) of length T. The fading is constant within each vector symbol but independent across the different vector symbols.

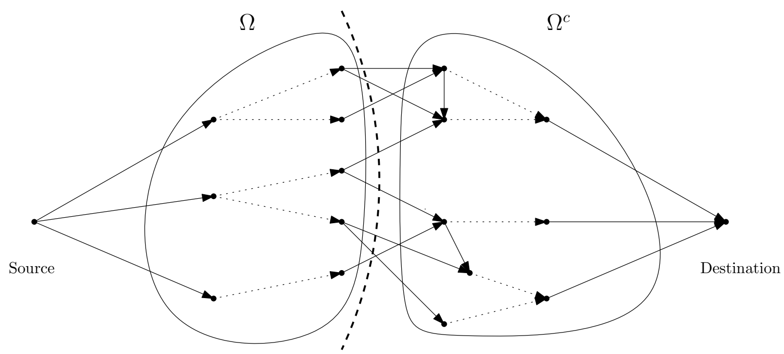


Figure B.5: A source-destination cut described by Ω in a general acyclic network. The set Ω has the nodes in the source side of the cut, the set Ω^c has the nodes in the destination side of the cut.

Let $L = |\Omega^c|$, let $(1), (2), \ldots, (L)$ be the nodes in the set Ω^c, the labeling of nodes is done with a partial ordering; any transmit symbols goes ONLY from a node with smaller numbering to larger numbering. Such a labeling exists since the network is acyclic. Let $X_{\text{in}(i)}$ denote all the transmit signals incoming to the node (i) and let X_{Ω^c} denote all the transmit signals in the destination side of the cut. We claim the following:

$$TR \leq \sum_i \left(h\left(Y_{(i)} \middle| Y_{(1)}, \ldots, Y_{(i-1)} \left(X_{\text{in}(i)} \bigcap X_{\Omega^c} \right) \right) - h\left(Y_{(i)} \middle| X_{\text{in}(i)} \right) \right) \tag{B.18}$$

and

$$TR \le \sum_i \left(h\left(Y_{(i)}\middle| Y_{(1)},\ldots,Y_{(i-1)},X_{(1)},\ldots,X_{(i-1)}\right) - h\left(Y_{(i)}\middle| X_{\text{in}(i)}\right)\right) \tag{B.19}$$

for some joint distribution on $X_i's$ and corresponding $Y_i's$ induced by the noncoherent channel. The proof is as follows.

Due to Fano's inequality, we have

$$\begin{aligned} nTR - n\epsilon_n &\le I\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)};M\right) \\ &= h\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)}\right) - h\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)}\middle| M\right) \end{aligned}$$

$$\begin{aligned} h\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)}\right) &= \sum_i h\left(Y^n_{(i)}\middle| Y^n_{(1)},\ldots,Y^n_{(i-1)}\right) &\text{(B.20)}\\ &\overset{\text{(i)}}{\le} \sum_i\sum_k h\left(Y_{(i)k}\middle| Y^n_{(1)},\ldots,Y^n_{(i-1)}\right) &\text{(B.21)}\\ &\overset{\text{(ii)}}{=} \sum_i\sum_k h\left(Y_{(i)k}\middle| Y^n_{(1)},\ldots,Y^n_{(i-1)},\left(X_{\text{in}(i)}\bigcap X_{\Omega^c}\right)_k\right) &\text{(B.22)}\\ &\le \sum_i\sum_k h\left(Y_{(i)k}\middle| Y_{(1)k},\ldots,Y_{(i-1)k},\left(X_{\text{in}(i)}\bigcap X_{\Omega^c}\right)_k\right), &\text{(B.23)} \end{aligned}$$

where (i) is because conditioning reduces entropy, (ii) is since $\left(X_{\text{in}(i)}\bigcap X_{\Omega^c}\right)_k$ is a function of $Y^n_{(1)},\ldots,Y^n_{(i-1)}$ because of the nature of labeling (instead we could have also used $X_{(1)k},\ldots,X_{(i-1)k}$ in the conditioning, which is also a function of $Y^n_{(1)},\ldots,Y^n_{(i-1)}$)

Remark B.1. *Note that IF we expanded*

$$h\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)}\right) = \sum_k h\left(Y_{(1)k},\ldots,Y_{(L)k}\middle| Y^{k-1}_{(1)},\ldots,Y^{k-1}_{(L)}\right)$$

as in the usual cut-set outer bound, then $X^k_{(1)},\ldots,X^k_{(L)}$ is NOT a function of $Y^{k-1}_{(1)},\ldots,Y^{k-1}_{(L)}$. Due to the block structure, $X^k_{(1)},\ldots,X^k_{(L)}$ is a function of $Y^k_{(1)},\ldots,Y^k_{(L)}$. This is similar to that we explain in the derivation for the diamond network in (B.6) on page 188.

Now,

$$h\left(Y^n_{(1)},Y^n_{(2)},\ldots,Y^n_{(L)}\middle| M\right) = \sum_i h\left(Y^n_{(i)}\middle| M,Y^n_{(1)},\ldots,Y^n_{(i-1)}\right) \tag{B.24}$$

$$= \sum_i \sum_k h\left(Y_{(i)k} \middle| M, Y_{(1)}^n, \ldots, Y_{(i-1)}^n, Y_{(i)}^{k-1}\right) \tag{B.25}$$

$$\overset{\text{(i)}}{\geq} \sum_i \sum_k h\left(Y_{(i)k} \middle| X_{\text{in}(i)k}, M, Y_{(1)}^n, \ldots, Y_{(i-1)}^n, Y_{(i)}^{k-1}\right) \tag{B.26}$$

$$\overset{\text{(ii)}}{=} \sum_i \sum_k h\left(Y_{(i)k} \middle| X_{\text{in}(i)k}\right), \tag{B.27}$$

where (i) is because conditioning reduces entropy and (ii) is because of the Markov Chain $Y_{(i),k} - X_{\text{in}(i)k} - \left(M, Y_{(1)}^n \ldots Y_{(i-1)}^n, Y_{(i)}^{k-1}\right)$. The Markovity holds because given $X_{\text{in}(i)k}$, $Y_{(i)k}$ is dependent only on the additive Gaussian noise and the fading in the incoming links which are independent of $\left(M, Y_{(1)}^n, \ldots, Y_{(i-1)}^n, Y_{(i)}^{k-1}\right)$. Using a time-sharing argument as in the usual cut-set outer bound we get

$$TR \leq \sum_i \left(h\left(Y_{(i)} \middle| Y_{(1)}, \ldots, Y_{(i-1)}, \left(X_{\text{in}(i)} \bigcap X_{\Omega^c}\right)\right) - h\left(Y_{(i)} \middle| X_{\text{in}(i)}\right)\right) \tag{B.28}$$

for some joint distribution on $X_i's$ and corresponding $Y_i's$ induced by the noncoherent channel. Similarly, if we had used $X_{(1)k}, \ldots, X_{(i-1)k}$ in (B.22) instead of $\left(X_{\text{in}(i)} \bigcap X_{\Omega^c}\right)_k$, we would have obtained

$$TR \leq \sum_i \left(h\left(Y_{(i)} \middle| Y_{(1)}, \ldots, Y_{(i-1)}, X_{(1)}, \ldots, X_{(i-1)}\right) - h\left(Y_{(i)} \middle| X_{\text{in}(i)}\right)\right). \tag{B.29}$$

Remark B.2. *The outer bound of the form*

$$TR \leq \sup_{p(X)} \min_{\Omega} \left\{r\left(p\left(X\right), \Omega\right)\right\} \tag{B.30}$$

with min *taken over all cuts and the* sup *taken over all probability distributions can be obtained, with rate expression* $r\left(p\left(X\right), \Omega\right)$ *of the form taken from the RHS of (B.28) or (B.29). Note that this would require different labeling of nodes depending on the cut, since to derive (B.28) and (B.29), the nodes are labeled depending on the cut.*

B.3 Proof of Theorem 3.3: structure of the optimizing distribution

The cut-set outer bound from (3.7), can be rewritten as

$$T\bar{C} = \sup_{p(X)} I_{out}\left(p\left(X\right)\right), \tag{B.31}$$

where

$$\begin{aligned} I_{out}\left(p\left(X\right)\right) = \min\Big\{ & I\left(X_{\mathrm{S}};Y_{\mathrm{R}}\right), I\left(X_{\mathrm{S}};Y_{\mathrm{R}_2}\right) + I\left(X_{\mathrm{R}_1};Y_{\mathrm{D}}\middle| X_{\mathrm{R}_2}\right), \\ & I\left(X_{\mathrm{S}};Y_{\mathrm{R}_1}\right) + I\left(X_{\mathrm{R}_2};Y_{\mathrm{D}}\middle| X_{\mathrm{R}_1}\right), I\left(X_{\mathrm{R}};Y_{\mathrm{D}}\right)\Big\} \end{aligned} \tag{B.32}$$

Lemma B.1. *(Invariance of $I_{out}\left(p\left(X\right)\right)$ to post-rotations of X). Suppose that X has a probability distribution $p_0\left(X\right)$ that generates some $I_{out}\left(p\left(X\right)\right)$. Then, for any unitary matrix Φ, the "post-rotated" probability distribution, $p_1\left(X\right) = p_0\left(X\Phi^{\dagger}\right)$ also generates $I_{out}\left(p\left(X\right)\right)$.*

Proof. This is an adaptation of the existing results for MIMO from [MH99, Lemma 1] to the network case. We prove this by taking each term of $I_{out}\left(p\left(X\right)\right)$ from (B.31). We give the sample proof showing $I\left(X_{1\mathrm{R}_1};Y_{1\mathrm{D}}\middle| X_{1\mathrm{R}_2}\right) = I\left(X_{0\mathrm{R}_1};Y_{0\mathrm{D}}\middle| X_{0\mathrm{R}_2}\right)$, where $\left(X_{1\mathrm{R}_1}, X_{1\mathrm{R}_2}, X_{1\mathrm{S}}\right)$ indicate a post-rotated version of $\left(X_{0\mathrm{R}_1}, X_{0\mathrm{R}_2}, X_{0\mathrm{S}}\right)$ with a given unitary $\Phi^{\dagger}$. Other terms follow the same arguments for proof. Let $\underline{g_{\mathrm{rd}}} = \left[\begin{array}{cc} g_{\mathrm{rd}1} & g_{\mathrm{rd}2}\end{array}\right]$ and $X_{0\mathrm{R}} = \mathrm{Transpose}\left(\left[X_{0\mathrm{R}_1}, X_{0\mathrm{R}_2}\right]\right)$, then $Y_{1\mathrm{D}} = \underline{g_{\mathrm{rd}}} X_{0\mathrm{R}}\Phi^{\dagger} + W_{\mathrm{D}}$. Hence we have

$$\begin{aligned} I\left(X_{1\mathrm{R}_1};Y_{1\mathrm{D}}\middle| X_{1\mathrm{R}_2}\right) &= h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}}\Phi^{\dagger} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}_2}\Phi^{\dagger}\right) - h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}}\Phi^{\dagger} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}}\Phi^{\dagger}\right) && \text{(B.33)}\\ &\overset{(i)}{=} h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\Phi\middle| X_{0\mathrm{R}_2}\Phi^{\dagger}\right) - h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\Phi\middle| X_{0\mathrm{R}}\Phi^{\dagger}\right) && \text{(B.34)}\\ &\overset{(ii)}{=} h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}_2}\Phi^{\dagger}\right) - h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}}\Phi^{\dagger}\right) && \text{(B.35)}\\ &\overset{(iii)}{=} h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}_2}\right) - h\left(\underline{g_{\mathrm{rd}}}X_{0\mathrm{R}} + W_{\mathrm{D}}\middle| X_{0\mathrm{R}}\right) && \text{(B.36)}\\ &= I\left(X_{0\mathrm{R}_1};Y_{0\mathrm{D}}\middle| X_{0\mathrm{R}_2}\right), && \text{(B.37)} \end{aligned}$$

where (i) is because unitary transformation preserves entropy, (ii) is because W_{D}, $W_{\mathrm{D}}\Phi$ have same distribution since W_{D} has i.i.d. $\mathcal{CN}\left(0,1\right)$ elements and Φ is unitary and (iii) is because Φ is a given unitary matrix and can be removed from conditioning. $\square$

Now, we show that he signal of the form $X = LQ$ with L being a lower triangular random matrix and Q being an isotropically distributed unitary matrix independent of L, achieves the outer bound in (B.31). This is also an adaptation of the existing results, we follow the techniques from [MH99, Theorem 2]. Let X_0 be a random variable which is optimal for the outer bound and I_0 be the corresponding mutual information achieved. Now X_0 can be decomposed as $X_0 = L\Phi'$ using LQ decomposition with L upper diagonal and Φ' unitary, but they could be jointly distributed and Φ' may not be isotropically unitary distributed. Let Θ be an isotropically distributed unitary matrix that is statistically independent of L and Φ'. Now, use $X_1 = X_0\Theta$ for signaling and let Y_1 be the corresponding received signal. Now,

$$I\left(X_{1\mathrm{R}_1}, \Theta; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) = I\left(X_{1\mathrm{R}_1}, \Theta; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) \tag{B.38}$$

$$I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) + I\left(\Theta; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2} X_{1\mathrm{R}_1}\right) = I\left(\Theta; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) + I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}, \Theta\right) \tag{B.39}$$

$$I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) + 0 \overset{(i)}{=} I\left(\Theta; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) + I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}, \Theta\right) \tag{B.40}$$

$$\begin{aligned} I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}\right) &\overset{(ii)}{\geq} I\left(X_{1\mathrm{R}_1}; Y_{1\mathrm{D}} \middle| X_{1\mathrm{R}_2}, \Theta\right) && \text{(B.41)} \\ &= I\left(X_{0\mathrm{R}_1}\Theta; Y_{1\mathrm{D}} \middle| X_{0\mathrm{R}_2}\Theta, \Theta\right) && \text{(B.42)} \\ &\overset{(iii)}{=} I\left(X_{0\mathrm{R}_1}; Y_{0\mathrm{D}} \middle| X_{0\mathrm{R}_2}\right), && \text{(B.43)} \end{aligned}$$

where (i) is because $Y_{1\mathrm{D}} - (X_{1\mathrm{R}_2}, X_{1\mathrm{R}_1}) - \Theta$ forms a Markov chain, (ii) is due to the nonnegativity of mutual information and (iii) is using Lemma B.1. Similarly it can be shown that each term of $I_{out}\left(p\left(X\right)\right)$ from (B.31) increases by choosing $X_1 = X_0\Theta$ for signaling instead of X_0. Hence without loss of generality, the signal of the form $LQ = L\Phi'\Theta$ with $Q = \Phi'\Theta$ is optimal for the outer bound. Now, $Q = \Phi'\Theta$ is also an isotropically distributed unitary matrix and independent of Φ' by the property of isotropically distributed matrices.

B.4 Proof of Discretization Lemma (Lemma 3.1)

We have

$$\begin{aligned}\psi_1 =& T\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + T\right)\right] \\ &- \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1\right)\right] && \text{(B.44)} \\ =& \mathbb{E}\left[f_1\left(|a|^2, |b|^2, |c|^2\right)\right], && \text{(B.45)} \\ \psi_2 =& (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + 1\right)\right] \\ &+ (T-1)\mathbb{E}\left[\log\left(\rho_{\mathrm{rd1}}^2 |c|^2 + T - 1\right)\right] \\ &- \mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1\right)\right] && \text{(B.46)} \\ =& (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + \mathbb{E}\left[f_2\left(|a|^2, |b|^2, |c|^2\right)\right], && \text{(B.47)}\end{aligned}$$

where we also included the straight-forward definition of $f_1(\cdot), f_2(\cdot)$ in the previous equations.

Note that

$$\begin{aligned}\left|\frac{\partial f_2}{\partial |a|^2}\right| \le& \frac{\rho_{\mathrm{rd2}}^2}{\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + 1} \\ &+ \frac{\rho_{\mathrm{rd2}}^2\left(1+\rho_{\mathrm{rd2}}^2 |c|^2\right)}{\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1} && \text{(B.48)} \\ \le& \rho_{\mathrm{rd2}}^2 + \frac{\rho_{\mathrm{rd2}}^2\left(1+\rho_{\mathrm{rd2}}^2 |c|^2\right)}{\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1} && \text{(B.49)} \\ =& \rho_{\mathrm{rd2}}^2 + \frac{\rho_{\mathrm{rd2}}^2}{1 + \rho_{\mathrm{rd2}}^2 |a|^2 + \frac{\rho_{\mathrm{rd1}}^2 |b|^2}{1+\rho_{\mathrm{rd2}}^2 |c|^2}} && \text{(B.50)} \\ \le& 2\rho_{\mathrm{rd2}}^2, && \text{(B.51)}\end{aligned}$$

$$\begin{aligned}\left|\frac{\partial f_2}{\partial |b|^2}\right| \le& \frac{\rho_{\mathrm{rd1}}^2}{\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + 1} \\ &+ \frac{\rho_{\mathrm{rd1}}^2}{\rho_{\mathrm{rd2}}^2 |a|^2 + \rho_{\mathrm{rd1}}^2 |b|^2 + \rho_{\mathrm{rd1}}^2 |c|^2 + \rho_{\mathrm{rd1}}^2\rho_{\mathrm{rd2}}^2 |c|^2 |a|^2 + 1} && \text{(B.52)} \\ \le& 2\rho_{\mathrm{rd1}}^2 && \text{(B.53)} \\ \le& 2\rho_{\mathrm{rd2}}^2, && \text{(B.54)}\end{aligned}$$

$$\left|\frac{\partial f_2}{\partial |c|^2}\right| \leq \frac{(T-1)\,\rho_{\text{rd1}}^2}{\rho_{\text{rd1}}^2\,|c|^2+T-1}+\frac{\rho_{\text{rd1}}^2}{1+\rho_{\text{rd1}}^2\,|c|^2+\frac{\rho_{\text{rd1}}^2|b|^2}{1+\rho_{\text{rd2}}^2|a|^2}} \tag{B.55}$$

$$\leq 2\rho_{\text{rd1}}^2 \tag{B.56}$$

$$\leq 2\rho_{\text{rd2}}^2. \tag{B.57}$$

Hence for $\left|\left|\left(|a|^2,|b|^2,|c|^2\right)-\left(|a'|^2,|b'|^2,|c'|^2\right)\right|\right|_2 \leq \sqrt{3}/\rho_{\text{rd2}}^2$,

$$\left|f_2\left(|a|^2,|b|^2,|c|^2\right)-f_2\left(|a'|^2,|b'|^2,|c'|^2\right)\right| \leq \left|\left|\left(2\rho_{\text{rd2}}^2,2\rho_{\text{rd2}}^2,2\rho_{\text{rd2}}^2\right)\right|\right|_2\left(\frac{\sqrt{3}}{\rho_{\text{rd2}}^2}\right) \tag{B.58}$$

$$=6 \tag{B.59}$$

and in a similar manner as above, it can be shown that

$$\left|f_1\left(|a|^2,|b|^2,|c|^2\right)-f_1\left(|a'|^2,|b'|^2,|c'|^2\right)\right| \leq 6. \tag{B.60}$$

Hence by considering a discrete version of the problem as

$$\mathcal{P}_2:\begin{cases}\underset{\mathbb{E}\left[|a|^2\right]\leq T,\mathbb{E}\left[|b|^2+|c|^2\right]\leq T}{\text{maximize}}\ \min\{\psi_1,\psi_2\}\\ \text{Support}\left(|a|^2,|b|^2,|c|^2\right)=\left\{0,\frac{1}{\rho_{\text{rd2}}^2},\frac{2}{\rho_{\text{rd2}}^2},\ldots,\infty\right\}^3,\end{cases} \tag{B.61}$$

the optimum value achieved is within 6 of the optimum value of $\mathcal{P}_1$ (refer to Theorem 3.4 on page 53 for definition of $\mathcal{P}_1$). Hence for an outer bound on degrees of freedom it is sufficient to solve $\mathcal{P}_2$.

$$\text{gDoF}\left(\mathcal{P}_1\right)=\text{gDoF}\left(\mathcal{P}_2\right) \tag{B.62}$$

Claim B.1. *The new optimization problem*

$$\mathcal{P}_3:\begin{cases}\underset{\mathbb{E}\left[|a|^2\right]\leq T,\mathbb{E}\left[|b|^2+|c|^2\right]\leq T}{\textit{maximize}}\ \textit{min}\{\psi_1,\psi_2\}\\ \textit{Support}\left(|a|^2,|b|^2,|c|^2\right)=\left\{0,\frac{1}{\rho_{rd2}^2},\frac{2}{\rho_{rd2}^2},\ldots,\frac{\lfloor\rho_{rd2}^4\rfloor}{\rho_{rd2}^2}\right\}^3=S_1\end{cases} \tag{B.63}$$

achieves the same degrees of freedom as $\mathcal{P}_2$.

