UNIVERSITY OF CALIFORNIA,
IRVINE

Optimal Use of Multiple Antennas in Interference Networks – MIMO, Interference Alignment and Beyond

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical Engineering

by

Bofeng Yuan

Dissertation Committee:
Professor Syed Jafar, Chair
Professor Ender Ayanoglu
Professor Weining Shen

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Professor Syed Jafar. I have benefited tremendously from his insight, vision and immense knowledge. My doctoral study and research has been a great adventure for me in the area of information theory which I could not accomplish without his omnipresent guidance and valuable feedback. In addition, the joy and enthusiasm he has for his research has inspired and motivated me for the past four years and will continue to be in the future. It has been a great honor for me to be his Ph.D. student.

I would also like to thank Professor Ender Ayanoglu and Professor Weining Shen for serving on my dissertation committee, and Professor Zhiying Wang, Professor Marco Levorato for serving on my qualifying examination committee. I am grateful to all my committee members for their insightful suggestions.

I would also like to extend my gratitude to Hua Sun and Arash Davoodi for collaborating and co-authoring papers with me. I would also like to thank my former and current colleagues including Chunhua Geng, Zhuqing Jia, Zhen Chen and Yao-Chia Chan. I am proud of being a part of such a wonderful group.

Finally, but by no means least, thanks go to my girlfriend Dongning and my parents for their unconditional support. They are the most important people in my world and I dedicate this dissertation to them.

CURRICULUM VITAE

Bofeng Yuan

EDUCATION

Doctor of Philosophy in Electrical Engineering **2018**
University of California, Irvine *Irvine, California*

Master of Science in Information and Communication Engineering **2014**
Xi'an Jiaotong University *Xi'an, China*

Bachelor of Engineering in Information Engineering **2011**
Xi'an Jiaotong University *Xi'an, China*

ABSTRACT OF THE DISSERTATION

Optimal Use of Multiple Antennas in Interference Networks – MIMO, Interference Alignment and Beyond

By

Bofeng Yuan

Doctor of Philosophy in Electrical Engineering

University of California, Irvine, 2018

Professor Syed Jafar, Chair

Degrees of freedom (DoF) studies of wireless networks have contributed many fundamental insights into their capacity limits. One of the most critical determinants of these capacity limits is the amount of channel state information at the transmitters (CSIT). In this dissertation, we consider the MIMO interference networks with both perfect CSIT and CSIT uncertainty. In particular, a novel class of replication-based outer bounds will first be presented for arbitrary rank-constrained MIMO interference networks with perfect CSIT. It creates a new perspective of the capacity problem, so that even simple arguments such as user cooperation become quite powerful when applied in the replicated network, giving rise to stronger outer bounds, than when applied directly in the original network. Then, we prove that when CSIT is not perfect, signal space partitioning schemes can be DoF optimal. An interesting idea that emerges from this study is “elevated multiplexing” where the signals are split into streams and transmitted from separate antennas at elevated power levels, which allows these signals to be jointly decoded at one receiver which has fewer spatial dimensions with lower interference floors, while another receiver is simultaneously able to separately decode these signals with a higher interference floor but across a greater number of spatial dimensions. Finally, we explore the compatibility of various approaches under CSIT uncertainty which only been studied in isolation before, such as Blind Interference Alignment

(BIA) and partial zero-forcing. Coding schemes are proposed that jointly exploit partial channel knowledge and reconfigurable antennas, demonstrating synergistic DoF gains over what is achievable with either BIA or beamforming by itself.

Chapter 1

Introduction

The focus of this dissertation is studying the optimal use of multiple antennas in interference networks from a information theoretical prospective, which involves ideas such as multiple-input-multiple-output (MIMO), Interference Alignment and also the search for new ideas beyond these as well.

1.1 Background

Channel capacity, i.e., the highest possible rate at which reliable communication can occur over a communication channel, has been the holy grail for information theorists over decades. The most commonly studied channel is the additive white Gaussian noise channel. The capacity of this channel takes the simple logarithmic form, i.e., $C = \frac{1}{2}\log(1 + SNR)$ in the units of bits/channel use. For a bandwidth B, the capacity is expressed as $C = B\log(1 + SNR)$ in the units bits/second, where B appears as the pre-log term and more generally referred to as the Degrees of Freedom (DoF). The significance of DoF lies in the fact that capacity grows only logarithmically with power, but essentially linearly with DoF.

Therefore, the most efficient way to increasing the capacity of a network is to increasing its DoF. However, there are two problems with this. First, bandwidth, or frequency spectrum, is a limited and very expensive resource. Second, as the same bandwidth B is shared among K users, e.g., by frequency division, then each user only gets B/K. The idea of using multiple antennas, known as MIMO, emerged in response to the first problem. It was shown that, for example, if there are M transmit antennas and M receive antennas then by joint processing of signals between the co-located antennas, the DoF improve by a factor of M. A solution to the second problem was the idea of Interference Alignment, which says that by designing signals carefully so that the interference terms overlap with each other, one can maximize the DoF available to the desired signal free from interference. So for example, in a K user interference channel, where we have K distributed transmit antennas, one at each transmitter, and K distributed receive antennas, one at each receiver, the DoF improve by $K/2$. This is also paraphrased as "everyone gets half the cake". Clearly there is a price for distributed processing, but that is no more than half. This is an extremely positive result, but it relies on an extremely optimistic assumption, i.e., perfect channel knowledge is available.