Proof. Here we show that it is sufficient to restrict $\text{Support}\left(|a|^2,|b|^2,|c|^2\right)=\{0,1/\rho_{\text{rd2}}^2,2/\rho_{\text{rd2}}^2,\ldots,\lfloor\rho_{\text{rd2}}^4\rfloor/\rho_{\text{rd2}}^2\}^3$, for a tight outer bound on gDoF. The main idea behind this claim is that outside this support, the points have very high power and hence due to the power constraints only very low probability can be assigned to those points. The probabilities assigned is low enough, so that the terms of the form $\mathbb{E}\left[\log\left(\rho_{\mathrm{rd2}}^2|a|^2+\rho_{\mathrm{rd1}}^2|b|^2+\rho_{\mathrm{rd1}}^2|c|^2\right)\right]$ do not contribute much from those points.

Let the optimum value of $\mathcal{P}_2$ be achieved by a probability distribution $\{p_i^*\}$ at the points $\{(l_{1i}^*/\rho_{\mathrm{rd2}}^2, l_{2i}^*/\rho_{\mathrm{rd2}}^2, l_{3i}^*/\rho_{\mathrm{rd2}}^2)\}$with $l_{ji}^*\in\mathbb{Z}$. Let

$$S_1=\left\{i:\max\left\{\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right\}\le\rho_{\mathrm{rd2}}^2\right\} \tag{B.64}$$

$$S_2=\left\{i:\max\left\{\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right\}>\rho_{\mathrm{rd2}}^2\right\} \tag{B.65}$$

and let $\max\{l_{1i}^*/\rho_{\mathrm{rd2}}^2, l_{2i}^*/\rho_{\mathrm{rd2}}^2, l_{3i}^*/\rho_{\mathrm{rd2}}^2\}=l_{Mi}^*/\rho_{\mathrm{rd2}}^2$ for labeling. Now,

$$\psi_2^*=(T-1)\log\left(\rho_{\mathrm{sr2}}^2\right)+\sum_{i\in S_1}p_i^*f_2\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right)+\sum_{i\in S_2}p_i^*f_2\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right) \tag{B.66}$$

and

$$\begin{aligned}&\sum_{i\in S_2}p_i^*f_2\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right)\\&\overset{(i)}{\le}\sum_{i\in S_2}p_i^*\left(\log\left(2\rho_{\mathrm{rd2}}^2\frac{l_{Mi}^*}{\rho_{\mathrm{rd2}}^2}+1\right)+(T-1)\log\left(\rho_{\mathrm{rd2}}^2\frac{l_{Mi}^*}{\rho_{\mathrm{rd2}}^2}+T-1\right)\right) &\text{(B.67)}\\&\le T\sum_{i\in S_2}p_i^*\log\left(2l_{Mi}^*+T\right), &\text{(B.68)}\end{aligned}$$

where (i) is because $\max\{l_{1i}^*/\rho_{\mathrm{rd2}}^2, l_{2i}^*/\rho_{\mathrm{rd2}}^2, l_{3i}^*/\rho_{\mathrm{rd2}}^2\}=l_{Mi}^*/\rho_{\mathrm{rd2}}^2$ for $i\in S_2$ and using the structure of the function $f_2(\cdot)$. Hence

$$\begin{aligned}&\sum_{i\in S_2}p_i^*f_2\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2},\frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right)\\&\le T\sum_{i\in S_2}p_i^*\log\left(2l_{Mi}^*+T\right) &\text{(B.69)}\\&\overset{(i)}{\le}T\sum_{i\in S_2}p_i^*\log\left(2\frac{\sum_{i\in S_2}p_i^*l_{Mi}^*}{\sum_{j\in S_2}p_j^*}+T\right) &\text{(B.70)}\\&\overset{(ii)}{\le}T\sum_{i\in S_2}p_i^*\log\left(4\frac{T\rho_{\mathrm{rd2}}^2}{\sum_{j\in S_2}p_j^*}+T\right) &\text{(B.71)}\end{aligned}$$

$$= T\sum_{i\in S_2} p_i^* \log\left(4T\rho_{\mathrm{rd2}}^2 + T\sum_{j\in S_2} p_j^*\right) - T\sum_{i\in S_2} p_i^* \log\left(\sum_{j\in S_2} p_j^*\right) \tag{B.72}$$

$$\overset{(iii)}{\leq} T\sum_{i\in S_2} p_i^* \log\left(4T\rho_{\mathrm{rd2}}^2 + T\right) + T\frac{\log(e)}{e} \tag{B.73}$$

$$\overset{(iv)}{\leq} T\frac{2T}{\rho_{\mathrm{rd2}}^2} \log\left(4T\rho_{\mathrm{rd2}}^2 + T\right) + T\frac{\log(e)}{e} \tag{B.74}$$

$$= T\frac{2T}{\rho_{\mathrm{rd2}}^2} \times \left(\log(T) + \log\left(4\rho_{\mathrm{rd2}}^2 + 1\right)\right) + T\frac{\log(e)}{e} \tag{B.75}$$

$$\overset{(v)}{\leq} T\frac{2T}{\rho_{\mathrm{rd2}}^2} \times \left(\log(T) + \left(4\rho_{\mathrm{rd2}}^2 + 1\right)\frac{\log(e)}{e}\right) + T\frac{\log(e)}{e} \tag{B.76}$$

$$\overset{(vi)}{\leq} 2T^2 \times \left(\log(T) + 5\frac{\log(e)}{e}\right) + T\frac{\log(e)}{e} \tag{B.77}$$

$$\overset{(vii)}{=} r_2(T), \tag{B.78}$$

where (i) is due to Jensen's inequality, (ii) is due to the power constraint $\sum_{i\in S_2} p_i^* \left(l_{Mi}^*/\rho_{\mathrm{rd2}}^2\right) \leq 2T \Rightarrow \sum_{i\in S_2} p_i^* l_{Mi}^* \leq 2T\rho_{\mathrm{rd2}}^2$, (iii) is due to the fact $0 \leq \sum_{i\in S_2} p_i^* \leq 1$ and $-x\log(x) \geq \log(e)/e$ for $x \in [0,1]$, (iv) is due to the fact $\sum_{i\in S_2} p_i^* \left(l_{Mi}^*/\rho_{\mathrm{rd2}}^2\right) \leq 2T$ (power constraint) and $\rho_{\mathrm{rd2}}^2 < (l_{Mi}^*/\rho_{\mathrm{rd2}}^2)$in S_2 and hence $\sum_{i\in S_2} p_i^* \rho_{\mathrm{rd2}}^2 \leq 2T$ and $\sum_{i\in S_2} p_i^* \leq 2T/\rho_{\mathrm{rd2}}^2$, (v) is due to the fact $(1/x)\log(x) \leq \log(e)/e$ for $x \in [1, +\infty)$, (vi) is assuming $\rho_{\mathrm{rd2}}^2 > 1$ (otherwise Relay R_2 does not contribute to gDoF and can be removed from the network), (vii) is by defining $r_2(T) = 2T^2 \times (\log(T) + 5\log(e)/e) + T\log(e)/e$.

Hence it follows that

$$\psi_2^* = (T-1)\log\left(\rho_{\mathrm{sr2}}^2\right) + \sum_{i\in S_1} p_i^* f_1\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2}, \frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2}, \frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right) + r_2(T) \tag{B.79}$$

and similarly it can be shown

$$\psi_1^* = \sum_{i\in S_1} p_i^* f_2\left(\frac{l_{1i}^*}{\rho_{\mathrm{rd2}}^2}, \frac{l_{2i}^*}{\rho_{\mathrm{rd2}}^2}, \frac{l_{3i}^*}{\rho_{\mathrm{rd2}}^2}\right) + r_1(T) \tag{B.80}$$

for some $r_1(T)$ independent of SNR. Hence it follows that

$$\mathcal{P}_3 : \begin{cases} \underset{\mathbb{E}\left[|a|^2\right]\leq T, \mathbb{E}\left[|b|^2+|c|^2\right]\leq T}{\text{maximize}} \min\{\psi_1, \psi_2\} \\ \text{Support}\left(|a|^2, |b|^2, |c|^2\right) = \left\{0, \frac{1}{\rho_{\mathrm{rd2}}^2}, \frac{2}{\rho_{\mathrm{rd2}}^2}, \ldots, \frac{\lfloor\rho_{\mathrm{rd2}}^4\rfloor}{\rho_{\mathrm{rd2}}^2}\right\}^3 = S_1 \end{cases} \tag{B.81}$$

achieves the same degrees of freedom as $\mathcal{P}_2$, because any nonzero probability outside S_1 in $\mathcal{P}_2$ can be assigned to $(0,0,0)$ in $\mathcal{P}_3$, changing the value of objective function only by a constant independent of SNR. $\square$

Hence

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right) = \text{gDoF}\left(\mathcal{P}_3\right). \tag{B.82}$$

Now, for

$$\mathcal{P}_4 : \begin{cases} \underset{\mathbb{E}\left[|a|^2+|b|^2+|c|^2\right]\leq 2T}{\text{maximize}} \quad \min\left\{\psi_1, \psi_2\right\} \\ \text{Support}\left(|a|^2, |b|^2, |c|^2\right) = \left\{0, \frac{1}{\rho_{\text{rd2}}^2}, \frac{2}{\rho_{\text{rd2}}^2}, \ldots, \frac{\left\lfloor \rho_{\text{rd2}}^4 \right\rfloor}{\rho_{\text{rd2}}^2}\right\}^3, \end{cases} \tag{B.83}$$

we have

$$\text{gDoF}\left(\mathcal{P}_3\right) \leq \text{gDoF}\left(\mathcal{P}_4\right). \tag{B.84}$$

In fact it can be easily shown that

$$\text{gDoF}\left(\mathcal{P}_3\right) = \text{gDoF}\left(\mathcal{P}_4\right) \tag{B.85}$$

by considering a new optimization problem with $\mathbb{E}\left[|a|^2 + |b|^2 + |c|^2\right] \leq T$ and using the fact that a constant scaling in a, b, c can be absorbed into SNR and using the behavior of $\log()$ under constant scaling. The detailed proof is omitted. We then will have

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right) = \text{gDoF}\left(\mathcal{P}_3\right) = \text{gDoF}\left(\mathcal{P}_4\right). \tag{B.86}$$

Now $\mathcal{P}_4$ is a linear program with finite number of variables and constraints (with a finite optimum value because of Jensen's inequality). The variables are $\{p_i\}$ and the maximum number of nontrivial active constraints on $\{p_i\}$ is 3, derived from

$$\psi_1 = \psi_2 \tag{B.87}$$

$$\mathbb{E}\left[|a|^2 + |b|^2 + |c|^2\right] = 2T \tag{B.88}$$

$$\sum p_i = 1. \tag{B.89}$$

Trivial constraints are $p_i \geq 0$. Hence using the theory of linear programming there exists an optimal $\{p_i^*\}$ with at most 3 nonzero values. Hence it follows that

$$\mathcal{P}_5 : \begin{cases} \underset{\sum_{i=1}^{3} p_i\left(|a_i|^2+|b_i|+|c_i|^2\right)\leq 2T}{\text{maximize}} \quad \min\Big\{ \sum_{i=1}^{3} p_i f_1\left(|a_i|^2, |b_i|^2, |c_i|^2\right) \\ \qquad\qquad (T-1)\log\left(\rho_{\text{sr2}}^2\right) + \sum_{i=1}^{3} p_i f_2\left(|a_i|^2, |b_i|^2, |c_i|^2\right) \Big\} \\ |a_i|^2, |b_i|^2, |c_i|^2 \geq 0 \end{cases} \tag{B.90}$$

has $(\mathcal{P}_5) \geq (\mathcal{P}_4)$. Note that we have allowed $\left(|a_i|^2, |b_i|^2, |c_i|^2\right)_{i=1}^{3}$ to be real positive variables to be optimized, instead of the discrete values. But it is also clear that $(\mathcal{P}_5) \leq (\mathcal{P}_1)$. Now, since $\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_4\right)$ it follows that

$$\text{gDoF}\left(\mathcal{P}_1\right) = \text{gDoF}\left(\mathcal{P}_2\right) = \text{gDoF}\left(\mathcal{P}_3\right) = \text{gDoF}\left(\mathcal{P}_4\right) = \text{gDoF}\left(\mathcal{P}_5\right). \tag{B.91}$$

Now, we consider solving $\mathcal{P}_5$. We have

$$\begin{aligned} & f_1\left(|a|^2, |b|^2, |c|^2\right) \\ & = T\log\left(\rho_{\text{rd2}}^2|a_i|^2 + \rho_{\text{rd1}}^2|b_i|^2 + \rho_{\text{rd1}}^2|c_i|^2 + T\right) && \text{(B.92)} \\ & \quad - \log\left(\rho_{\text{rd2}}^2|a_i|^2 + \rho_{\text{rd1}}^2|b_i|^2 + \rho_{\text{rd1}}^2|c_i|^2 + \rho_{\text{rd1}}^2\rho_{\text{rd2}}^2|c_i|^2|a_i|^2 + 1\right) && \text{(B.93)} \end{aligned}$$

$$\begin{aligned} & f_2\left(|a_i|^2, |b_i|^2, |c_i|^2\right) && \text{(B.94)} \\ & = \log\left(\rho_{\text{rd2}}^2|a_i|^2 + \rho_{\text{rd1}}^2|b_i|^2 + 1\right) + (T-1)\log\left(\rho_{\text{rd1}}^2|c_i|^2 + T - 1\right) \\ & \quad - \log\left(\rho_{\text{rd2}}^2|a_i|^2 + \rho_{\text{rd1}}^2|b_i|^2 + \rho_{\text{rd1}}^2|c_i|^2 + \rho_{\text{rd1}}^2\rho_{\text{rd2}}^2|c_i|^2|a_i|^2 + 1\right). && \text{(B.95)} \end{aligned}$$

If $\rho_{\text{rd2}}^2|a|^2 \geq \max\left(\rho_{\text{rd1}}^2|b_i|^2,\ \rho_{\text{rd1}}^2|c_i|^2\right)$, then it can be easily seen that setting $|b_i'|^2 = 0$ decreases f_1 and f_2 by at most a constant independent of SNR and then we get

$$\begin{aligned} f_1\left(|a_i|^2, |b_i'|^2, |c_i|^2\right) & \doteq T\log\left(\rho_{\text{rd2}}^2|a_i|^2 + T\right) \\ & \quad - \log\left(\rho_{\text{rd2}}^2|a_i|^2 + \rho_{\text{rd1}}^2|c_i|^2 + \rho_{\text{rd1}}^2\rho_{\text{rd2}}^2|c_i|^2|a_i|^2 + 1\right) && \text{(B.96)} \\ & \doteq (T-1)\log\left(\rho_{\text{rd2}}^2|a_i|^2 + 1\right) - \log\left(\rho_{\text{rd1}}^2|c_i|^2 + 1\right) && \text{(B.97)} \end{aligned}$$

$$f_2\left(|a_i|^2, |b_i'|^2, |c_i|^2\right) \doteq \log\left(\rho_{\text{rd2}}^2|a_i|^2 + 1\right) + (T-1)\log\left(\rho_{\text{rd1}}^2|c_i|^2 + T - 1\right)$$

$$- \log\left(\rho_{\text{rd2}}^2 |a_i|^2 + \rho_{\text{rd1}}^2 |c_i|^2 + \rho_{\text{rd1}}^2\rho_{\text{rd2}}^2 |c_i|^2 |a_i|^2 + 1\right) \tag{B.98}$$

$$\doteq (T-2)\log\left(\rho_{\text{rd1}}^2 |c_i|^2 + 1\right). \tag{B.99}$$

If $\rho_{\text{rd2}}^2 |a|^2 < \max\left(\rho_{\text{rd1}}^2 |b_i|^2,\ \rho_{\text{rd1}}^2 |c_i|^2\right)$, then setting $|b_i'|^2 = |c_i'|^2 = \left(|b_i|^2 + |c_i|^2\right)/2 = |d_i|^2$, $|a_i|^2 = 0$ decreases f_1 and f_2 by at most a constant independent of SNR and in this case

$$f_1\left(|a_i'|^2 = 0, |b_i'|^2 = |d_i|^2, |c_i'|^2 = |d_i|^2\right)$$

$$\doteq T\log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right)$$

$$- \log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right) \tag{B.100}$$

$$\doteq (T-1)\log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right) \tag{B.101}$$

$$f_2\left(|a_i'|^2 = 0, |b_i'|^2 = |d_i|^2, |c_i'|^2 = |d_i|^2\right) \tag{B.102}$$

$$\doteq \log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right) + (T-1)\log\left(\rho_{\text{rd1}}^2 |d_i|^2 + T - 1\right)$$

$$- \log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right) \tag{B.103}$$

$$\doteq (T-1)\log\left(\rho_{\text{rd1}}^2 |d_i|^2 + 1\right). \tag{B.104}$$

Hence for the following optimization problem $\mathcal{P}_6$ with mass points $\left(|a_i|^2, 0, |c_i|^2\right)$ with probability p_{1i} and mass points $\left(0, |d_i|^2, |d_i|^2\right)$ with probability p_{2i},

$$\mathcal{P}_6 : \begin{cases} \text{maximize } \min\Bigg\{ \sum_{i=1}^{3} p_{1i}\left((T-1)\log\left(\rho_{\text{rd2}}^2 |a_{1i}|^2 + 1\right) - \log\left(\rho_{\text{rd1}}^2 |c_{1i}|^2 + 1\right)\right) \\ \qquad + \sum_{i=1}^{3} p_{2i}\left((T-1)\log\left(\rho_{\text{rd1}}^2 |d_{2i}|^2 + 1\right)\right), \\ \qquad (T-1)\log\left(\rho_{\text{sr2}}^2\right) + \sum_{i=1}^{3} p_{1i}\left((T-2)\log\left(\rho_{\text{rd1}}^2 |c_{1i}|^2 + 1\right)\right) \\ \qquad + \sum_{i=1}^{3} p_{2i}\left((T-1)\log\left(\rho_{\text{rd1}}^2 |d_{2i}|^2 + 1\right)\right) \Bigg\} \\ \sum_{i=1}^{3} p_{1i}\left(|a_{1i}|^2 + |c_{1i}|^2\right) + \sum_{i=1}^{3} 2p_{2i} |d_{2i}|^2 \le 2T \\ |a_{1i}|^2, |c_{1i}|^2, |d_{2i}|^2 \ge 0 \end{cases} \tag{B.105}$$

we have

$$\text{gDoF}(\mathcal{P}_1) = \text{gDoF}(\mathcal{P}_2) = \text{gDoF}(\mathcal{P}_3) = \text{gDoF}(\mathcal{P}_4) = \text{gDoF}(\mathcal{P}_5) = \text{gDoF}(\mathcal{P}_6). \tag{B.106}$$

Now, we claim that multiple mass points of the form $\left(|a_{1i}|^2, 0, |c_{1i}|^2\right)$ with probability p_{1i} can be replaced by a single point $\left(|a_1|^2, 0, |c_1|^2\right)$.

Claim B.2. *There exists c_1 such that $\sum_i p_{1i} \log\left(\rho_{rd1}^2 |c_1|^2 + 1\right) = \sum_i p_{1i} \log\left(\rho_{rd1}^2 |c_{1i}|^2 + 1\right)$ with $\sum_i p_{1i} |c_1|^2 \leq \sum_i p_{1i} |c_{1i}|^2$.*

Proof. We have by Jensen's inequality

$$\sum_i p_{1i} \log\left(\rho_{\text{rd1}}^2 \frac{\sum_j p_{1j} |c_{1j}|^2}{\sum_j p_{1j}} + 1\right) \geq \sum_i p_{1i} \log\left(\rho_{\text{rd1}}^2 |c_{1i}|^2 + 1\right). \tag{B.107}$$

Hence there exists c_1 with

$$|c_1|^2 \leq \frac{\sum_j p_{1j} |c_{1j}|^2}{\sum_j p_{1j}} \tag{B.108}$$

such that

$$\sum_i p_{1i} \log\left(\rho_{\text{rd1}}^2 |c_1|^2 + 1\right) = \sum_i p_{1i} \log\left(\rho_{\text{rd1}}^2 |c_{1i}|^2 + 1\right). \tag{B.109}$$

Also, due to $|c_1|^2 \leq \left(\sum_j p_{1j} |c_{1j}|^2\right) / \left(\sum_j p_{1j}\right)$, we have $\sum_j p_{1j} |c_1|^2 \leq \sum_j p_{1j} |c_{1j}|^2$, hence the power constraint is not violated. □

Hence we reduce $\{c_{1i}\}_{i=1}^3$ to a single point c_1. Similar procedure can be carried out with a_{1i} and d_{2i} and we get

$$\mathcal{P}_7 : \begin{cases} \text{maximize } \min \Big\{ p_1 \left((T-1)\log\left(\rho_{\text{rd2}}^2 |a_1|^2 + 1\right) - \log\left(\rho_{\text{rd1}}^2 |c_1|^2 + 1\right)\right) \\ \qquad + (T-1)\, p_2 \log\left(\rho_{\text{rd1}}^2 |d_2|^2 + 1\right),\ (T-1)\log\left(\rho_{\text{sr2}}^2\right) \\ \qquad + (T-2)\, p_1 \log\left(\rho_{\text{rd1}}^2 |c_1|^2 + 1\right) \\ \qquad + (T-1)\, p_2 \log\left(\rho_{\text{rd1}}^2 |d_2|^2 + 1\right) \Big\} \\ p_1\left(|a_1|^2 + |c_1|^2\right) + 2p_2 |d_2|^2 \leq 2T \\ |a_1|^2, |c_1|^2, |d_2|^2 \geq 0, \end{cases} \tag{B.110}$$

$$\text{gDoF}\left(\mathcal{P}_1\right)=\text{gDoF}\left(\mathcal{P}_2\right)=\cdots=\text{gDoF}\left(\mathcal{P}_6\right)=\text{gDoF}\left(\mathcal{P}_7\right). \tag{B.111}$$

$\mathcal{P}_7$ has $\left(|a_1|^2, 0, |c_1|^2\right)$ with probability p_1 and $\left(0, |d_2|^2, |d_2|^2\right)$ with probability p_2. Since a constant power scaling does not affect gDoF for the problem, with $\mathcal{P}_8$ defined as

$$\mathcal{P}_8 : \begin{cases} \text{maximize } \min\Big\{ p_1\left((T-1)\log\left(\rho_{\text{rd2}}^2|a_1|^2+1\right)-\log\left(\rho_{\text{rd1}}^2|c_1|^2+1\right)\right) \\ \qquad\qquad +(T-1)\,p_2\log\left(\rho_{\text{rd1}}^2|d_2|^2+1\right),\ (T-1)\log\left(\rho_{\text{sr2}}^2\right) \\ \qquad\qquad +(T-2)\,p_1\log\left(\rho_{\text{rd1}}^2|c_1|^2+1\right) \\ \qquad\qquad +(T-1)\,p_2\log\left(\rho_{\text{rd1}}^2|d_2|^2+1\right)\Big\} \\ p_1|a_1|^2\le T, p_1|c_1|^2\le T, p_2|d_1|^2\le T/2 \\ |a_1|^2, |c_1|^2, |d_2|^2\ge 0, \end{cases} \tag{B.112}$$

we can show that

$$\text{gDoF}\left(\mathcal{P}_1\right)=\text{gDoF}\left(\mathcal{P}_2\right)=\cdots=\text{gDoF}\left(\mathcal{P}_7\right)=\text{gDoF}\left(\mathcal{P}_8\right). \tag{B.113}$$

Now, with $p_1|a_1|^2\le T$,

$$\begin{aligned} p_1(T-1)\log\left(\rho_{\text{rd2}}^2|a_1|^2+1\right) &\le p_1(T-1)\log\left(\rho_{\text{rd2}}^2\frac{T}{p_1}+1\right) \\ &= p_1(T-1)\log\left(\rho_{\text{rd2}}^2T+p_1\right)-p_1(T-1)\log\left(p_1\right) \\ &\overset{(i)}{\le} p_1(T-1)\log\left(\rho_{\text{rd2}}^2T+1\right)+(T-1)\frac{\log(e)}{e}, \end{aligned} \tag{B.114}$$

where (i) is using $-p_1\log(p_1)\le\log(e)/e$. Hence it suffices to use $|a_1|^2\le T$ for the optimal value without losing gDoF. Choosing a larger value does not improve gDoF due to (B.114). Similarly keeping $|c_1|^2\le T$, $|d_1|^2\le T/2$ is sufficient to achieve the gDoF. Note that for $\mathcal{P}_8$ the objective function is increasing in $|a_1|^2$, $|d_2|^2$. Hence by choosing $|a_1|^2=T$, $|d_2|^2=T/2$, we get a gDoF-optimal solution. Hence by choosing $|a_1|^2=T$, $|d_2|^2=T/2$ and including the extra constraint $|c_1|^2\le T$ (which makes the constraint $p_1|c_1|^2\le T$ inactive), and also

$\rho_{\mathrm{rd}i}^2 = \mathsf{SNR}^{\gamma_{\mathrm{rd}i}}$, $\rho_{\mathrm{sr}i}^2 = \mathsf{SNR}^{\gamma_{\mathrm{sr}i}}$, we obtain an equivalent optimization problem:

$$\mathcal{P}_9 : \begin{cases} \text{maximize } \min \Big\{ p_1 \left((T-1)\,\gamma_{\mathrm{rd}2} \log(\mathsf{SNR}) - \log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd}1}} |c_1|^2 + 1\right)\right) \\ \qquad\qquad\qquad + (T-1)\, p_2 \gamma_{\mathrm{rd}1} \log(\mathsf{SNR}),\ (T-1)\,\gamma_{\mathrm{sr}2} \log(\mathsf{SNR}) \\ \qquad\qquad\qquad + (T-2)\, p_1 \log\left(\mathsf{SNR}^{\gamma_{\mathrm{rd}1}} |c_1|^2 + 1\right) \\ \qquad\qquad\qquad + (T-1)\, p_2 \gamma_{\mathrm{rd}1} \log(\mathsf{SNR}) \Big\} \\ |c_1|^2 \le T, p_1 + p_2 = 1, |c_1|^2 \ge 0. \end{cases} \tag{B.115}$$

We relabel $p_1 = p_\lambda$, $p_2 = 1 - p_\lambda$ and complete the proof.

B.5 Proof of Theorem 3.14

Following the notation from the statement of Theorem 3.14 on page 80, we can equivalently use

$$X = \begin{bmatrix} \begin{bmatrix} \alpha_1 & 0 & 0 & . & . & 0 \end{bmatrix} Q_1 \\ \begin{bmatrix} \alpha_2 & 0 & 0 & . & . & 0 \end{bmatrix} Q_2 \end{bmatrix} \tag{B.116}$$

where Q_1, Q_2 are independent $T \times T$ isotropically distributed unitary matrices and α_1, α_2 are chosen independently as

$$\alpha_1 \sim a_1 \sqrt{\frac{1}{2}\chi^2(2T)}, \tag{B.117}$$

$$\alpha_1 \sim a_2 \sqrt{\frac{1}{2}\chi^2(2T)}, \tag{B.118}$$

where $\chi^2(k)$ is chi-squared distributed. This choice will induce $\begin{bmatrix} \alpha_i & 0 & 0 & . & . & 0 \end{bmatrix} Q_i = \alpha_i q_i$ to be T dimensional random vectors with i.i.d. $a_i \mathcal{CN}(0,1)$ components, where q_i are T dimensional unitary isotropic distributed row vectors (see Section 3.4.1.1 on page 67 for details on chi-squared distribution).

With this choice we have

$$\mathbb{E}\left[Y^\dagger Y \middle| X_1, X_2 \right] \overset{(i)}{=} Q_1^\dagger K_1 Q_1 + Q_2^\dagger K_2 Q_2 + I_{T\times T} \tag{B.119}$$

$$h\left(Y|\,X\right) \doteq \mathbb{E}\left[\log\left(\det\left(Q_1^\dagger K_1 Q_1 + Q_2^\dagger K_2 Q_2 + I_{T\times T}\right)\right)\right] \tag{B.120}$$

$$\overset{(ii)}{=} \mathbb{E}\left[\log\left(\det\left(K_1 + Q_2^\dagger K_2 Q_2 + I_{T\times T}\right)\right)\right], \tag{B.121}$$

where in step (i) we have

$$K_1 = \begin{bmatrix} \rho_{11}^2\left|\alpha_1\right|^2 & 0 & 0 & . & . & 0 \\ 0 & 0 & & & & . \\ . & & . & & & . \\ 0 & . & & . & . & 0 \end{bmatrix}, \ K_2 = \begin{bmatrix} \rho_{12}^2\left|\alpha_2\right|^2 & 0 & 0 & . & . & 0 \\ 0 & 0 & & & & . \\ . & & . & & & . \\ 0 & . & & . & . & 0 \end{bmatrix}$$

and in step (ii) Q_1 is absorbed using properties of determinants and unitary matrices. Now,

$$\Delta = \det\left(K_1 + Q_2^\dagger K_2 Q_2 + I_{T\times T}\right) \tag{B.122}$$

$$\overset{(i)}{=} \rho_{11}^2\left|\alpha_1\right|^2 \det\left(\text{Cofactor}\left(Q_2^\dagger K_2 Q_2 + I_{T\times T}, 1, 1\right)\right) + \det\left(Q_2^\dagger K_2 Q_2 + I_{T\times T}\right) \tag{B.123}$$

$$= \rho_{11}^2\left|\alpha_1\right|^2 \det\left(\text{Cofactor}\left(Q_2^\dagger K_2 Q_2 + I_{T\times T}, 1, 1\right)\right) + \rho_{12}^2\left|\alpha_2\right|^2 + 1, \tag{B.124}$$

where (i) is due of the structure of K_1 and the property of determinants. Now, with q_2 being the first row of Q_2 (q_2 will be an isotropically distributed unit vector), we get

$$Q_2^\dagger K_2 Q_2 = q_2^\dagger\left(\rho_{12}^2\left|\alpha_2\right|^2 q_2\right). \tag{B.125}$$

Hence

$$\text{Cofactor}\left(Q_2^\dagger K_2 Q_2 + I_{T\times T}, 1, 1\right) = \eta_2^\dagger\left(\rho_{12}^2\left|\alpha_2\right|^2 \eta_2\right) + I_{(T-1)\times(T-1)}, \tag{B.126}$$

where η_1 is the row vector formed with the last $T-1$ components of q_2. And hence

$$\det\left(\text{Cofactor}\left(Q_2^\dagger K_2 Q_2 + I_{T\times T}, 1, 1\right)\right) = \det\left(\eta_2^\dagger\left(\rho_{12}^2\left|\alpha_2\right|^2 \eta_2\right) + I_{(T-1)\times(T-1)}\right) \tag{B.127}$$

$$= \rho_{12}^2\left|\alpha_2\right|^2 \eta_2\eta_2^\dagger + 1, \tag{B.128}$$

where the last step followed due to matrix theory results on determinants of matrices of the form (identity+column·row). Hence

$$
\begin{aligned}
\Delta &= \rho_{11}^2 \left|\alpha_1\right|^2 + \rho_{21}^2 \left|\alpha_2\right|^2 + \rho_{11}^2 \left|\alpha_1\right|^2 \rho_{21}^2 \left|\alpha_2\right|^2 \eta_2 \eta_2^\dagger + 1 && \text{(B.129)}\\
h\left(Y|\, X\right) &\doteq \mathbb{E}\left[\log\left(\rho_{11}^2 \left|\alpha_1\right|^2 + \rho_{21}^2 \left|\alpha_2\right|^2 + \rho_{11}^2 \left|\alpha_1\right|^2 \rho_{21}^2 \left|\alpha_2\right|^2 \eta_2 \eta_2^\dagger + 1\right)\right] && \text{(B.130)}\\
&\overset{(i)}{\leq} \mathbb{E}\left[\log\left(\rho_{11}^2 \left|\alpha_1\right|^2 + \rho_{21}^2 \left|\alpha_2\right|^2 + \rho_{11}^2 \left|\alpha_1\right|^2 \rho_{21}^2 \left|\alpha_2\right|^2 + 1\right)\right] && \text{(B.131)}\\
&= \mathbb{E}\left[\log\left(\left(1 + \rho_{11}^2 \left|\alpha_1\right|^2\right)\left(1 + \rho_{21}^2 \left|\alpha_2\right|^2\right)\right)\right] && \text{(B.132)}\\
&\overset{(ii)}{\doteq} \log\left(\left(1 + \rho_{11}^2 \left|a_1\right|^2\right)\left(1 + \rho_{21}^2 \left|a_2\right|^2\right)\right), && \text{(B.133)}
\end{aligned}
$$

where (i) followed since $\eta_2\eta_2^\dagger \leq 1$ because η_2 was a subvector of a unit vector, (ii) is because $\alpha_i \sim a_i\sqrt{\frac{1}{2}\chi^2\left(2T\right)}$ and using Fact 3.3 for chi-squared distributed random variables. Hence

$$
h\left(Y|\, X\right) \dot{\leq} \log\left(\left(1 + \rho_{11}^2 \left|a_1\right|^2\right)\left(1 + \rho_{21}^2 \left|a_2\right|^2\right)\right). \tag{B.134}
$$

B.6 Proof of Lemma 3.4

In this appendix, we prove that $\log\left(\mathbb{E}\left[\left|w\right|^2 / \left(1 + \left|g+w\right|^2\right)\right]\right) \dot{\leq} \log\left(1/\rho^2\right)$. We have

$$
\begin{aligned}
\mathbb{E}\left[\frac{\left|w\right|^2}{1 + \left|g+w\right|^2}\right] &\overset{(i)}{=} \mathbb{E}\left[\frac{\left|w\right|^2}{1 + \left|w\right|^2 + \left|g\right|^2 + 2\left|w\right|\left|g\right|\cos\left(\theta\right)}\right] && \text{(B.135)}\\
&\overset{(ii)}{=} \mathbb{E}\left[\frac{2\pi\left|w\right|^2}{\sqrt{1 + 2\left(\left|w\right|^2 + \left|g\right|^2\right) + \left(\left|w\right|^2 - \left|g\right|^2\right)^2}}\right] && \text{(B.136)}\\
&\leq \mathbb{E}\left[\frac{2\pi\left|w\right|^2}{\sqrt{1 + \left(\left|w\right|^2 - \left|g\right|^2\right)^2}}\right], && \text{(B.137)}
\end{aligned}
$$

where (i) is using the property of independent circularly symmetric Gaussians w, g to introduce θ uniformly distributed in $[0, 2\pi]$ independent of $\left|w\right|, \left|g\right|$ and (ii) is using the Tower property of expectation and by integrating over θ (integration can be easily verified in Mathematica).

Hence

$$\mathbb{E}\left[\frac{1}{2\pi}\frac{|w|^2}{1+|g+w|^2}\right] \leq \mathbb{E}\left[\frac{|w|^2}{\sqrt{1+\left(|w|^2-|g|^2\right)^2}}\right] \tag{B.138}$$

$$\begin{aligned}
&\leq \mathbb{E}\left[\frac{|w|^2}{|g|^2-|w|^2}\mathbb{1}_{\left\{|g|^2>|w|^2+1\right\}}\right] \\
&\quad + \mathbb{E}\left[\frac{|w|^2}{|w|^2-|g|^2}\mathbb{1}_{\left\{|w|^2>|g|^2+1\right\}}\right] \\
&\quad + \mathbb{E}\left[|w|^2\,\mathbb{1}_{\left\{\left||w|^2-|g|^2\right|\leq 1\right\}}\right]
\end{aligned} \tag{B.139}$$

$$\begin{aligned}
&\overset{(i)}{=} \frac{\rho^2\cdot\Gamma\left(0,\frac{1}{\rho^2}\right)}{(\rho^2+1)^2} + \mathbb{E}\left[\frac{|w|^2}{|w|^2-|g|^2}\mathbb{1}_{\left\{|w|^2>|g|^2+1\right\}}\right] \\
&\quad + \frac{-e^{-1/\rho^2}\rho^4+\rho^4-\frac{3\rho^2}{e}+2\rho^2-\frac{2}{e}+1}{(\rho^2+1)^2}
\end{aligned} \tag{B.140}$$

$$\begin{aligned}
&\overset{(ii)}{\leq} \frac{\rho^2 e^{-\frac{1}{\rho^2}}\ln\left(1+\rho^2\right)}{(\rho^2+1)^2} + \mathbb{E}\left[\frac{|w|^2}{|w|^2-|g|^2}\mathbb{1}_{\left\{|w|^2>|g|^2+1\right\}}\right] \\
&\quad + \frac{-e^{-\frac{1}{\rho^2}}\rho^4+\rho^4-\frac{3\rho^2}{e}+2\rho^2-\frac{2}{e}+1}{(\rho^2+1)^2},
\end{aligned} \tag{B.141}$$

where (i) is obtained by evaluating $\mathbb{E}\left[\frac{|w|^2}{|g|^2-|w|^2}\mathbb{1}_{\{|g|^2>|w|^2+1\}}\right]$ and $\mathbb{E}\left[|w|^2\,\mathbb{1}_{\{||w|^2-|g|^2|<1\}}\right]$ (integration can be easily verified in Mathematica). Also, $\Gamma(0,x)$ is the incomplete gamma function, (ii) is using the inequality $\Gamma(0,x)=E_1(x)\leq e^{-x}\ln(1+1/x)$, where $E_1(x)=\int_x^\infty \frac{e^{-t}}{t}dt$ is the exponential integral.

Now,

$$\mathbb{E}\left[\frac{|w|^2}{|w|^2-|g|^2}\mathbb{1}_{\left\{|w|^2>|g|^2+1\right\}}\right] = \int_{s=0}^{\infty}\left(\int_{r=s+1}^{\infty}\frac{r}{r-s}e^{-r}\frac{1}{\rho^2}e^{-\frac{s}{\rho^2}}dr\right)ds \tag{B.142}$$

$$= \int_{s=0}^{\infty}\frac{1}{\rho^2}e^{-\frac{s}{\rho^2}}\left(\int_{r=s+1}^{\infty}e^{-r}dr+\int_{r=s+1}^{\infty}\frac{s}{r-s}e^{-r}dr\right)ds \tag{B.143}$$

$$= \int_{s=0}^{\infty}\frac{1}{\rho^2}e^{-\frac{s}{\rho^2}}\left(e^{-s-1}+\int_{r=s+1}^{\infty}\frac{se^{-s}}{r-s}e^{-r+s}dr\right)ds \tag{B.144}$$

$$\overset{(i)}{=} \int_{s=0}^{\infty}\frac{1}{\rho^2}e^{-\frac{s}{\rho^2}}\left(e^{-s-1}+se^{-s}E_1(1)\right)ds \tag{B.145}$$

$$= \frac{1}{1+\rho^2} - \frac{\rho^2 E_1(1)}{(\rho^2+1)^2}, \tag{B.146}$$

where (i) is using change of variables and the formula for Exponential integral $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$. Also, $E_1(1) \approx 0.219384$.

Hence it follows that

$$\mathbb{E}\left[\frac{|w|^2}{1+|g+w|^2}\right] \leq \frac{\rho^2 e^{-\frac{1}{\rho^2}} \ln(1+\rho^2)}{(\rho^2+1)^2} + \frac{1}{1+\rho^2} - \frac{\rho^2 E_1(1)}{(\rho^2+1)^2} + \frac{-e^{-\frac{1}{\rho^2}}\rho^4 + \rho^4 - \frac{3\rho^2}{e} + 2\rho^2 - \frac{2}{e} + 1}{(\rho^2+1)^2} \tag{B.147}$$

and hence

$$\log\left(\mathbb{E}\left[\frac{|w|^2}{1+|g+w|^2}\right]\right) \dot{\leq} \log\left(\frac{1}{\rho^2}\right). \tag{B.148}$$

APPENDIX C

Proofs for Chapter 4

C.1 Proof of achievability for non-feedback case

We evaluate the term in the first inner bound inequality (5.5a). The other terms can be similarly evaluated.

$$I\left(X_1; Y_1, \underline{g_1} | U_2\right) \overset{(a)}{=} I\left(X_1; Y_1 | U_2, \underline{g_1}\right) \tag{C.1}$$

$$= h\left(Y_1 | U_2, \underline{g_1}\right)$$

$$- h\left(Y_1 | X_1, U_2, \underline{g_1}\right), \tag{C.2}$$

$$h\left(Y_1 | U_2, \underline{g_1}\right) = h\left(g_{11}X_1 + g_{21}X_2 + Z_1 | U_2, \underline{g_1}\right) \tag{C.3}$$

$$= h\left(g_{11}X_1 + g_{21}X_{p2} + Z_1 | \underline{g_1}\right), \tag{C.4}$$

$$\text{variance}\left(g_{11}X_1 + g_{21}X_{p2} + Z_1 | \underline{g_1}\right) = |g_{11}|^2 + \lambda_{p2} |g_{21}|^2 + 1,$$

$$\therefore h\left(Y_1 | U_2, \underline{g_1}\right) = \mathbb{E}\left[\log\left(|g_{11}|^2 + \lambda_{p2} |g_{21}|^2 + 1\right)\right]$$

$$+ \log(2\pi e), \tag{C.5}$$

$$h\left(Y_1 | X_1, U_2, \underline{g_1}\right)$$

$$= h\left(g_{11}X_1 + g_{21}X_2 + Z_1 | X_1, U_2, \underline{g_1}\right) \tag{C.6}$$

$$= h\left(g_{21}X_{p2} + Z_1 | \underline{g_1}\right) \tag{C.7}$$

$$= \mathbb{E}\left[\log\left(1 + \lambda_{p2} |g_{21}|^2\right)\right] + \log(2\pi e)$$

$$\overset{(b)}{\leq} \mathbb{E}\left[\log\left(1 + |g_{21}|^2 / INR_2\right)\right] + \log(2\pi e) \tag{C.8}$$

$$\overset{(c)}{\leq} \log(2) + \log(2\pi e) \tag{C.9}$$

$$\therefore I\left(X_1;Y_1,\underline{g_1}|U_2\right) \geq \mathbb{E}\left[\log\left(|g_{11}|^2+\lambda_{p2}|g_{21}|^2+1\right)\right] \\ -1,$$

where (a) uses independence, (b) is because $\lambda_{pi} \leq \frac{1}{INR_i}$, and (c) follows from Jensen's inequality.

C.2 Proof of outer bounds for non-feedback case

Note that we have the notation $\underline{g} = [g_{11}, g_{21}, g_{22}, g_{12}]$, $S_1 = g_{12}X_1 + Z_2$, and $S_2 = g_{21}X_2 + Z_1$. Our outer bounding steps are valid while allowing X_{1i} to be a function of $\left(W_1, \underline{g^n}\right)$, thus letting transmitters have instantaneous and future CSIT. On choosing a uniform distribution of messages we get

$$
\begin{aligned}
& n(R_1+2R_2-\epsilon_n) \\
& \leq I\left(W_1;Y_1^n,S_1^n,\underline{g^n}\right)+I\left(W_2;Y_2^n,\underline{g^n}\right) \\
& \quad +I\left(W_2;Y_2^n,S_2^n,X_1^n,\underline{g^n}\right) && \text{(C.10)} \\
& = I\left(W_1;Y_1^n,S_1^n|\,\underline{g^n}\right)+I\left(W_2;Y_2^n|\,\underline{g^n}\right) \\
& \quad +I\left(W_2;Y_2^n,S_2^n|\,X_1^n,\underline{g^n}\right) && \text{(C.11)} \\
& = I\left(W_1;S_1^n|\,\underline{g^n}\right)+I\left(W_1;Y_1^n|\,S_1^n,\underline{g^n}\right)+I\left(W_2;Y_2^n|\,\underline{g^n}\right) \\
& \quad +I\left(W_2;S_2^n|\,X_1^n,\underline{g^n}\right)+I\left(W_2;Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right) \\
& = h\left(S_1^n|\,\underline{g^n}\right)-h\left(S_1^n|\,W_1,\underline{g^n}\right)+h\left(Y_1^n|\,S_1^n,\underline{g^n}\right) \\
& \quad -h\left(Y_1^n|\,W_1,S_1^n,\underline{g^n}\right)+h\left(Y_2^n|\,\underline{g^n}\right)-h\left(Y_2^n|\,W_2,\underline{g^n}\right) \\
& \quad +h\left(S_2^n|\,X_1^n,\underline{g^n}\right)-h\left(S_2^n|\,X_1^n,W_2,\underline{g^n}\right) \\
& \quad +h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right)-h\left(Y_2^n|\,X_1^n,W_2,S_2^n,\underline{g^n}\right) && \text{(C.12)} \\
& = h\left(S_1^n|\,\underline{g^n}\right)-h\left(Z_2^n\right)+h\left(Y_1^n|\,S_1^n,\underline{g^n}\right)-h\left(S_2^n|\,\underline{g^n}\right) \\
& \quad +h\left(Y_2^n|\,\underline{g^n}\right)-h\left(S_1^n|\,\underline{g^n}\right)+h\left(S_2^n|\,\underline{g^n}\right) \\
& \quad -h\left(Z_1^n\right)+h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right)-h\left(Z_2^n\right) && \text{(C.13)} \\
& = h\left(Y_1^n|\,S_1^n,\underline{g^n}\right)+h\left(Y_2^n|\,\underline{g^n}\right)+h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right)
\end{aligned}
$$

$$
\begin{aligned}
&- h\left(Z_1^n\right) - 2h\left(Z_2^n\right) && \text{(C.14)}\\
&\overset{(a)}{\leq} \sum \left[h\left(Y_{1i}|S_{1i},\underline{g^n}\right) - h\left(Z_{1i}\right)\right] + \sum \left[h\left(Y_{2i}|\underline{g^n}\right) - h\left(Z_{2i}\right)\right]\\
&\quad + \sum \left[h\left(Y_{2i}|X_{1i},S_{2i},\underline{g^n}\right) - h\left(Z_{2i}\right)\right] && \text{(C.15)}\\
&= \mathbb{E}_{\underline{g^n}}\left[\sum \left(h\left(Y_{1i}|S_{1i},\underline{g^n}\right) - h\left(Z_{1i}\right)\right)\right]\\
&\quad + \mathbb{E}_{\underline{g^n}}\left[\sum \left(h\left(Y_{2i}|\underline{g^n}\right) - h\left(Z_{2i}\right)\right)\right]\\
&\quad + \mathbb{E}_{\underline{g^n}}\left[\sum \left(h\left(Y_{2i}|X_{1i},S_{2i},\underline{g^n}\right) - h\left(Z_{2i}\right)\right)\right] && \text{(C.16)}\\
&\overset{(b)}{\leq} n\mathbb{E}\left[\log\left(1+|g_{21}|^2+|g_{11}|^2/\left(1+|g_{12}|^2\right)\right)\right]\\
&\quad + n\mathbb{E}\left[\log\left(1+|g_{12}|^2+|g_{22}|^2\right)\right]\\
&\quad + n\mathbb{E}\left[\log\left(1+|g_{22}|^2/\left(1+|g_{21}|^2\right)\right)\right], && \text{(C.17)}
\end{aligned}
$$

where (a) is due to the fact that conditioning reduces entropy and (b) follows from Equations [ETW08, (50)] , [ETW08, (51)] and [ETW08, (52)]. Note that in the calculation of step (b) we allow the symbols X_{1i}, X_{2i} to depend on $\underline{g^n}$, but since $\underline{g^n}$ is available in conditioning the calculation proceeds similar to that in [ETW08].

$$
\begin{aligned}
&n(R_1+R_2-\epsilon_n) && \text{(C.18)}\\
&\leq I\left(W_1;Y_1^n,\underline{g^n}\right) + I\left(W_2;Y_2^n,S_2^n,X_1^n,\underline{g^n}\right) && \text{(C.19)}\\
&= I\left(W_1;Y_1^n|\,\underline{g^n}\right) + I\left(W_2;Y_2^n,S_2^n|\,X_1^n,\underline{g^n}\right) && \text{(C.20)}\\
&= I\left(W_1;Y_1^n|\,\underline{g^n}\right) + I\left(W_2;S_2^n|\,X_1^n,\underline{g^n}\right)\\
&\quad + I\left(W_2;Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right) && \text{(C.21)}\\
&= h\left(Y_1^n|\,\underline{g^n}\right) - h\left(Y_1^n|\,W_1,\underline{g^n}\right)\\
&\quad + h\left(S_2^n|\,X_1^n,\underline{g^n}\right) - h\left(S_2^n|\,X_1^n,W_2,\underline{g^n}\right)\\
&\quad + h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right) - h\left(Y_2^n|\,X_1^n,W_2,S_2^n,\underline{g^n}\right) && \text{(C.22)}\\
&= h\left(Y_1^n|\,\underline{g^n}\right) - h\left(S_2^n|\,\underline{g^n}\right) + h\left(S_2^n|\,\underline{g^n}\right)\\
&\quad - h\left(Z_1^n\right) + h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right) - h\left(Z_2^n\right) && \text{(C.23)}\\
&= h\left(Y_1^n|\,\underline{g^n}\right) + h\left(Y_2^n|\,X_1^n,S_2^n,\underline{g^n}\right) - h\left(Z_1^n\right) - h\left(Z_2^n\right) && \text{(C.24)}\\
&\overset{(a)}{\leq} \sum \left[h\left(Y_{1i}|\underline{g^n}\right) - h\left(Z_{1i}\right)\right]\\
&\quad + \sum \left[h\left(Y_{2i}|X_{1i},S_{2i},\underline{g^n}\right) - h\left(Z_{2i}\right)\right] && \text{(C.25)}
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{\underline{g}^n}\left[\sum \left(h\left(Y_{1i}|\underline{g}^n\right) - h\left(Z_{1i}\right)\right)\right] \\
&\quad + \mathbb{E}_{\underline{g}^n}\left[\sum \left(h\left(Y_{2i}|X_{1i}, S_{2i}, \underline{g}^n\right) - h\left(Z_{2i}\right)\right)\right] && \text{(C.26)} \\
&\overset{(b)}{\leq} n\mathbb{E}\left[\log\left(1 + |g_{21}|^2 + |g_{11}|^2\right)\right] \\
&\quad + n\mathbb{E}\left[\log\left(1 + |g_{22}|^2 / \left(1 + |g_{21}|^2\right)\right)\right], && \text{(C.27)}
\end{aligned}
$$

where (a) is due to the fact that conditioning reduces entropy and (b) again follows from Equations [ETW08, (51)] and [ETW08, (52)].

C.3 Proof of Lemma 4.2

We have $F(w) \leq aw^b$ for $w \in [0, \epsilon]$, where $a \geq 0, b > 0, 1 \geq \epsilon > 0$. Now using integration by parts we get

$$
\begin{aligned}
\mathbb{E}\left[\ln(W)\right] &\geq \int_0^1 f(w)\ln(w)\,dw && \text{(C.28)} \\
&= \int_0^\epsilon f(w)\ln(w)\,dw + \int_\epsilon^1 f(w)\ln(w)\,dw && \text{(C.29)} \\
&= \left[F(w)\ln(w)\right]_0^\epsilon - \int_0^\epsilon F(w)\frac{1}{w}dw + \int_\epsilon^1 f(w)\ln(w)\,dw && \text{(C.30)} \\
&\geq \left[aw^b\ln(w)\right]_0^\epsilon - \int_0^\epsilon aw^b\frac{1}{w}dw + \ln(\epsilon) && \text{(C.31)} \\
&\geq a\epsilon^b\ln(\epsilon) - \frac{a\epsilon^b}{b} + \ln(\epsilon). && \text{(C.32)}
\end{aligned}
$$

Note that $\ln(w)$ is negative in the range $[0, 1)$, thus we get the desired inequalities in the last two steps.

C.4 Proof of Corollary 4.9

The rate region of non-feedback case in given in Equation (4.13) can be reduced to the rate region for a channel without fading. Let $\mathcal{R}'_{NFB}$ be the approximately optimal Han-Kobayashi rate region of IC [ETW08] with equivalent channel strengths $SNR_i := \mathbb{E}\left[|g_{ii}|^2\right]$ for $i = 1, 2$, and $INR_i := \mathbb{E}\left[|g_{ij}|^2\right]$ for $i \neq j$. Then for a constant c' we have

$$\mathcal{R}'_{NFB} \supseteq \mathcal{R}_{NFB} \supseteq \mathcal{R}'_{NFB} - c'. \tag{C.33}$$

This can be verified by proceeding through each inner bound equation. For example, consider the first inner bound Equation (4.13a) $R_1 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 1$. The corresponding equation in $\mathcal{R}'_{NFB}$ is $R_1 \leq \log\left(1 + SNR_1 + \lambda_{p2}INR_1\right) - 1$. Now

$$\log\left(1 + SNR_1 + \lambda_{p2}INR_1\right) - 1 \overset{(a)}{\geq} \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 1 \tag{C.34}$$

$$\overset{(b)}{\geq} \left(\log\left(1 + SNR_1 + \lambda_{p2}INR_1\right) - 1\right) - 2c_{JG}, \tag{C.35}$$

where (a) is due to Jensen's inequality and (b) is using logarithmic Jensen's gap result twice. Due to (C.34), (C.35) it follows that the first inner bound equation for fading case is in constant gap with that of static case. Similarly, by proceeding through each inner bound equation, it follows that $\mathcal{R}'_{NFB} \supseteq \mathcal{R}_{NFB} \supseteq \mathcal{R}'_{NFB} - c'$ for a constant c'.

C.5 Proof of Theorem 4.10 (Fast fading interference multiple access channel)

We have the following achievable rate region for from fast fading interference multiple access channel, by using the scheme from [PDT09] by considering $\left(Y_1, \underline{g_1}\right)$, $\left(Y_2, \underline{g_2}\right)$ as the outputs at the receivers.

$$R_1 \leq I\left(X_1; Y_1, \underline{g_1} | X_2\right) \tag{C.36}$$

$$R_2 \leq I\left(X_2; Y_2, \underline{g_2} | U_1\right) \tag{C.37}$$

$$R_2 \leq I\left(X_2; Y_1, \underline{g_2} | X_1\right) \tag{C.38}$$

$$R_1 + R_2 \leq I\left(X_1 X_2; Y_1, \underline{g_1}\right) \tag{C.39}$$

$$R_1 + R_2 \leq I\left(X_2, U_1; Y_2, \underline{g_2}\right) + I\left(X_1; Y_1, \underline{g_1} | U_1, X_2\right) \tag{C.40}$$

$$R_1 + 2R_2 \leq I\left(X_2, U_1; Y_2, \underline{g_2}\right) + I\left(X_1, X_2; Y_1, \underline{g_1} | U_1\right) \tag{C.41}$$

with mutually independent Gaussian input distributions U_1, X_{p1}, X_2

$$U_1 \sim \mathcal{CN}\left(0, \lambda_{c1}\right), \quad X_{p1} \sim \mathcal{CN}\left(0, \lambda_{p1}\right), \tag{C.42}$$

$$X_1 = U_1 + X_{p1}, \quad X_2 \sim \mathcal{CN}(0,1), \tag{C.43}$$

where $\lambda_{c1} + \lambda_{p1} = 1$ and $\lambda_{p1} = \min\left(\frac{1}{INR_1}, 1\right)$. Evaluating the achievable region we obtain

$$R_1 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2\right)\right] \tag{C.44}$$

$$R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + \lambda_{p1}|g_{12}|^2\right)\right] - 1 \tag{C.45}$$

$$R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{21}|^2\right)\right] \tag{C.46}$$

$$R_1 + R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + |g_{21}|^2\right)\right] \tag{C.47}$$

$$R_1 + R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2\right)\right] - 1 \tag{C.48}$$

$$R_1 + 2R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + |g_{21}|^2\right)\right] - 1. \tag{C.49}$$

The calculations are similar to that of FF-IC (subsection 4.4.2 on page 107). Now we claim the following outer bounds.

$$R_1 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2\right)\right] \tag{C.50}$$

$$R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2\right)\right] \tag{C.51}$$

$$R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{21}|^2\right)\right] \tag{C.52}$$

$$R_1 + R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{11}|^2 + |g_{21}|^2\right)\right] \tag{C.53}$$

$$R_1 + R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{|g_{11}|^2}{1 + |g_{12}|^2}\right)\right] \tag{C.54}$$

$$R_1 + 2R_2 \leq \mathbb{E}\left[\log\left(1 + |g_{22}|^2 + |g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{|g_{11}|^2}{1 + |g_{12}|^2} + |g_{21}|^2\right)\right]. \tag{C.55}$$

With the above outer bounds it can be shown that the capacity region can be achieved within $1 + \frac{1}{2}c_{JG}$ bits per channel use. The computations are similar to that of FF-IC (claim 4.2 on page 4.2).

The outer bound (C.50) can be derived by giving side information W_2 at Rx1 and requiring W_1 to be decoded, outer bound (C.51) can be derived by giving side information W_1 at Rx2 and requiring W_2 to be decoded, outer bound (C.52) can be derived by giving side information W_1 at Rx1 and requiring W_2 to be decoded, outer bound (C.53) can be derived by giving side information requiring W_1, W_2 to be decoded at Rx1 with no side information. We derive (C.54) below. Note that we have the notation $\underline{g} = [g_{11}, g_{21}, g_{22}, g_{12}]$, $S_1 = g_{12}X_1 + Z_2$, and $S_2 = g_{21}X_2 + Z_1$.

$$
\begin{aligned}
& n(R_2 + R_1 - \epsilon_n) && \text{(C.56)}\\
& \leq I\left(W_2; Y_2^n, \underline{g^n}\right) + I\left(W_1; Y_1^n, S_1^n, X_2^n, \underline{g^n}\right) && \text{(C.57)}\\
& = I\left(W_2; Y_2^n \middle| \underline{g^n}\right) + I\left(W_1; Y_1^n, S_1^n \middle| X_2^n, \underline{g^n}\right) && \text{(C.58)}\\
& = I\left(W_2; Y_2^n \middle| \underline{g^n}\right) + I\left(W_1; S_1^n \middle| X_2^n, \underline{g^n}\right) + I\left(W_1; Y_1^n \middle| X_2^n, S_1^n, \underline{g^n}\right) && \text{(C.59)}\\
& = h\left(Y_2^n \middle| \underline{g^n}\right) - h\left(Y_2^n \middle| W_2, \underline{g^n}\right) + h\left(S_1^n \middle| X_2^n, \underline{g^n}\right) - h\left(S_1^n \middle| X_2^n, W_1, \underline{g^n}\right) \\
& \qquad + h\left(Y_1^n \middle| X_2^n, S_1^n, \underline{g^n}\right) - h\left(Y_1^n \middle| X_2^n, W_1, S_1^n, \underline{g^n}\right) && \text{(C.60)}\\
& = h\left(Y_2^n \middle| \underline{g^n}\right) - h\left(S_1^n \middle| \underline{g^n}\right) + h\left(S_1^n \middle| \underline{g^n}\right) - h\left(Z_2^n\right) + h\left(Y_1^n \middle| X_2^n, S_1^n, \underline{g^n}\right) - h\left(Z_1^n\right) && \text{(C.61)}\\
& = h\left(Y_2^n \middle| \underline{g^n}\right) + h\left(Y_1^n \middle| X_2^n, S_1^n, \underline{g^n}\right) - h\left(Z_2^n\right) - h\left(Z_1^n\right) && \text{(C.62)}\\
& \overset{(b)}{\leq} \sum \left[h\left(Y_{2i} | \underline{g^n}\right) - h\left(Z_{2i}\right)\right] + \sum \left[h\left(Y_{1i} | X_{2i}, S_{1i}, \underline{g^n}\right) - h\left(Z_{1i}\right)\right] && \text{(C.63)}\\
& = \mathbb{E}_{\underline{g^n}}\left[\sum \left(h\left(Y_{2i} | \underline{g^n}\right) - h\left(Z_{2i}\right)\right)\right] + \mathbb{E}_{\underline{g^n}}\left[\sum \left(h\left(Y_{1i} | X_{2i}, S_{1i}, \underline{g^n}\right) - h\left(Z_{1i}\right)\right)\right] && \text{(C.64)}\\
& \overset{(c)}{\leq} n\mathbb{E}\left[\log\left(1 + |g_{12}|^2 + |g_{22}|^2\right)\right] + n\mathbb{E}\left[\log\left(1 + \frac{|g_{11}|^2}{1 + |g_{12}|^2}\right)\right], && \text{(C.65)}
\end{aligned}
$$

where (a) is because $\left(X_2^n, \underline{g^n}\right)$ is independent of W_1, (b) is due to the fact that conditioning reduces entropy and (c) follows from Equations [ETW08, (51)] and [ETW08, (52)]. Now we derive (C.55) below using the fact that W_2 has to be decoded at both receivers:

$$
\begin{aligned}
& n(R_1 + 2R_2 - \epsilon_n) \\
& \leq I\left(W_1; Y_1^n, S_1^n, \underline{g^n}\right) + I\left(W_2; Y_2^n, \underline{g^n}\right) + I\left(W_2; Y_1^n, S_2^n, \underline{g^n}\right) && \text{(C.66)}\\
& = I\left(W_1; Y_1^n, S_1^n \middle| \underline{g^n}\right) + I\left(W_2; Y_2^n \middle| \underline{g^n}\right) + I\left(W_2; Y_1^n, S_2^n \middle| \underline{g^n}\right) && \text{(C.67)}\\
& = I\left(W_1; S_1^n \middle| \underline{g^n}\right) + I\left(W_1; Y_1^n \middle| S_1^n, \underline{g^n}\right) + I\left(W_2; Y_2^n \middle| \underline{g^n}\right) \\
& \qquad + I\left(W_2; S_2^n \middle| \underline{g^n}\right) + \underbrace{I\left(W_2; Y_1^n \middle| S_2^n, \underline{g^n}\right)}_{=0} && \text{(C.68)}\\
& = h\left(S_1^n \middle| \underline{g^n}\right) - h\left(S_1^n \middle| W_1, \underline{g^n}\right) + h\left(Y_1^n \middle| S_1^n, \underline{g^n}\right) - h\left(Y_1^n \middle| W_1, S_1^n, \underline{g^n}\right) \\
& \qquad + h\left(Y_2^n \middle| \underline{g^n}\right) - h\left(Y_2^n \middle| W_2, \underline{g^n}\right) + h\left(S_2^n \middle| \underline{g^n}\right) - h\left(S_2^n \middle| W_2, \underline{g^n}\right) && \text{(C.69)}\\
& = h\left(S_1^n \middle| \underline{g^n}\right) - h\left(Z_2^n\right) + h\left(Y_1^n \middle| S_1^n, \underline{g^n}\right) - h\left(S_2^n \middle| \underline{g^n}\right) + h\left(Y_2^n \middle| \underline{g^n}\right) - h\left(S_1^n \middle| \underline{g^n}\right)
\end{aligned}
$$

$$+ h\left(S_2^n \middle| \underline{g}^n\right) - h\left(Z_1^n\right) \tag{C.70}$$

$$= h\left(Y_1^n \middle| S_1^n, \underline{g}^n\right) + h\left(Y_2^n \middle| \underline{g}^n\right) - h\left(Z_1^n\right) - h\left(Z_2^n\right) \tag{C.71}$$

$$\overset{(a)}{\leq} \sum \left[h\left(Y_{1i}|S_{1i}, \underline{g}^n\right) - h\left(Z_{1i}\right)\right] + \sum \left[h\left(Y_{2i}|\underline{g}^n\right) - h\left(Z_{2i}\right)\right] \tag{C.72}$$

$$= \mathbb{E}_{\underline{g}^n}\left[\sum \left(h\left(Y_{1i}|S_{1i}, \underline{g}^n\right) - h\left(Z_{1i}\right)\right)\right] + \mathbb{E}_{\underline{g}^n}\left[\sum \left(h\left(Y_{2i}|\underline{g}^n\right) - h\left(Z_{2i}\right)\right)\right] \tag{C.73}$$

$$\overset{(b)}{\leq} n\mathbb{E}\left[\log\left(1 + |g_{21}|^2 + \frac{|g_{11}|^2}{1 + |g_{12}|^2}\right)\right] + n\mathbb{E}\left[\log\left(1 + |g_{12}|^2 + |g_{22}|^2\right)\right], \tag{C.74}$$

where (a) is due to the fact that conditioning reduces entropy and (b) follows from Equations [ETW08, (50)] , [ETW08, (51)] and [ETW08, (52)]. Note that in the calculation of step (b) we allow the symbols X_{1i}, X_{2i} to depend on $\underline{g}^n$, but since $\underline{g}^n$ is available in conditioning the calculation proceeds similar to that in [ETW08].

C.6 Proof of Corollary 4.13

Let $\mathcal{R}'_{NFB}$ be the approximately optimal Han-Kobayashi rate region of feedback IC [ST11] with equivalent channel strengths $SNR_i := \mathbb{E}\left[|g_{ii}|^2\right]$ for $i = 1, 2$, and $INR_i := \mathbb{E}\left[|g_{ij}|^2\right]$ for $i \neq j$. Then for a constant c'' we have

$$\mathcal{R}'_{FB} \supseteq \mathcal{R}_{FB} \supseteq \mathcal{R}'_{FB} - c''. \tag{C.75}$$

This can be verified by proceeding through each inner bound equation. For example, consider the first inner bound Equation (4.23a) $R_1 \leq \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|\rho|^2 \operatorname{Re}\left(g_{11}g_{21}^*\right) + 1\right)\right] - 1$. The corresponding equation in $\mathcal{R}'_{NFB}$ is $R_1 \leq \log\left(1 + SNR_1 + INR_2 + 2|\rho|^2 \sqrt{SNR_1 \cdot INR_2} + 1\right) - 1$. Now

$$\mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|\rho|^2 \operatorname{Re}\left(g_{11}g_{21}^*\right) + 1\right)\right]$$

$$\overset{(a)}{\leq} \log\left(1 + SNR_1 + INR_2\right) \tag{C.76}$$

$$\leq \log\left(1 + SNR_1 + INR_2 + 2|\rho|^2 \sqrt{SNR_1 \cdot INR_2} + 1\right), \tag{C.77}$$

where (a) is due to Jensen's inequality and independence of g_{11}, g_{21}. Also

$$\mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|\rho|^2 \operatorname{Re}\left(g_{11}g_{21}^*\right) + 1\right)\right]$$

$$\overset{(a)}{=} \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2|g_{11}||g_{21}||\rho|\cos(\theta) + 1\right)\right] \tag{C.78}$$

$$\overset{(b)}{\geq} \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 1\right)\right] - 1 \tag{C.79}$$

$$\overset{(c)}{\geq} \log\left(SNR_1 + INR_2 + 1\right) - 1 - 2c_{JG} \tag{C.80}$$

$$\overset{(d)}{\geq} \log\left(SNR_1 + INR_2 + 2|\rho|^2\sqrt{SNR_1 \cdot INR_2} + 1\right) - 2 - 2c_{JG}, \tag{C.81}$$

where (a) is because phases of g_{11}, g_{12} are independently uniformly distributed in $[0, 2\pi]$ yielding $\mathrm{Re}\left(g_{11}g_{21}^*\right) = |g_{11}||g_{21}|\cos(\theta)$ with an independent $\theta \sim \mathrm{Unif}\,[0, 2\pi]$, (b) is using the fact that for $p > q$ $\frac{1}{2\pi}\int_0^{2\pi}\log(p + q\cos(\theta))\,d\theta = \log\left(\frac{p+\sqrt{p^2-q^2}}{2}\right) \geq \log(p) - 1$, (c) is using logarithmic Jensen's gap result twice and (d) is because $SNR_1 + INR_2 \geq 2|\rho|^2\sqrt{SNR_1 \cdot INR_2}$. It follows from Equations (C.77) and (C.81), that the first inner bound for fading case is within constant gap with the first inner bound of the static case.

Now consider the second inner bound Equation (4.23b)

$$R_1 \leq \mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right] - 2 \tag{C.82}$$

and the corresponding equation

$$R_1 \leq \log\left(1 + \left(1 - |\rho|^2\right)INR_1\right) + \log\left(1 + \lambda_{p1}SNR_1 + \lambda_{p2}INR_2\right) - 3c_{JG} - 2 \tag{C.83}$$

from $\mathcal{R}'_{FB}$. We have

$$\begin{aligned}&\mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right]\\ &\leq \log\left(1 + \left(1 - |\rho|^2\right)INR_1\right) + \log\left(1 + \lambda_{p1}SNR_1 + \lambda_{p2}INR_2\right)\end{aligned} \tag{C.84}$$

due to Jensen's inequality. And

$$\begin{aligned}&\mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \lambda_{p1}|g_{11}|^2 + \lambda_{p2}|g_{21}|^2\right)\right]\\ &\geq \log\left(1 + \left(1 - |\rho|^2\right)INR_1\right) + \log\left(1 + \lambda_{p1}SNR_1 + \lambda_{p2}INR_2\right) - 3c_{JG}\end{aligned} \tag{C.85}$$

using logarithmic Jensen's gap result thrice. It follows from Equations (C.84) and (C.85), that the second inner bound for fading case is within constant gap with the second inner bound of the static case. Similarly, by proceeding through each inner bound equation, it follows that

$$\mathcal{R}'_{FB} \supseteq \mathcal{R}_{FB} \supseteq \mathcal{R}'_{FB} - c''$$

for a constant c''.

C.7 Proof of achievability for feedback case

We evaluate the term in the first inner bound inequality $(5.14a)$. The other terms can be similarly evaluated.

$$I\left(U,U_2,X_1;Y_1,\underline{g_1}\right) \overset{(a)}{=} I\left(U,U_2,X_1;Y_1|\,\underline{g_1}\right) \tag{C.86}$$

$$= h\left(Y_1|\underline{g_1}\right) - h\left(Y_1|\underline{g_1},U,U_2,X_1\right), \tag{C.87}$$

$$\text{variance}\left(Y_1|\underline{g_1}\right) = \text{variance}\left(g_{11}X_1 + g_{21}X_2 + Z_1|\,g_{11},g_{21}\right) \tag{C.88}$$

$$= |g_{11}|^2 + |g_{21}|^2 + g_{11}^* g_{21}\mathbb{E}\left[X_1^* X_2\right] + g_{11}g_{21}^*\mathbb{E}\left[X_1 X_2^*\right] + 1 \tag{C.89}$$

$$= |g_{11}|^2 + |g_{21}|^2 + 2\,|\rho|^2\,\mathrm{Re}\left(g_{11}g_{21}^*\right) + 1, \tag{C.90}$$

$$h\left(Y_1|\underline{g_1},U,U_2,X_1\right) = h\left(g_{11}X_1 + g_{21}X_2 + Z_1|\,\underline{g_1},U,U_2,X_1\right) \tag{C.91}$$

$$= h\left(g_{21}X_{p2} + Z_1|\,\underline{g_1}\right) \tag{C.92}$$

$$= \mathbb{E}\left[\log\left(1 + \lambda_{p2}\,|g_{21}|^2\right)\right] + \log\left(2\pi e\right) \tag{C.93}$$

$$\overset{(b)}{\leq} \mathbb{E}\left[\log\left(1 + \frac{1}{INR_2}\,|g_{21}|^2\right)\right] + \log\left(2\pi e\right) \tag{C.94}$$

$$\overset{(c)}{\leq} \log\left(2\right) + \log\left(2\pi e\right) \tag{C.95}$$

$$= 1 + \log\left(2\pi e\right), \tag{C.96}$$

$$\therefore I\left(U,U_2,X_1;Y_1,\underline{g_1}\right) \geq \mathbb{E}\left[\log\left(|g_{11}|^2 + |g_{21}|^2 + 2\,|\rho|^2\,\mathrm{Re}\left(g_{11}g_{21}^*\right) + 1\right)\right] - 1, \tag{C.97}$$

where (a) uses independence, (b) is because $\lambda_{pi} \leq \frac{1}{INR_i}$, and (c) follows from Jensen's inequality.

C.8 Proof of outer bounds for feedback case

Following the methods in [ST11], we let $\mathbb{E}\left[X_1X_2^*\right]=\rho$. We have the notation $\underline{g_1}=[g_{11},g_{21}]$, $\underline{g_2}=[g_{22},g_{12}]$, $\underline{g}=[g_{11},g_{21},g_{22},g_{12}]$, $S_1=g_{12}X_1+Z_2$, and $S_2=g_{21}X_2+Z_1$. We let $\mathbb{E}\left[X_1X_2^*\right]=\rho=|\rho|\,e^{i\theta}$. All of our outer bounding steps are valid while allowing X_{1i} to be a function of $\left(W_1,Y_1^{i-1},\underline{g_1^n}\right)$, thus letting transmitters have full CSIT along with feedback. On choosing a uniform distribution of messages we get

$$
\begin{aligned}
n(R_1-\epsilon_n) &\overset{(a)}{\le} I\left(W_1;Y_1^n|\,\underline{g_1^n}\right) && \text{(C.98)}\\
&\overset{(b)}{\le} \sum\left(h\left(Y_{1i}|\underline{g_{1i}}\right)-h\left(Z_{1i}\right)\right) && \text{(C.99)}\\
&=\sum\left(\mathbb{E}_{\underline{\tilde{g}_{1i}}}\left[h\left(Y_{1i}|\underline{g_{1i}}=\underline{\tilde{g}_{1i}}\right)-h\left(Z_{1i}\right)\right]\right) && \text{(C.100)}\\
&\overset{(c)}{=}\mathbb{E}_{\underline{\tilde{g}_1}}\left[\sum\left(h\left(Y_{1i}|\underline{g_{1i}}=\underline{\tilde{g}_1}\right)-h\left(Z_{1i}\right)\right)\right] && \text{(C.101)}\\
\therefore R_1 &\le \mathbb{E}\left[\log\left(|g_{11}|^2+|g_{21}|^2+\left(\rho^*g_{11}^*g_{21}+\rho g_{11}g_{21}^*\right)+1\right)\right], && \text{(C.102)}
\end{aligned}
$$

where (a) follows from Fano's inequality, (b) follows from the fact that conditioning reduces entropy, and (c) follows from the fact that $\underline{\tilde{g}_{1i}}$ are i.i.d. Now we bound R_1 in a second way as done in [ST11]:

$$
\begin{aligned}
n(R_1-\epsilon_n) &\le I\left(W_1;Y_1^n,\underline{g_1^n}\right) && \text{(C.103)}\\
&\le I\left(W_1;Y_1^n,\underline{g_1^n},Y_2^n,\underline{g_2^n},W_2\right) && \text{(C.104)}\\
&=I\left(W_1;\underline{g}^n,W_2\right)+I\left(W_1;Y_1^n,Y_2^n|\underline{g}^n,W_2\right) && \text{(C.105)}\\
&=0+I\left(W_1;Y_1^n,Y_2^n|\underline{g}^n,W_2\right) && \text{(C.106)}\\
&=h\left(Y_1^n,Y_2^n|\underline{g}^n,W_2\right)-h\left(Y_1^n,Y_2^n|\underline{g}^n,W_1,W_2\right) && \text{(C.107)}\\
&=\sum\left[h\left(Y_{1i},Y_{2i}|\underline{g}^n,W_2,Y_1^{i-1},Y_2^{i-1}\right)\right]-\sum\left[h\left(Z_{1i}\right)+h\left(Z_{2i}\right)\right] && \text{(C.108)}\\
&=\sum\left[h\left(Y_{2i}|\underline{g}^n,W_2,Y_1^{i-1},Y_2^{i-1}\right)\right]+\sum\left[h\left(Y_{1i}|\underline{g}^n,W_2,Y_1^{i-1},Y_2^{i}\right)\right]\\
&\qquad-\sum\left[h\left(Z_{1i}\right)+h\left(Z_{2i}\right)\right] && \text{(C.109)}\\
&\overset{(a)}{=}\sum\left[h\left(Y_{2i}|\underline{g}^n,W_2,Y_1^{i-1},Y_2^{i-1},X_2^i\right)\right]+\sum\left[h\left(Y_{1i}|\underline{g}^n,W_2,Y_1^{i-1},Y_2^{i},S_{1i},X_2^i\right)\right]\\
&\qquad-\sum\left[h\left(Z_{1i}\right)+h\left(Z_{2i}\right)\right] && \text{(C.110)}
\end{aligned}
$$

$$\overset{(b)}{\leq} \sum \left[h\left(Y_{2i}|\underline{g_i}, X_{2i}\right) - h\left(Z_{2i}\right)\right] + \sum \left[h\left(Y_{1i}|\underline{g_i}, S_{1i}, X_{2i}\right) - h\left(Z_{1i}\right)\right] \tag{C.111}$$

$$\overset{(c)}{=} \mathbb{E}_{\underline{\tilde{g}}}\left[\sum \left(h\left(Y_{2i}|X_{2i}, \underline{g_i} = \underline{\tilde{g}}\right) - h\left(Z_{2i}\right)\right)\right]$$

$$+ \mathbb{E}_{\underline{\tilde{g}}}\left[\sum \left(h\left(Y_{1i}|S_{1i}, X_{2i}, \underline{g_i} = \underline{\tilde{g}}\right) - h\left(Z_{1i}\right)\right)\right], \tag{C.112}$$

$$\therefore R_1 \overset{(d)}{\leq} \mathbb{E}\left[\log\left(1 + \left(1 - |\rho|^2\right)|g_{12}|^2\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{\left(1 - |\rho|^2\right)|g_{11}|^2}{1 + \left(1 - |\rho|^2\right)|g_{12}|^2}\right)\right], \tag{C.113}$$

where (a) follows from the fact that X_2^i is a function of $\left(W_2, Y_2^{i-1}, \underline{g}^n\right)$ and S_{1i} is a function of $\left(Y_2^i, X_2^i, \underline{g}^n\right)$, (b) follows from the fact that conditioning reduces entropy, (c) follows from the fact that $\underline{\tilde{g}_i}$ are i.i.d., and (d) follows from [ST11, (43)]. The other outer bounds can be derived similarly following [ST11] and making suitable changes to account for fading as we illustrated in the previous two derivations.

C.9 Fading matrix

The calculations are given in Equations (C.114),(C.115).

$$\begin{aligned}
&\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)|\right)\right] \\
&= \mathbb{E}\left[\log\left(\left(|g_{11}(n)|^2 + |g_{21}(n)|^2\left(\frac{|g_{12}(n-1)|^2 + 1}{1 + INR}\right) + 1\right)|K_{\mathbf{Y_1}}(n-1)|\right.\right. \\
&\qquad \left.\left. - \frac{|g_{11}(n-1)|^2 |g_{21}(n)|^2 |g_{12}(n-1)|^2}{1 + INR}|K_{\mathbf{Y_1}}(n-2)|\right)\right] \qquad \text{(C.114)} \\
&\geq \mathbb{E}\left[\log\left((1 + INR + SNR)|K_{\mathbf{Y_1}}(n-1)|\right.\right. \\
&\qquad \left.\left. - \frac{INR \cdot INR\,|g_{11}(n-1)|^2}{1 + INR}|K_{\mathbf{Y_1}}(n-2)|\right)\right] - 3c_{JG}. \qquad \text{(C.115)}
\end{aligned}$$

The first step (C.114), is by expanding the determinant. We use the logarithmic Jensen's gap property thrice in the second step (C.115). This is justified because the coefficients of $\left\{|g_{11}(n)|^2, |g_{12}(n-1)|^2, |g_{21}(n)|^2\right\}$ from Equation (C.114) are non-negative (due to the fact that all the matrices involved are covariance matrices), and the coefficients themselves are independent of $\left\{|g_{11}(n)|^2, |g_{12}(n-1)|^2, |g_{21}(n)|^2\right\}$. (Note that $|K_{\mathbf{Y_1}}(n-1)|$ depend on

$|g_{12}(n-2)|^2$ but not on $|g_{12}(n-1)|^2$). This procedure can be carried out n times and it follows that:

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\log\left(|K_{\mathbf{Y_1}}(n)|\right)\right] \geq \lim_{n\to\infty}\frac{1}{n}\log\left(\left|\hat{K}_{\mathbf{Y_1}}(n)\right|\right) - 3c_{JG}, \tag{C.116}$$

where $\hat{K}_{\mathbf{Y_1}}(n)$ is obtained from $K_{\mathbf{Y_1}}(n)$ by replacing $g_{12}(i)$'s, $g_{21}(i)$'s with $\sqrt{INR}$ and $g_{11}(i)$'s with $\sqrt{SNR}$.

C.10 Matrix determinant: asymptotic behavior

The following recursion easily follows:

$$|A_n| = |a|\,|A_{n-1}| - |b|^2\,|A_{n-2}| \tag{C.117}$$

with $|A_1| = |a|, |A_2| = |a|^2 - |b|^2$. Also $|A_0|$ can be consistently defined to be 1. The characteristic equation for this recursive relation is given by: $\lambda^2 - |a|\,\lambda + |b|^2 = 0$ and the characteristic roots are given by:

$$\lambda_1 = \frac{|a| + \sqrt{|a|^2 - 4\,|b|^2}}{2}, \lambda_2 = \frac{|a| - \sqrt{|a|^2 - 4\,|b|^2}}{2}. \tag{C.118}$$

Now the solution of the recursive system is given by $|A_n| = c_1\lambda_1^n + c_2\lambda_2^n$ with the boundary conditions $1 = c_1 + c_2, \quad |a| = c_1\lambda_1 + c_2\lambda_2$. It can be easily seen that $c_1 > 0$, $\lambda_1 > \lambda_2 > 0$ since $|a|^2 > 4\,|b|^2$ by assumption of Lemma 4.4. Now

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log\left(|A_n|\right) &= \lim_{n\to\infty}\frac{1}{n}\log\left(c_1\lambda_1^n + c_2\lambda_2^n\right) && \text{(C.119)}\\
&\stackrel{(a)}{=} \log\left(\lambda_1\right) && \text{(C.120)}\\
&= \log\left(|a| + \sqrt{|a|^2 - 4\,|b|^2}\right) - 1. && \text{(C.121)}
\end{aligned}$$

The step (a) follows because $\lambda_1 > \lambda_2 > 0$ and $c_1 > 0$.

C.11 Approximate capacity using n phase schemes

We have the following outer bounds from Theorem 4.11.

$$R_1, R_2 \leq \mathbb{E}\left[\log\left(|g_d|^2 + |g_c|^2 + 1\right)\right] \tag{C.122}$$

$$R_1 + R_2 \leq \mathbb{E}\left[\log\left(1 + \frac{|g_d|^2}{1 + |g_c|^2}\right)\right] + \mathbb{E}\left[\log\left(|g_d|^2 + |g_c|^2 + 2|g_d||g_c| + 1\right)\right]. \tag{C.123}$$

The above outer bound region is a polytope with the following two non-trivial corner points:

$$\left\{\begin{array}{ccc} R_1 & = & \mathbb{E}\left[\log\left(|g_d|^2 + |g_c|^2 + 1\right)\right] \\ R_2 & = & \mathbb{E}\left[\log\left(1 + \frac{|g_d|^2}{1+|g_c|^2}\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{2|g_d||g_c|}{1+|g_d|^2+|g_c|^2}\right)\right] \end{array}\right\}$$

$$\left\{\begin{array}{ccc} R_1 & = & \mathbb{E}\left[\log\left(1 + \frac{|g_d|^2}{1+|g_c|^2}\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{2|g_d||g_c|}{1+|g_d|^2+|g_c|^2}\right)\right] \\ R_2 & = & \mathbb{E}\left[\log\left(|g_d|^2 + |g_c|^2 + 1\right)\right] \end{array}\right\}.$$

We can achieve these rate points within $2 + 3c_{JG}$ bits per channel use for each user using the n-phase schemes since

$$(R_1, R_2) = \left(\log\left(1 + SNR + INR\right) - 2 - 3c_{JG}, \mathbb{E}\left[\log^+\left[\frac{|g_d|^2}{1 + INR}\right]\right]\right) \tag{C.124}$$

$$(R_1, R_2) = \left(\mathbb{E}\left[\log^+\left[\frac{|g_d|^2}{1 + INR}\right]\right], \log\left(1 + SNR + INR\right) - 2 - 3c_{JG}\right). \tag{C.125}$$

are achievable and since using Jensen's inequality

$$\mathbb{E}\left[\log\left(|g_d|^2 + |g_c|^2 + 1\right)\right] \leq \log\left(1 + SNR + INR\right). \tag{C.126}$$

The only important point left to verify is in the following claim.

Claim C.1. $\mathbb{E}\left[\log\left(1 + \frac{|g_d|^2}{1+|g_c|^2}\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{2|g_d||g_c|}{1+|g_d|^2+|g_c|^2}\right)\right] - \mathbb{E}\left[\log^+\left[\frac{|g_d|^2}{1+INR}\right]\right] \leq 2 + c_{JG}$

Proof. We have $\frac{2|g_d||g_c|}{|g_d|^2+|g_c|^2} \leq 1$ due to AM-GM inequality. Hence,

$$\mathbb{E}\left[\log\left(1 + \frac{2|g_d||g_c|}{1 + |g_d|^2 + |g_c|^2}\right)\right] \leq 1. \tag{C.127}$$

Also

$$\mathbb{E}\left[\log\left(1+\frac{|g_d|^2}{1+|g_c|^2}\right)\right] \leq \mathbb{E}\left[\log\left(1+\frac{|g_d|^2}{1+INR}\right)\right] + c_{JG} \tag{C.128}$$

using logarithmic Jensen's gap property. Hence, it only remains to show $\log\left(1+\frac{|g_d|^2}{1+INR}\right) - \log^+\left[\frac{|g_d|^2}{1+INR}\right] \leq 1$ to complete the proof.

If $\log^+\left[\frac{|g_d|^2}{1+INR}\right] = 0$ then $\frac{|g_d|^2}{1+INR} \leq 1$ and Hence, $\log\left(1+\frac{|g_d|^2}{1+INR}\right) \leq \log(2) = 1$.

If $\log^+\left[\frac{|g_d|^2}{1+INR}\right] > 0$ then $\frac{|g_d|^2}{1+INR} > 1$ and Hence, again

$$\log\left(1+\frac{|g_d|^2}{1+INR}\right) - \log^+\left[\frac{|g_d|^2}{1+INR}\right] = \log\left(1+\frac{1+INR}{|g_d|^2}\right) < 1. \tag{C.129}$$

□

C.12 Analysis for the 2-tap fading ISI channel

We have for the outer bound

$$n\left(R-\epsilon_n\right) \leq I\left(Y^n, g_d^n, g_c^n; W\right) \tag{C.130}$$

$$= I\left(Y^n; W \middle| g_d^n, g_c^n\right) \tag{C.131}$$

$$= h\left(Y^n \middle| g_d^n, g_c^n\right) - h\left(Z^n\right) \tag{C.132}$$

$$\leq \sum h\left(Y_i \middle| g_{d,i}, g_{c,i}\right) - h\left(Z^n\right) \tag{C.133}$$

$$\overset{(a)}{\leq} \sum \mathbb{E}\left[\log\left(1+P_i|g_d|^2+P_{i-1}|g_c|^2+2|g_d||g_c|\sqrt{P_iP_{i-1}}\right)\right] \tag{C.134}$$

$$\leq \sum \left(\mathbb{E}\left[\log\left(1+P_i|g_d|^2+P_{i-1}|g_c|^2\right)\right]+1\right), \tag{C.135}$$

where (a) is using P_i as the power for i^{th} symbol and using Cauchy Schwarz inequality to bound $|\mathbb{E}[X_iX_{i-1}]| \leq \sqrt{P_iP_{i-1}}$. Now using Jensen's inequality it follows that

$$R-\epsilon_n \leq \mathbb{E}\left[\log\left(1+|g_d|^2+|g_c|^2\right)\right]+1 \tag{C.136}$$

$$\leq \log\left(1+SNR+INR\right)+1. \tag{C.137}$$

For the inner bound similar to the scheme in subsection (4.6), using Gaussian codebooks and n phases we obtain that

$$R = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[\log \left(\frac{|K_{\mathbf{Y}}(n)|}{\left|K_{\mathbf{Y}|\mathbf{X}}(n)\right|} \right) \right] \tag{C.138}$$

is achievable, where $\boldsymbol{X}(n)$ is n-length Gaussian vector with i.i.d $\mathcal{CN}(0,1)$ elements and $\boldsymbol{Y}(n)$ is generated from $\boldsymbol{X}(n)$ by the ISI channel (from Equation (4.68)). Here $\left|K_{\mathbf{Y}|\mathbf{X}}(n)\right| = |K_{\mathbf{Z}}(n)| = 1$ because Z is AWGN. Hence,

$$R = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[\log \left(|K_{\mathbf{Y}}(n)| \right) \right] \tag{C.139}$$

is achievable. Hence, it follows that

$$R \geq \log \left(1 + SNR + INR \right) - 1 - 3c_{JG} \tag{C.140}$$

is achievable due to Lemma 4.3 and Lemma 4.4.

APPENDIX D

Proofs for Chapter 5

D.1 Proof of Claim 5.1

$$
\begin{aligned}
I\left(X_1;Y_1|U_2\right) &= h\left(Y_1|U_2\right) - h\left(Y_1|U_2,X_1\right) \\
&= h\left(g_{11}X_1 + g_{21}X_2 + Z_1\middle| U_2\right) - h\left(g_{11}X_1 + g_{21}X_2 + Z_1\middle| U_2, X_1\right) \qquad \text{(D.1)}
\end{aligned}
$$

$$
\begin{aligned}
h\left(g_{11}X_1 + g_{21}X_2 + Z_1|U_2\right) &= \sum_{i=1}^{T} h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| \left\{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\right\}_{j=1}^{i-1}, U_2\right) \\
&\overset{(i)}{\dot{\geq}} h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11}\middle| X_{11}, X_{21}, U_2\right) \\
&\quad + \sum_{i=2}^{T} h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| U_{2i}, g_{21}, g_{11}\right) \\
&\overset{(ii)}{\dot{\geq}} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + (T-1)\log\left(1+\mathsf{SNR}\right) \qquad \text{(D.2)}
\end{aligned}
$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\right) - \left(U_{2i}, g_{21}, g_{11}\right) - \left(\left\{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\right\}_{j=1}^{i-1}, U_2\right)$ and (ii) is using Gaussianity for terms $h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11}\middle| X_{11}, X_{21}, U_2\right)$ and $h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| U_{2i}, g_{21}, g_{11}\right)$. Now we will show that

$$
h\left(g_{11}X_1 + g_{21}X_2 + Z_1\middle| U_2, X_1\right) \dot{\leq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + \log\left(1 + \mathsf{INR}\right) \qquad \text{(D.3)}
$$

and this will complete our proof for $I\left(X_1;Y_1|U_2\right) \dot{\geq} (T-1)\log\left(1+\mathsf{SNR}\right) - \log\left(1+\mathsf{INR}\right)$. For (D.3), we have

$$
h\left(g_{11}X_1 + g_{21}X_2 + Z_1|U_2, X_1\right)
$$

$$
\begin{aligned}
&\leq h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle|\, U_{21},X_{11}\right)\\
&\quad + h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_2,X_1\right) \qquad \text{(D.4)}\\
&\quad + \sum_{i=3}^{T} h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},g_{11}X_{12}+g_{21}X_{22}+Z_{12},U_2,X_1\right)\\
&\dot{\leq} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_2,X_1\right)\\
&\quad + \sum_{i=3}^{T} h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},g_{11}X_{12}+g_{21}X_{22}+Z_{12},U_2,X_1\right),
\end{aligned}
$$

$$\text{(D.5)}$$

Considering the second term in the above expression,

$$
\begin{aligned}
&h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_2,X_1\right)\\
&= h\left(g_{11}X_{11}X_{12}+g_{21}X_{11}X_{22}+X_{11}Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_2,X_1\right)-\mathbb{E}\left[\log\left(|X_{11}|\right)\right]\\
&\overset{(i)}{\leq} h\left(g_{11}X_{11}X_{12}+g_{21}X_{11}X_{22}+X_{11}Z_{12}-X_{12}\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\right)\right)-\mathbb{E}\left[\log\left(|X_{11}|\right)\right]\\
&= h\left(g_{21}X_{11}X_{22}+X_{11}Z_{12}-X_{12}\left(g_{21}X_{21}+Z_{11}\right)\right)-\mathbb{E}\left[\log\left(|X_{11}|\right)\right]\\
&= h\left(g_{21}\left(X_{11}X_{22}-X_{21}X_{12}\right)+X_{11}Z_{12}-X_{12}Z_{11}\right)-\mathbb{E}\left[\log\left(|X_{11}|\right)\right] \qquad \text{(D.6)}\\
&\dot{\leq} \log\left(\mathbb{E}\left[\left|g_{21}\left(X_{11}X_{22}-X_{21}X_{12}\right)+X_{11}Z_{12}-X_{12}Z_{11}\right|^2\right]\right)-\frac{1}{2}\mathbb{E}\left[\log\left(|X_{11}|^2\right)\right]\\
&\overset{(ii)}{\doteq} \log\left(1+\mathsf{INR}\right), \qquad \text{(D.7)}
\end{aligned}
$$

where (i) is by subtracting $X_{12}\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\right)$ which is available from conditioning and then using the fact that conditioning reduces entropy, (ii) is by using property of Gaussians for i.i.d. $g_{21},X_{11},X_{22},X_{21},X_{12},Z_{12},Z_{11}$ and Fact 5.1 for $\mathbb{E}\left[\log\left(|X_{11}|^2\right)\right]$ since $|X_{11}|^2$ is exponentially distributed with mean 1. Now for $i \geq 3$ we will show that

$$h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},g_{11}X_{12}+g_{21}X_{22}+Z_{12},U_2,X_1\right) \dot{\leq} 0. \qquad \text{(D.8)}$$

Using (D.8) and (D.7) in (D.5) yields us (D.3) and will complete the proof. For (D.8), we have

$$h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},g_{11}X_{12}+g_{21}X_{22}+Z_{12},U_2,X_1\right)$$

$$\leq h\left(g_{21}\left(X_{11}X_{2i}-X_{21}X_{1i}\right)+X_{11}Z_{1i}-X_{1i}Z_{11}\middle|\,|g_{21}\left(X_{11}X_{22}-X_{21}X_{12}\right)+X_{11}Z_{12}-X_{12}Z_{11},U_2,X_1\right)$$
$$-\mathbb{E}\left[\log\left(|X_{11}|\right)\right]. \tag{D.9}$$

Now we have

$$\begin{aligned}&g_{21}\left(X_{11}X_{2i}-X_{21}X_{1i}\right)+X_{11}Z_{1i}-X_{1i}Z_{11}\\&=g_{21}\left(X_{11}U_{2i}-U_{21}X_{1i}\right)+\left(g_{21}\left(X_{11}X_{2pi}-X_{2p1}X_{1i}\right)+X_{11}Z_{1i}-X_{1i}Z_{11}\right)\end{aligned}$$

in the entropy expression. And in the conditioning the term

$$g_{21}\left(X_{11}U_{22}-U_{21}X_{12}\right)+\left(g_{21}\left(X_{11}X_{2p2}-X_{2p1}X_{12}\right)+X_{11}Z_{12}-X_{12}Z_{11}\right)$$

and U_2, X_1 are available. Hence by elimination we can get

$$\begin{aligned}\xi=&\left(X_{11}U_{22}-U_{21}X_{12}\right)\left(g_{21}\left(X_{11}X_{2pi}-X_{2p1}X_{1i}\right)+X_{11}Z_{1i}-X_{1i}Z_{11}\right)\\&-\left(X_{11}U_{2i}-U_{21}X_{1i}\right)\left(g_{21}\left(X_{11}X_{2p2}-X_{2p1}X_{12}\right)+X_{11}Z_{12}-X_{12}Z_{11}\right)\end{aligned} \tag{D.10}$$

in the entropy expression. Let ξ be expanded into a sum of product form

$$\xi=\sum_{i=1}^{L}\xi_i \tag{D.11}$$

$$=X_{11}U_{22}g_{21}X_{11}X_{2pi}+\left(-X_{11}U_{22}g_{21}X_{2p1}X_{1i}\right)+\cdots \tag{D.12}$$

where ξ_i is in a simple product form. Now due to generalized mean inequality, we have

$$\left|\sum_{i=1}^{L}\xi_i\right|^2\leq L\left(\sum_{i=1}^{L}|\xi_i|^2\right). \tag{D.13}$$

Hence

$$\mathbb{E}\left[|\xi|^2\right]=\mathbb{E}\left[\left|\sum_{i=1}^{L}\xi_i\right|^2\right] \tag{D.14}$$

$$\leq L\left(\sum_{i=1}^{L}\mathbb{E}\left[|\xi_i|^2\right]\right). \tag{D.15}$$

Now for example consider the term $\mathbb{E}\left[|X_{11}U_{22}g_{21}X_{11}X_{2pi}|^2\right]$ in the last equation

$$\mathbb{E}\left[|X_{11}U_{22}g_{21}X_{11}X_{2pi}|^2\right]=\mathbb{E}\left[|X_{11}|^4\right]\mathbb{E}\left[|U_{22}|^2\right]\mathbb{E}\left[|g_{21}|^2\right]\mathbb{E}\left[|X_{2pi}|^2\right] \tag{D.16}$$

$$= 2 \times \left(1 - \frac{1}{\mathsf{INR}}\right) \times \mathsf{INR} \times \frac{1}{\mathsf{INR}} \tag{D.17}$$

$$\leq 2 \tag{D.18}$$

Each of $\mathbb{E}\left[|\xi_i|^2\right]$ will be bounded by a constant since g_{21} always appears coupled with X_{2pi}. Hence the power scaling $\mathbb{E}\left[|g_{21}|^2\right] = \mathsf{INR}$ gets canceled with the scaling $\mathbb{E}\left[|X_{2pi}|^2\right] = 1/\mathsf{INR}$. Hence, by analyzing each of $\mathbb{E}\left[|\xi_i|^2\right]$ together with maximum entropy results it can be shown that $\mathbb{E}\left[|\xi_i|^2\right] \dot{\leq} 0$ and hence, $h(\xi) \dot{\leq} 0$. Thus (D.8) is proved and it completes our proof for the main result.

D.2 Proof of Claim 5.3

We have

$$I(X_1; Y_1 | U_1, U_2) = h(Y_1 | U_1, U_2) - h(Y_1 | X_1, U_1, U_2), \tag{D.19}$$

$$\begin{aligned}
h(Y_1|U_1,U_2) &= h\left(g_{11}X_1 + g_{21}X_2 + Z_1 \middle| U_1, U_2\right) && \text{(D.20)}\\
&= \sum_i h\left(g_{11}X_{1i} + g_{21}X_{2i} + \middle| \{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\}_{j=1}^{i-1}, U_1, U_2\right) && \text{(D.21)}\\
&\overset{(i)}{\dot{\geq}} h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11} \middle| X_{11}, X_{21}, U_1, U_2\right) \\
&\quad + h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12} \middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_1, U_2\right) && \text{(D.22)}\\
&\quad + \sum_{i=3}^{T} h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i} \middle| U_{1i}, U_{2i}, g_{21}, g_{11}\right) && \text{(D.23)}\\
&\overset{(ii)}{\dot{\geq}} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12} \middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_1, U_2\right) \\
&\quad + (T-2)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right), && \text{(D.24)}
\end{aligned}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}) - (U_{1i}, U_{2i}, g_{21}, g_{11}) - \left(\{g_{11}X_{1j} + g_{21}X_{2j} + Z_{1j}\}_{j=1}^{i-1}, U_1, U_2\right)$ and (ii) is using Gaussianity for terms $h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11} \middle| X_{11}, X_{21}\right)$ and

$h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle| U_{1i},U_{2i},g_{21},g_{11}\right)$. Now,

$$
\begin{aligned}
& h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle| g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2\right) && \text{(D.25)}\\
& \geq h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle| g_{11}X_{11}+g_{21}X_{21}+Z_{11},X_1,X_2,U_1,U_2\right) && \text{(D.26)}\\
& = h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12},g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle| X_1,X_2,U_1,U_2\right) && \text{(D.27)}\\
& \qquad - h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle| X_1,X_2,U_1,U_2\right) && \text{(D.28)}\\
& \overset{(i)}{\doteq} \mathbb{E}\left[\log\left(\begin{vmatrix} \mathsf{SNR}\left|X_{12}\right|^2+\mathsf{INR}\left|X_{22}\right|^2+1 & \mathsf{SNR}X_{12}X_{11}^{\dagger}+\mathsf{INR}X_{22}X_{21}^{\dagger}\\ \left(\mathsf{SNR}X_{12}X_{11}^{\dagger}+\mathsf{INR}X_{22}X_{21}^{\dagger}\right)^{\dagger} & \mathsf{SNR}\left|X_{11}\right|^2+\mathsf{INR}\left|X_{21}\right|^2+1 \end{vmatrix}\right)\right]\\
& \quad -\log\left(1+\mathsf{INR}+\mathsf{SNR}\right) && \text{(D.29)}\\
& \geq \mathbb{E}\left[\log\left(\mathsf{SNR}\cdot\mathsf{INR}\left(\left|X_{11}\right|^2\left|X_{22}\right|^2+\left|X_{12}\right|^2\left|X_{21}\right|^2-2\mathrm{Re}\left(X_{12}X_{11}^{\dagger}X_{22}^{\dagger}X_{21}\right)\right)\right)\right]\\
& \quad -\log\left(1+\mathsf{INR}+\mathsf{SNR}\right) && \text{(D.30)}\\
& = \mathbb{E}\left[\log\left(\mathsf{SNR}\cdot\mathsf{INR}\left|X_{11}X_{22}-X_{12}X_{21}\right|^2\right)\right]-\log\left(1+\mathsf{INR}+\mathsf{SNR}\right) && \text{(D.31)}\\
& \doteq \log\left(\frac{\mathsf{SNR}\cdot\mathsf{INR}}{1+\mathsf{INR}+\mathsf{SNR}}\right)+\mathbb{E}\left[\log\left(\mathsf{SNR}\cdot\mathsf{INR}\left|X_{11}X_{22}-X_{12}X_{21}\right|^2\right)\right] && \text{(D.32)}\\
& \overset{(ii)}{\doteq} \log\left(\frac{\mathsf{SNR}\cdot\mathsf{INR}}{1+\mathsf{INR}+\mathsf{SNR}}\right) && \text{(D.33)}\\
& \overset{(iii)}{\doteq} \log\left(\mathsf{INR}\right), && \text{(D.34)}
\end{aligned}
$$

where (i) is using property of Gaussians, (ii) is using Fact 5.1 and Tower property of Expectation for $\mathbb{E}\left[\log\left(\left|X_{11}X_{22}-X_{12}X_{21}\right|^2\right)\right]$, (iii) is because $\mathsf{INR}\leq\mathsf{SNR}$. Also

$$
\begin{aligned}
& h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle| g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2\right)\\
& \overset{(i)}{\geq} h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle| g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,g_{11},g_{21}\right) && \text{(D.35)}\\
& \overset{(ii)}{\doteq} h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle| U_1,U_2,g_{11},g_{21}\right) && \text{(D.36)}\\
& \overset{(iii)}{\doteq} h\left(g_{11}X_{p12}+g_{21}X_{p22}+Z_{12}\middle| g_{11},g_{21}\right) && \text{(D.37)}\\
& = \mathbb{E}\left[\log\left(2\pi e\left(1+\frac{1}{\mathsf{INR}}\left|g_{11}\right|^2+\frac{1}{\mathsf{INR}}\left|g_{21}\right|^2\right)\right)\right] && \text{(D.38)}\\
& \overset{(iv)}{\dot{\geq}} \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right), && \text{(D.39)}
\end{aligned}
$$

where (i) is using the fact that conditioning reduces entropy, (ii) is due to the Markov chain $(g_{11}X_{12} + g_{21}X_{22} + Z_{12}) - (U_{12}, U_{22}, g_{21}, g_{11}) - (g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_1, U_2)$, (iii) is because the private message parts X_{p12}, X_{p22} are independent of the common message parts U_1, U_2, (iv) is using Fact 5.1. Now combining (D.34), (D.39), we get

$$h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12} \middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_1, U_2\right) \dot{\geq} \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right). \quad \text{(D.40)}$$

Hence substituting the above in D.24, we get

$$\begin{aligned} h\left(Y_1|U_1, U_2\right) \dot{\geq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) \\ + (T-2)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right). \end{aligned} \quad \text{(D.41)}$$

Also from (D.3) for $h\left(Y_1 \middle| X_1, U_2\right)$ in Appendix D.1 on page 225, we have

$$h\left(Y_1|X_1, U_1, U_2\right) \leq h\left(Y_1|X_1, U_2\right) \quad \text{(D.42)}$$

$$\dot{\leq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + \log\left(1 + \mathsf{INR}\right). \quad \text{(D.43)}$$

Hence using the above two equations we get,

$$I\left(X_1; Y_1|U_1, U_2\right) \dot{\geq} (T-2)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right) + \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR}\right) - \log\left(\mathsf{INR}\right). \quad \text{(D.44)}$$

D.3 Proof of Claim 5.5

$$\begin{aligned} I\left(X_1; Y_1|U_2\right) &= h\left(Y_1|U_2\right) - h\left(Y_1|U_2, X_1\right), \\ &= h\left(g_{11}X_1 + g_{21}X_2 + Z_1 \middle| U_2\right) - h\left(g_{11}X_1 + g_{21}X_2 + Z_1 \middle| U_2, X_1\right). \end{aligned}$$

We have

$$h\left(g_{11}X_1 + g_{21}X_2 + Z_1|U_2\right) \dot{\geq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + (T-1)\log\left(1 + \mathsf{SNR}\right)$$

following (D.2) in Appendix D.1 on page 225. Now we will show that

$$h\left(g_{11}X_1 + g_{21}X_2 + Z_1 \middle| U_2, X_1\right) \dot{\leq} \log\left(1 + \mathsf{SNR} + \mathsf{INR}\right) + \log\left(1 + \mathsf{SNR}\right) \quad \text{(D.45)}$$

and this will complete the proof for

$$I\left(X_1;Y_1|U_2\right) \dot{\geq} (T-1)\log\left(1+\mathsf{SNR}\right) - \log\left(1+\mathsf{SNR}\right) \doteq (T-2)\log\left(1+\mathsf{SNR}\right).$$

We have

$$\begin{aligned}
&h\left(g_{11}X_1 + g_{21}X_2 + Z_1|U_2, X_1\right)\\
&\overset{(i)}{\leq} h\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11}\middle| U_{21}, X_{11}\right) + h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_2, X_1\right)\\
&\qquad + \sum h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, g_{11}X_{12} + g_{21}X_{22} + Z_{12}, U_2, X_1\right)\\
&\doteq \log\left(1+\mathsf{SNR}+\mathsf{INR}\right) + h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_2, X_1\right)\\
&\qquad + \sum_{i=3}^{T} h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, g_{11}X_{12} + g_{21}X_{22} + Z_{12}, U_2, X_1\right),
\end{aligned} \tag{D.46}$$

where (i) was using the fact that conditioning reduces entropy.

$$\begin{aligned}
&h\left(g_{11}X_{12} + g_{21}X_{22} + Z_{12}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_2, X_1\right)\\
&= h\left(g_{11}U_{21}X_{12} + g_{21}U_{21}X_{22} + U_{21}Z_{12}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, U_2, X_1\right) - \mathbb{E}\left[\log\left(|U_{21}|\right)\right]\\
&\overset{(i)}{\leq} h\left(g_{11}U_{21}X_{12} + g_{21}U_{21}X_{22} + U_{21}Z_{12} - U_{22}\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11}\right)\right) - \mathbb{E}\left[\log\left(|U_{21}|\right)\right]\\
&= h\left(g_{11}\left(U_{21}X_{12} - X_{11}U_{22}\right) + g_{21}\left(U_{21}X_{p22} - U_{22}X_{p21}\right) + U_{21}Z_{12} - U_{22}Z_{11}\right)\\
&\quad - \mathbb{E}\left[\log\left(|U_{21}|\right)\right]
\end{aligned} \tag{D.47}$$

$$\begin{aligned}
&\dot{\leq} \log\left(\mathbb{E}\left[\left|g_{11}\left(U_{21}X_{12} - X_{11}U_{22}\right) + g_{21}\left(U_{21}X_{p22} - U_{22}X_{p21}\right) + U_{21}Z_{12} - U_{22}Z_{11}\right|^2\right]\right)\\
&\quad - \frac{1}{2}\mathbb{E}\left[\log\left(|U_{21}|^2\right)\right]\\
&\overset{(ii)}{\doteq} \log\left(1+\mathsf{SNR}\right),
\end{aligned} \tag{D.48}$$

where (i) is by subtracting $X_{12}\left(g_{11}X_{11} + g_{21}X_{21} + Z_{11}\right)$ which is available from conditioning and then using the fact that conditioning reduces entropy, (ii) is by using properties of i.i.d. Gaussians to evaluate the second moments and Fact 5.1 for $\mathbb{E}\left[\log\left(|U_{21}|^2\right)\right]$ since $|U_{21}|^2$ is exponentially distributed with mean $1 - 1/\mathsf{INR}$. Now for $i \geq 3$ we claim that

$$h\left(g_{11}X_{1i} + g_{21}X_{2i} + Z_{1i}\middle| g_{11}X_{11} + g_{21}X_{21} + Z_{11}, g_{11}X_{12} + g_{21}X_{22} + Z_{12}, U_2, X_1\right) \dot{\leq} 0. \tag{D.49}$$

This follows with the same steps as for the low interference case as in (D.8) on page 226. Using (D.49) and (D.48) in (D.46) yields (D.45) and completes the proof.

D.4 Proof of Claim 5.10

We have

$$\begin{aligned}
h\left(Y_1\middle|\,U_1,U_2,U\right) =& h\left(g_{11}X_1+g_{21}X_2+Z_1\middle|\,U_1,U_2,U\right)\\
=&\sum_i h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\,\left\{g_{11}X_{1j}+g_{21}X_{2j}+Z_{1j}\right\}_{j=1}^{i-1},U_1,U_2,U\right)\\
\overset{(i)}{\dot{\geq}}& h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle|\,X_{21},X_1,U_1,U_2,U\right)\\
&+h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\,g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&+\sum_{i=3}^{T} h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\,U_1,U_2,U,g_{21},g_{11}\right)\\
\overset{(ii)}{\dot{\geq}}& \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)\\
&+h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\,g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&+(T-2)\,\mathbb{E}\left[\log\left(1+\lambda_{p2}\left|g_{22}\right|^2+\lambda_{p1}\left|g_{12}\right|^2\right)\right],\\
\overset{(iii)}{\doteq}& \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)\\
&+h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\,g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right),
\end{aligned} \tag{D.50}$$

where (i) is due to the fact that conditioning reduces entropy and Markovity $(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i})-(U_1,U_2,U,g_{21},g_{11})-\left(\left\{g_{11}X_{1j}+g_{21}X_{2j}+Z_{1j}\right\}_{j=1}^{i-1},U_1,U_2,U\right)$ and (ii) is using Gaussianity for terms $h\left(g_{11}X_{11}+g_{21}X_{21}+Z_{11}\middle|\,X_{21},X_1,U_1,U_2,U\right)$ and $h\left(g_{11}X_{1i}+g_{21}X_{2i}+Z_{1i}\middle|\,U_1,U_2,U,g_{21},g_{11}\right)$. The step (iii) is using Fact 5.1. Now

$$\begin{aligned}
&h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\,g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&\geq h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\,g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U,X_1,X_2\right)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(i)}{\doteq} \log\left(\frac{\mathsf{SNR}\cdot\mathsf{INR}}{1+\mathsf{SNR}+\mathsf{INR}}\right)\\
&\doteq \log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right), \qquad \text{(D.51)}
\end{aligned}
$$

where (i) is using similar calculations as (D.33) on page 229. Also

$$
\begin{aligned}
&h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&\dot{\geq} h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U,g_{11},g_{21}\right)\\
&=\mathbb{E}\left[\log\left(1+\lambda_{p2}\left|g_{22}\right|^2+\lambda_{p1}\left|g_{12}\right|^2\right)\right]\\
&\doteq \log\left(1+\lambda_{p2}\mathsf{SNR}+\lambda_{p1}\mathsf{INR}\right)\\
&\dot{\geq} \log\left(1+\lambda_{p2}\mathsf{SNR}\right)\\
&\doteq \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right). \qquad \text{(D.52)}
\end{aligned}
$$

Using (D.51), (D.52) we get

$$
\begin{aligned}
&h\left(g_{11}X_{12}+g_{21}X_{22}+Z_{12}\middle|\, g_{11}X_{11}+g_{21}X_{21}+Z_{11},U_1,U_2,U\right)\\
&\dot{\geq} \log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right).
\end{aligned}
$$

Using the above equation in (D.50), we get

$$
\begin{aligned}
h\left(Y_1\middle|\, U_1,U_2,U\right) \dot{\geq} &\log\left(1+\mathsf{SNR}+\mathsf{INR}\right)\\
&+\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)\\
&+(T-2)\log\left(1+\frac{\mathsf{SNR}}{\mathsf{INR}}\right),
\end{aligned}
$$

Also

$$
\begin{aligned}
h\left(Y_1\middle|\, U,U_1,U_2,X_1\right) &\leq h\left(Y_1\middle|\, U,U_2,X_1\right)\\
&\dot{\leq} \log\left(1+\mathsf{SNR}+\mathsf{INR}\right)+\log\left(1+\min\left(\mathsf{SNR},\mathsf{INR}\right)\right)
\end{aligned}
$$

from (5.18) on page 143. Using the above two equations, we get

$$
I\left(X_1;Y_1|U_1,U_2,U\right)
$$

$$\dot{\geq} \log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}} + \min\left(\mathsf{SNR}, \mathsf{INR}\right)\right) + (T-2)\log\left(1 + \frac{\mathsf{SNR}}{\mathsf{INR}}\right)$$
$$- \log\left(1 + \min\left(\mathsf{SNR}, \mathsf{INR}\right)\right).$$

APPENDIX E

Details of Schemes for Backscatter Systems

E.1 Backscatter system with ISI taps

We consider the backscatter communication system in which Emitter and Reader can collaborate (Figure 6.3). We focus on the 2-tap ISI channel, but all the concepts extend to arbitrary number of ISI taps. Let $\{g_{\mathrm{t}1}, g_{\mathrm{t}2}\}$ be the taps from Emitter to Tag and $\{g_{\mathrm{r}1}, g_{\mathrm{r}2}\}$ be the taps from Tag to Reader. Emitter is assumed to have an average power constraint of P. Tag is assumed to use ON-OFF keying for data transmission. For simplicity of design, we assume that Emitter sends a periodic sequence $\{s_1, \ldots, s_N\}$ and Tag sends cyclic-suffixed symbols of the form $\{x_{j1}, \ldots, x_{jN}, x_{j1}\}$ with j being the block index. For every block, the first symbol received at Reader is ignored to remove ISI. We have for each block j and $i \in [2 : N]$ that

$$\begin{aligned} y_{ji} = {} & g_{\mathrm{r}1} x_{ji} \left(g_{\mathrm{t}1} s_i + g_{\mathrm{t}2} s_{i-1}\right) \\ & + g_{\mathrm{r}2} x_{ji-1} \left(g_{\mathrm{t}1} s_{i-1} + g_{\mathrm{t}2} s_{i-2}\right) + w_{ji} \end{aligned} \tag{E.1}$$

with $x_{ji} \in \{0, 1\}$ drawn equiprobably. The choice of $x_{ji} \in \{0, 1\}$ corresponds to ON-OFF keying, this is commonly used in backscatter communications, since Tag can choose to not reflect the carrier, or to reflect the carrier, by adjusting its impedence, thus creating the ON and OFF states. Adjusting the impedence in more levels can lead x_{ji} to take discrete values from a larger set. The noise w_{ij} is circularly symmetric complex Gaussian distributed and independent across i, j. We use the notation $\mathcal{CN}\left(\mu, \sigma^2\right)$ for circularly symmetric complex Gaussian distribution with mean μ and variance σ^2. We assume the power constraint $\frac{1}{N} \sum_{i=1}^{N} |s_i|^2 \leq P$ on the carrier sequence. Note that we can scale $g_{\mathrm{r}2}, g_{\mathrm{t}2}, P$ to assume without loss of generality, that $g_{\mathrm{r}1} = g_{\mathrm{t}1} = 1$ and $w_{ji} \sim \mathcal{CN}(0,1)$. For a given power P, the problem is to design the sequence $\{s_1, \ldots, s_N\}$ to maximize the mutual information rate achievable. Alternately for a fixed rate, we can optimize the sequence $\{s_1, \ldots, s_N\}$ to use minimal power. We also investigate the use of channel codes and the performance in terms of bit-error rate.

Now we derive the form of the effective channel between Tag and Reader for our proposed model and formulate the optimization problem for maximizing the mutual information between Tag and Reader. We also deal with the issue of channel training to obtain the necessary parameters for optimization and formulate approximate optimization techniques. We assume $N = 5$ for simplicity of analysis, but it can be easily extended to arbitrary N case. Emitter sends the sequence $s_1, s_2, s_3, s_4, s_5, s_1, s_2, s_3, s_4, s_5, s_1, s_2, s_3, s_4, s_5, \ldots$ and thus Tag receives a periodic version of

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix} = \begin{pmatrix} g_{\mathrm{t}1} & 0 & 0 & 0 & g_{\mathrm{t}2} \\ g_{\mathrm{t}2} & g_{\mathrm{t}1} & 0 & 0 & 0 \\ 0 & g_{\mathrm{t}2} & g_{\mathrm{t}1} & 0 & 0 \\ 0 & 0 & g_{\mathrm{t}2} & g_{\mathrm{t}1} & 0 \\ 0 & 0 & 0 & g_{\mathrm{t}2} & g_{\mathrm{t}1} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{pmatrix}. \tag{E.2}$$

Tag sends blocks of cyclic-suffixed symbols of the form $x_{j1}, x_{j2}, x_{j3}, x_{j4}, x_{j1}$ and every first symbol received at Reader is ignored to remove the ISI from the previous block. Hence

$$\begin{aligned} Y_j &= \begin{pmatrix} g_{\mathrm{r}2} & g_{\mathrm{r}1} & 0 & 0 & 0 \\ 0 & g_{\mathrm{r}2} & g_{\mathrm{r}1} & 0 & 0 \\ 0 & 0 & g_{\mathrm{r}2} & g_{\mathrm{r}1} & 0 \\ 0 & 0 & 0 & g_{\mathrm{r}2} & g_{\mathrm{r}1} \end{pmatrix} \begin{pmatrix} u_1 x_{j1} \\ u_2 x_{j2} \\ u_3 x_{j3} \\ u_4 x_{j4} \\ u_5 x_{j1} \end{pmatrix} + W_j \\ &= \begin{pmatrix} g_{\mathrm{r}2} u_1 & g_{\mathrm{r}1} u_2 & 0 & 0 & 0 \\ 0 & g_{\mathrm{r}2} u_2 & g_{\mathrm{r}1} u_3 & 0 & 0 \\ 0 & 0 & g_{\mathrm{r}2} u_3 & g_{\mathrm{r}1} u_4 & 0 \\ 0 & 0 & 0 & g_{\mathrm{r}2} u_4 & g_{\mathrm{r}1} u_5 \end{pmatrix} \end{aligned}$$

$$
\times \begin{pmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \\ x_{j4} \\ x_{j1} \end{pmatrix} + W_j
$$

$$
\begin{aligned}
&= \begin{pmatrix} g_{\mathrm{r}2}u_1 & g_{\mathrm{r}1}u_2 & 0 & 0 \\ 0 & g_{\mathrm{r}2}u_2 & g_{\mathrm{r}1}u_3 & 0 \\ 0 & 0 & g_{\mathrm{r}2}u_3 & g_{\mathrm{r}1}u_4 \\ g_{\mathrm{r}1}u_5 & 0 & 0 & g_{\mathrm{r}2}u_4 \end{pmatrix} \begin{pmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \\ x_{j4} \end{pmatrix} + W_j \\
&= GX_j + W_j, \qquad \text{(E.3)}
\end{aligned}
$$

where

$$
G = \begin{pmatrix} g_4s_5 + g_3s_1 & g_2s_1 + g_1s_2 & 0 & 0 \\ 0 & g_4s_1 + g_3s_2 & g_2s_2 + g_1s_3 & 0 \\ 0 & 0 & g_4s_2 + g_3s_3 & g_2s_3 + g_1s_4 \\ g_2s_4 + g_1s_5 & 0 & 0 & g_4s_3 + g_3s_4 \end{pmatrix},
$$

$$
X_j = \begin{pmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \\ x_{j4} \end{pmatrix}
$$

with $g_1 = g_{\mathrm{r}1}g_{\mathrm{t}1}, g_2 = g_{\mathrm{r}1}g_{\mathrm{t}2}, g_3 = g_{\mathrm{r}2}g_{\mathrm{t}1}$ and $g_4 = g_{\mathrm{r}2}g_{\mathrm{t}2}$. Here X_j has i.i.d. elements equiprobably drawn from $\{0,1\}$ *i.e.,* ON-OFF keying is used. The noise W_j has i.i.d. $\mathcal{CN}(0,1)$ elements. Now the optimization problem is

$$
\underset{s_i|_{i=1}^{5}}{\text{maximize}} \ \frac{1}{4} I(X;Y). \qquad \text{(E.4)}
$$

The $\{X_i\}$'s are taken from the corners of 4-dimensional hypercube $\{0,1\}^4$, we have

$$
\begin{aligned}
&h(GX+W) \\
&= -\int \sum_i \frac{1}{2^4}\frac{1}{\pi^4} \exp\left(-(Y-GX_i)^\dagger (Y-GX_i)\right) \times \\
&\quad \log\left(\sum_j \frac{1}{2^4}\frac{1}{\pi^4} \exp\left(-(Y-GX_j)^\dagger (Y-GX_j)\right)\right) dY
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_i \int \frac{1}{2^4}\frac{1}{\pi^4} \exp\left(-\left(Y - GX_i\right)^\dagger \left(Y - GX_i\right)\right) \times \\
&\qquad \log\left(\sum_j \frac{1}{2^4}\frac{1}{\pi^4} \exp\left(-\left(Y - GX_j\right)^\dagger \left(Y - GX_j\right)\right)\right) dY \\
&= -\sum_i I_i
\end{aligned} \tag{E.5}
$$

as a sum of integrals I_i and $h\left(Y\middle| X\right) = h\left(W\right) = 4 \times \log\left(\pi e\right)$.

E.1.1 Approximate optimization techniques

We now deal with approximation techniques for the optimization problem, since we do not have a closed form expression for the integrals in (E.5). We first obtain an approximation for $h\left(GX + W\right) = -\sum_i I_i$ from (E.5). In order to evaluate I_i, we use Taylor expansion on $\log\left(\sum_j \left(1/2^4\right)\left(1/\pi^4\right) \exp\left(-\left(Y - GX_j\right)^\dagger \left(Y - GX_j\right)\right)\right)$around X_i (up to order 2) and perform the integral I_i. Thus we need 2^4 different Taylor expansions. With this approximation for $h\left(GX + W\right)$, we can optimize it over $s_i|_{i=1}^5$. As a proxy we propose the following simpler optimization problems:

$$
\underset{s_i|_{i=1}^5}{\text{maximize}} \ \log\left(\det\left(G^\dagger G + I\right)\right) \tag{E.6}
$$

or

$$
\underset{s_i|_{i=1}^5}{\text{maximize}} \ \log\left(\det\left(G^\dagger G\right)\right). \tag{E.7}
$$

After solving the simper optimization problems, we can substitute the approximate solutions into the actual function to be optimized (we evaluate the performance numerically in the next section). Note that $\det\left(G^\dagger G\right) = \det\left(G^\dagger\right)\det\left(G\right)$. We have

$$
\begin{aligned}
&\det\left(G\right) \\
&= \left(g_4 s_5 + g_3 s_1\right)\left(g_4 s_1 + g_3 s_2\right)\left(g_4 s_2 + g_3 s_3\right)\left(g_4 s_3 + g_3 s_4\right) \\
&\quad - \left(g_2 s_1 + g_1 s_2\right)\left(g_2 s_2 + g_1 s_3\right)\left(g_2 s_3 + g_1 s_4\right)\left(g_2 s_4 + g_1 s_5\right).
\end{aligned}
$$

We now try to evaluate the gradient of $G^\dagger G$ for simulations (we do not have a simple closed form expression for $\det\left(G^\dagger G + I\right)$). Looking at the term

$(g_4s_5+g_3s_1)(g_4s_1+g_3s_2)(g_4s_2+g_3s_3)(g_4s_3+g_3s_4)$, we define

$$\begin{aligned}
\bar{t_1} &= (g_4s_1+g_3s_2)(g_4s_2+g_3s_3)(g_4s_3+g_3s_4)\\
\bar{t_2} &= (g_4s_5+g_3s_1)\,s_2\,(g_4s_2+g_3s_3)(g_4s_3+g_3s_4)\\
&\vdots
\end{aligned}$$

etc and looking at the term $(g_4s_5+g_3s_1)(g_4s_1+g_3s_2)(g_4s_2+g_3s_3)(g_4s_3+g_3s_4)$, we define

$$\begin{aligned}
\underline{t_1} &= (g_2s_2+g_1s_3)(g_2s_3+g_1s_4)(g_2s_4+g_1s_5)\\
\underline{t_2} &= (g_2s_1+g_1s_2)(g_2s_3+g_1s_4)(g_2s_4+g_1s_5)\\
&\vdots
\end{aligned}$$

Then we have

$$\begin{aligned}
\nabla_s\det(G) &= \begin{bmatrix} g_3\bar{t_1}+g_4\bar{t_2}-g_2\underline{t_1}\\ g_3\bar{t_2}+g_4\bar{t_3}-g_1\underline{t_1}-g_2\underline{t_2}\\ g_3\bar{t_3}+g_4\bar{4}-g_1\underline{t_2}-g_2\underline{t_3}\\ g_3\bar{t_4}-g_1\underline{t_3}-g_2\underline{t_4}\\ g_4\bar{t_1}-g_1\underline{t_4}\end{bmatrix}\\
&= G_\nabla.
\end{aligned}$$

Hence we have

$$\begin{aligned}
\nabla_{\mathrm{Re}(s)}\left(\det(G)^\dagger\det(G)\right) &= G_\nabla^\dagger\det(G)+\det(G)^\dagger G_\nabla\\
&= 2\times\mathrm{Real}\left(G_\nabla^\dagger\det(G)\right)\\
\nabla_{\mathrm{Im}(s)}\left(\det(G)^\dagger\det(G)\right) &= -1iG_\nabla^\dagger\det(G)+1i\times\det(G)^\dagger G_\nabla\\
&= 2\times\mathrm{Real}\left(-1iG_\nabla^\dagger\det(G)\right).
\end{aligned}$$

E.1.2 Observations on non-optimality of constant carrier

We provide two examples when a constant carrier sequence can be shown to be not optimal. We have

$$\det(G)$$

Table E.1: Training scheme

Time slot	1	2	3	4
Emitter signal	$\sqrt{P}$	$\sqrt{P}$	0	0
Tag state	ON	OFF	ON	OFF
Reflected signal	$g_{t1}\sqrt{P}$	0	$g_{t2}\sqrt{P}$	0
Signal received at reader	$g_{r1}g_{t1}\sqrt{P}+w_1$	$g_{r2}g_{t1}\sqrt{P}+w_2$	$g_{r1}g_{t2}\sqrt{P}+w_3$	$g_{r2}g_{t2}\sqrt{P}+w_4$

$$= (g_4s_5 + g_3s_1)(g_4s_1 + g_3s_2)(g_4s_2 + g_3s_3)(g_4s_3 + g_3s_4)$$
$$- (g_2s_1 + g_1s_2)(g_2s_2 + g_1s_3)(g_2s_3 + g_1s_4)(g_2s_4 + g_1s_5)$$

1. Suppose $g_{\mathrm{r1}} = g_{\mathrm{r2}}$, since $g_1 = g_{\mathrm{r1}}g_{\mathrm{t1}}, g_2 = g_{\mathrm{r1}}g_{\mathrm{t2}}, g_3 = g_{\mathrm{r2}}g_{\mathrm{t1}}, g_4 = g_{\mathrm{r2}}g_{\mathrm{t2}}$. This implies that $g_4 = g_2$, $g_3 = g_1$. Then
$$\det(G) = (g_4 + g_3)(s_1 + s_2)(s_2 + s_3)(s_3 + s_4)(s_1 - s_4).$$
Then it is clear that setting $s_1 = -s_4, s_5 = 0$ is better than setting $s_1 = s_4 = s_5$ for maximizing $\det(G)$.

2. Suppose $g_{\mathrm{t1}} = -g_{\mathrm{t2}}$. Then $g_4 = -g_3$ and $g_2 = -g_1$. Then it is clear that setting all s_i's to be zero yields G to be a zero matrix.

E.1.3 Simulation with the channel training

We need $\{g_{\mathrm{r1}}g_{\mathrm{t1}}, g_{\mathrm{r2}}g_{\mathrm{t1}}, g_{\mathrm{r1}}g_{\mathrm{t2}}, g_{\mathrm{r2}}g_{\mathrm{t2}}\}$ for the optimization problem. This is obtained by training as follows: let Emitter send a sequence $\sqrt{P}, \sqrt{P}, 0, 0$ and Tag follow the sequence {ON,OFF,ON,OFF}, then Reader receives $\left\{g_{\mathrm{r1}}g_{\mathrm{t1}}\sqrt{P} + w_1, g_{\mathrm{r2}}g_{\mathrm{t1}}\sqrt{P} + w_2, g_{\mathrm{r1}}g_{\mathrm{t2}}\sqrt{P} + w_3, \quad g_{\mathrm{r2}}g_{\mathrm{t2}}\sqrt{P} + w_4\right\}$. The training scheme is illustrated in Table E.1. The training can be repeated K times to refine the coefficients to $\left\{g_{\mathrm{r1}}g_{\mathrm{t1}}\sqrt{P} + \frac{1}{K}w_1, g_{\mathrm{r2}}g_{\mathrm{t1}}\sqrt{P} + \frac{1}{K}w_2, g_{\mathrm{r1}}g_{\mathrm{t2}}\sqrt{P} + \frac{1}{K}w_3, \quad g_{\mathrm{r2}}g_{\mathrm{t2}}\sqrt{P} + \frac{1}{K}w_4\right\}$.

With this training scheme, only the approximate coefficients are known to the collaborating Emitter and Reader. Hence the carrier sequence is designed using the approximate coefficients substituted to the optimization problems. Also the decoder uses the approximate channel coefficients to decode the data.

E.1.4 Simulation details

We obtain the results averaging over $\{g_{t1}, g_{t2}\} = \left\{1, \frac{1}{2}\mathcal{CN}(0,1)\right\}$ and $\{g_{r1}, g_{r2}\} = \left\{1, \frac{1}{2}\mathcal{CN}(0,1)\right\}$ and block length $N = 5$. We obtain the results for optimizing the mutual information in Figure 6.4. Note that the maximum rate achievable is 1 since we use ON-OFF keying. Compared to using a constant carrier, we have a gain of 3.51 dB at rate 0.8 using the actual optimization problem (E.4), and we have a gain of 3.36 dB at rate 0.8, when we use $\det\left(G^{\dagger}G + I\right)$ to approximate mutual information. Also we have a gain of 3.47 dB at rate 0.8, when we use $\det\left(G^{\dagger}G\right)$ to approximate mutual information. This suggests that using $\det\left(G^{\dagger}G\right)$ for optimization can give most of the gain and since it is numerically simpler, we use it in the subsequent bit error rate (BER) simulations.

For the figures 6.5 and 6.6, we send 10,000 packets per channel over 100 randomly generated channels. Each packet is of size 57 bits and is encoded using a (57,63) Hamming code. In Figure 6.5 we use ON-OFF keying for modulation and a zero-forcing (ZF) channel matrix equalizer. The channel is assumed to be known perfectly at the receiver and the transmitter. At BER of 10^{-3}, we observe a gain of about 5 dB.

In Figure 6.6 we use the same setup as above, except that the channel is assumed to be obtained by training once, with additive $\mathcal{CN}(0,1)$ noise as described in Section E.1.3. At BER of 10^{-3}, we still observe a gain of about 5 dB, thus the robustness of our optimization technique is demonstrated.

References

[ADT11] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse. "Wireless Network Information Flow: A Deterministic Approach." *IEEE Transactions on Information Theory*, **57**(4):1872–1905, April 2011.

[ADT15] A. S. Avestimehr, S. N. Diggavi, Chao Tian, and D. N. C. Tse. "An Approximation Approach to Network Information Theory." *Foundations and Trends in Communications and Information Theory*, **12**(1–2):1–183, 2015.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables.* Number 55. Courier Corporation, 1964.

[ASC09] V. Aggarwal, L. Sankar, A. R. Calderbank, and H. V. Poor. "Ergodic layered erasure one-sided interference channels." In *IEEE Information Theory Workshop*, pp. 574–578, Oct 2009.

[ATS01] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. "The capacity of discrete-time memoryless Rayleigh-fading channels." *IEEE Transactions on Information Theory*, **47**(4):1290–1301, May 2001.

[Bat08] N. Batir. "Inequalities for the gamma function." *Archiv der Mathematik*, **91**(6):554–563, 2008.

[BLM14] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Sukhavasi, Chirag Patel, and Stefan Geirhofer. "Network densification: the dominant theme for wireless evolution into 5G." *IEEE Communications Magazine*, **52**(2):82–89, 2014.

[BPY15] G. Bassi, P. Piantanida, and S. Yang. "Capacity Bounds for a Class of Interference Relay Channels." *IEEE Transactions on Information Theory*, **61**(7):3698–3721, July 2015.

[BR14] C. Boyer and S. Roy. "Backscatter Communication and RFID: coding, energy, and MIMO Analysis." *IEEE Transactions on Communications*, **62**(3):770–785, March 2014.

[CJ08] V. R. Cadambe and S. A. Jafar. "Interference Alignment and Degrees of Freedom of the K-User Interference Channel." *IEEE Transactions on Information Theory*, **54**(8):3425–3441, Aug 2008.

[CMG08] H-F. Chong, M. Motani, H. K. Garg, and H. El Gamal. “On the Han-Kobayashi region for the interference channel.” *IEEE Transactions on Information Theory*, **54**(7):3188–3194, 2008.

[CT12] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2012.

[CTK14] M. Cardone, D. Tuninetti, R. Knopp, and U. Salim. “Gaussian Half-Duplex Relay Networks: Improved Constant Gap and Connections With the Assignment Problem.” *IEEE Transactions on Information Theory*, **60**(6):3559–3575, Jun 2014.

[Dob12] D.M. Dobkin. *The RF in RFID: UHF RFID in practice*. Elsevier Science, 2012.

[ETW08] R. H. Etkin, D. N. C. Tse, and H. Wang. “Gaussian interference channel capacity to within one bit.” *IEEE Transactions on Information Theory*, **54**(12):5534–5562, 2008.

[Far13] R. K. Farsani. “The capacity region of the wireless ergodic fading Interference Channel with partial CSIT to within one bit.” In *IEEE International Symposium on Information Theory*, pp. 759–763, July 2013.

[GCS16] S. Gherekhloo, A. Chaaban, and A. Sezgin. “Cooperation for Interference Management: A GDoF Perspective.” *IEEE Transactions on Information Theory*, **62**(12):6986–7029, Dec 2016.

[GK00] P. Gupta and P. R. Kumar. “The capacity of wireless networks.” *IEEE Transactions on Information Theory*, **46**(2):388–404, Mar 2000.

[GY14] R. H. Gohary and H. Yanikomeroglu. “Grassmannian Signalling Achieves Tight Bounds on the Ergodic High-SNR Capacity of the Noncoherent MIMO Full-Duplex Relay Channel.” *IEEE Transactions on Information Theory*, **60**(5):2480–2494, May 2014.

[HK81] T. S. Han and K. Kobayashi. “A new achievable rate region for the interference channel.” *IEEE Transactions on Information Theory*, **27**(1):49–60, Jan 1981.

[IDM11] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel. “Coordinated multipoint: Concepts, performance, and field trial results.” *IEEE Communications Magazine*, **49**(2):102–111, February 2011.

[KC13] M. G. Kang and W. Choi. “Ergodic Interference Alignment With Delayed Feedback.” *IEEE Signal Processing Letters*, **20**(5):511–514, May 2013.

[KD17] C. Karakus and S. N. Diggavi. “Enhancing Multiuser MIMO Through Opportunistic D2D Cooperation.” *IEEE Transactions on Wireless Communications*, **16**(9):5616–5629, Sept 2017.

[KK13] T. Koch and G. Kramer. “On Noncoherent Fading Relay Channels at High Signal-to-Noise Ratio.” *IEEE Transactions on Information Theory*, **59**(4):2221–2241, April 2013.

[KOG15] R. Kolte, A. Ozgur, and A. El Gamal. “Capacity Approximations for Gaussian Relay Networks.” *IEEE Transactions on Information Theory*, **61**(9):4721–4734, Sept 2015.

[KRD14] G. Koliander, E. Riegler, G. Durisi, and F. Hlawatsch. “Degrees of Freedom of Generic Block-Fading MIMO Channels Without apriori Channel State Information.” *IEEE Transactions on Information Theory*, **60**(12):7760–7781, Dec 2014.

[Lap05] A. Lapidoth. “On the high-SNR capacity of noncoherent networks.” *IEEE Transactions on Information Theory*, **51**(9):3025–3036, Sept 2005.

[LKG11] S. H. Lim, Y. H. Kim, A. El Gamal, and S. Y. Chung. “Noisy Network Coding.” *IEEE Transactions on Information Theory*, **57**(5):3132–3152, May 2011.

[LM03] A. Lapidoth and S. M. Moser. “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels.” *IEEE Transactions on Information Theory*, **49**(10):2426–2467, 2003.

[MDG12] I. Maric, R. Dabora, and A. J. Goldsmith. “Relaying in the Presence of Interference: Achievable Rates, Interference Forwarding, and Outer Bounds.” *IEEE Transactions on Information Theory*, **58**(7):4342–4354, July 2012.

[MH99] T. L. Marzetta and B. M. Hochwald. “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading.” *IEEE Transactions on Information Theory*, **45**(1):139–157, 1999.

[MJS12] H. Maleki, S. A. Jafar, and S. Shamai. “Retrospective Interference Alignment Over Interference Networks.” *IEEE Journal of Selected Topics in Signal Processing*, **6**(3):228–240, June 2012.

[MRY13] V. I. Morgenshtern, E. Riegler, W. Yang, G. Durisi, S. Lin, B. Sturmfels, and H. Bolcskei. “Capacity Pre-Log of Noncoherent SIMO Channels Via Hironaka’s Theorem.” *IEEE Transactions on Information Theory*, **59**(7):4213–4229, July 2013.

[MT10] M. A. Maddah-Ali and D. N. C. Tse. "On the degrees of freedom of MISO broadcast channels with delayed feedback." *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2010-122*, 2010.

[ND10] U. Niesen and S. N. Diggavi. "Non-coherent hierarchical cooperation." In *Annual Allerton Conference on Communication, Control, and Computing*, pp. 507–513, Sept 2010.

[ND13] U. Niesen and S. N. Diggavi. "The Approximate Capacity of the Gaussian n-Relay Diamond Network." *IEEE Transactions on Information Theory*, **59**(2):845–859, Feb 2013.

[NGJ12] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath. "Ergodic interference alignment." *IEEE Transactions on Information Theory*, **58**(10):6355–6371, 2012.

[NOF14] C. Nazaroglu, A. Ozgur, and C. Fragouli. "Wireless Network Simplification: The Gaussian N-Relay Diamond Network." *IEEE Transactions on Information Theory*, **60**(10):6329–6341, Oct 2014.

[NYG17] K. Ngo, S. Yang, and M. Guillaud. "An achievable DoF region for the two-user non-coherent MIMO broadcast channel with statistical CSI." In *IEEE Information Theory Workshop*, pp. 604–608, Nov 2017.

[OD10] A. Ozgur and S. N. Diggavi. "Approximately achieving Gaussian relay network capacity with lattice codes." In *2010 IEEE International Symposium on Information Theory*, pp. 669–673, June 2010.

[OD13] A. Ozgur and S. N. Diggavi. "Approximately Achieving Gaussian Relay Network Capacity With Lattice-Based QMF Codes." *IEEE Transactions on Information Theory*, **59**(12):8275–8294, Dec 2013.

[OLT07] A. Ozgur, O. Leveque, and D. N. C. Tse. "Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks." *IEEE Transactions on Information Theory*, **53**(10):3549–3572, Oct 2007.

[PDT09] E. Perron, S. N. Diggavi, and E. Telatar. "The Interference-Multiple-Access Channel." In *2009 IEEE International Conference on Communications*, pp. 1–5, June 2009.

[PV11] V. M. Prabhakaran and P. Viswanath. "Interference Channels With Source Cooperation." *IEEE Transactions on Information Theory*, **57**(1):156–186, Jan 2011.

[SD18a] J. Sebastian and S. N. Diggavi. "Generalized Degrees of Freedom of Noncoherent

Diamond Networks." *arXiv*, 2018.

[SD18b] J. Sebastian and S. N. Diggavi. "Generalized Degrees of Freedom of Noncoherent Interference Channel." *http://www.seas.ucla.edu/~joyson/Documents/Non_coherent_IC.pdf*, 2018.

[SD18c] J. Sebastian and S. N. Diggavi. "On Capacity of Noncoherent MIMO with Asymmetric Link Strengths." *arXiv*, 2018.

[SE07] O. Sahin and E. Erkip. "Achievable Rates for the Gaussian Interference Relay Channel." In *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1627–1631, Nov 2007.

[SED18] J. Sebastian, Y. Ezzeldin, S. N. Diggavi, and C. Fragouli. "Carrier optimization for backscatter communication." *http://www.seas.ucla.edu/~joyson/Documents/Backscattering_letter.pdf*, 2018.

[SG00] B. Schein and R. Gallager. "The Gaussian parallel relay network." In *IEEE International Symposium on Information Theory*, p. 22, June 2000.

[SK66] J. Schalkwijk and T. Kailath. "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint." *IEEE Transactions on Information Theory*, **12**(2):172–182, 1966.

[SKD17] J. Sebastian, C. Karakus, and S. N. Diggavi. "Approximate Capacity of Fast Fading Interference Channels with no Instantaneous CSIT." *http://arxiv.org/abs/1706.03659*, 2017.

[SKD18] J. Sebastian, C. Karakus, and S. N. Diggavi. "Approximate Capacity of Fast Fading Interference Channels with No Instantaneous CSIT." *IEEE Transactions on Communications*, preprint, 2018.

[SSE11] L. Sankar, X. Shang, E. Erkip, and H. V. Poor. "Ergodic Fading Interference Channels: Sum-Capacity and Separability." *IEEE Transactions on Information Theory*, **57**(5):2605–2626, May 2011.

[ST11] 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, May 2011.

[SWF14] A. Sengupta, I. H. Wang, and C. Fragouli. "Cooperative Relaying at Finite SNR; Role of Quantize-Map-and-Forward." *IEEE Transactions on Wireless Communications*, **13**(9):4857–4870, Sept 2014.

[TE97] G. Taricco and M. Elia. "Capacity of fading channel with no side information." *Electronics Letters*, **33**(16):1368–1370, Jul 1997.

[TMP13] R. Tandon, S. Mohajer, H. V. Poor, and S. Shamai. "Degrees of Freedom Region of the MIMO Interference Channel With Output Feedback and Delayed CSIT." *IEEE Transactions on Information Theory*, **59**(3):1444–1457, March 2013.

[TY11] Y. Tian and A. Yener. "The Gaussian Interference Relay Channel: Improved Achievable Rates and Sum Rate Upperbounds Using a Potent Relay." *IEEE Transactions on Information Theory*, **57**(5):2865–2879, May 2011.

[VMA14] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr. "Capacity Results for Binary Fading Interference Channels With Delayed CSIT." *IEEE Transactions on Information Theory*, **60**(10):6093–6130, Oct 2014.

[VMA17] A. Vahid, M. A. Maddah-Ali, A. S. Avestimehr, and Y. Zhu. "Binary Fading Interference Channel With No CSIT." *IEEE Transactions on Information Theory*, **63**(6):3565–3578, June 2017.

[VSA12] A. Vahid, C. Suh, and S. Avestimehr. "Interference Channels With Rate-Limited Feedback." *IEEE Transactions on Information Theory*, **58**(5):2788–2812, May 2012.

[VV12] 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, July 2012.

[WHH05] N.A. Weiss, P.T. Holmes, and M. Hardy. *A Course in Probability*. Pearson Addison Wesley, 2005.

[WSD13] H. Wang, C. Suh, S. N. Diggavi, and P. Viswanath. "Bursty interference channel with feedback." In *IEEE International Symposium on Information Theory*, pp. 21–25. IEEE, 2013.

[WT11a] I. H. Wang and D. N. C. Tse. "Interference Mitigation Through Limited Receiver Cooperation." *IEEE Transactions on Information Theory*, **57**(5):2913–2940, May 2011.

[WT11b] I. H. Wang and D. N. C. Tse. "Interference Mitigation Through Limited Transmitter Cooperation." *IEEE Transactions on Information Theory*, **57**(5):2941–2965, May 2011.

[WZH15] Jun Wu, Zhifeng Zhang, Yu Hong, and Yonggang Wen. "Cloud radio access network (C-RAN): a primer." *IEEE Network*, **29**(1):35–41, 2015.

[XYV14] L. Xie, Y. Yin, A. V. Vasilakos, and S. Lu. "Managing RFID data: challenges, opportunities and solutions." *IEEE Communications Surveys Tutorials*, **16**(3):1294–1311, Third 2014.

[YDR13] W. Yang, G. Durisi, and E. Riegler. "On the Capacity of Large-MIMO Block-Fading Channels." *IEEE Journal on Selected Areas in Communications*, **31**(2):117–132, February 2013.

[ZG11] Y. Zhu and D. Guo. "Ergodic Fading Z-Interference Channels Without State Information at Transmitters." *IEEE Transactions on Information Theory*, **57**(5):2627–2647, May 2011.

[ZT02] L. Zheng and D. N. C. Tse. "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel." *IEEE Transactions on Information Theory*, **48**(2):359–383, Feb 2002.

[ZT03] Lizhong Zheng and D. N. C. Tse. "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels." *IEEE Transactions on Information Theory*, **49**(5):1073–1096, May 2003.