As a key ingredient for interference management techniques, channel state information at transmitter (CSIT) can affect not only the capacity but also the DoF of interference networks. Although the DoF for a variety of networks has been well explored with perfect channel state information, the effect of the absence of channel information is still less understood. The DoF of MIMO interference channel (IC) with no CSIT is first studied in [21, 46]. Recently, Davoodi et al. in [6, 9] studies the finite precision CSIT setting for broadcast channel (BC). It is shown that when channel knowledge is absent or available only to finite precision, the DoF collapse, i.e., the gains of MIMO beamforming and interference alignment are lost.

In this dissertation, we will study the DoF for MIMO interference networks under different CSIT settings, i.e., perfect channel knowledge and channel uncertainty. This will leads us to the fundamental understanding of signal dimensions and power levels, the tradeoffs between joint processing and distributed processing of signals.

1.2 Overview of the Dissertation

In Chapter 2, in order to gain new insights into MIMO interference networks, the optimality of $\sum_{k=1}^{K} M_k/2$ (half the cake per user) degrees of freedom is explored for a K-user MIMO interference channel where the cross-channels have arbitrary rank constraints, and the k^{th} transmitter and receiver are equipped with M_k antennas each. The result consolidates and significantly generalizes results from prior studies by Krishnamurthy et al., of rank-deficient interference channels where all users have M antennas; and by Tang et al., of full rank interference channels where the k^{th} user pair has M_k antennas. The broader outcome of this work is a novel class of replication-based outer bounds for arbitrary rank-constrained MIMO interference networks where replicas of existing users are added as auxiliary users and the network connectivity is chosen to ensure that any achievable scheme for the original network also works in the new network. The replicated network creates a new perspective of the problem, so that even simple arguments such as user cooperation become quite powerful when applied in the replicated network, giving rise to stronger outer bounds, than when applied directly in the original network. Remarkably, the replication based bounds are broadly applicable not only to MIMO interference channels with arbitrary rank-constraints, but much more broadly, even beyond Gaussian settings.

In Chapter 3, the 2 user MIMO interference channel with arbitrary antenna configurations is studied under arbitrary levels of partial CSIT for each of the channels, to find the DoF achievable by User 2 while User 1 achieves his full interference-free DoF. The goal is to gain new insights due to the inclusion of MIMO into the signal space partitioning schemes associated with partial CSIT. An interesting idea that emerges from this study is "elevated multiplexing" where the signals are split into streams and transmitted from separate antennas at elevated power levels, which allows these signals to be jointly decoded at one receiver which has fewer spatial dimensions with lower interference floors, while another receiver is simultaneously able to separately decode these signals with a higher interference floor but across a greater number of spatial dimensions. Through elevated multiplexing we find that there is a DoF benefit from increasing the number of antennas at a transmitter even if that transmitter already has more antennas than its desired receiver and has no CSIT.

In Chapter 4, the DoF region of the two user MIMO interference channel (IC) are characterized under arbitrary levels of partial CSIT and arbitrary antenna configurations (M_1, M_2, N_1, N_2). This result bridges the gap between the DoF region of this channel under finite precision CSIT and perfect CSIT. The outer bounds of this paper are accomplished by a novel utilization of sum-set inequalities from Aligned Image Sets (AIS) approach.

In Chapter 5, we consider the two-user MIMO IC with reconfigurable antennas at Transmitter 1 and partial CSIT at Transmitter 2. Each node has arbitrary number of antennas. If reconfigurable antennas and partial CSIT are both useful, i.e., comparing with no CSIT one can achieve more DoF with either reconfigurable antennas or partial CSIT, our result indicate that there is always a synergistic DoF gain, i.e., we can achieve more by jointly exploiting reconfigurable antennas and partial CSIT than we could by exploiting each of them individually. Our new achievable schemes allow User 2 to carefully design the transmitted signal, so that BIA and partial zero-forcing approach can be used simultaneously. Then by analyzing the intuition behind the new schemes, we further conjecture that the sufficient condition for synergistic DoF gain to exist is also necessary. Furthermore, by introducing a tight DoF outer bound, the complete DoF region of this channel is characterized under the setting $N_1 < M_2 \leq N_2$.

1.3 Notations and Abbreviations

The sets $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{R}^n$ and $\mathbb{Q}^n$ stand for the sets of real numbers, rational numbers, all n-tuples of real numbers and all n-tuples of rational numbers, respectively. Moreover, the set $\mathbb{R}^{2+}$ is defined as the set of all pairs of non-negative numbers. We define $(A)^+ = \max(A, 0)$. $\min^+(A, B)$ is defined as follows $\min^+(A, B) = \max[\min(A, B), 0]$. If A is a set of random variables, then $H(A)$ refers to the joint entropy of the random variables in A. Conditional entropies, mutual information and joint and conditional probability densities of sets of random variables are similarly interpreted. Moreover, we use the Landau $O(\cdot)$ and $o(\cdot)$ notations as follows. For functions $f(x), g(x)$ from $\mathbb{R}$ to $\mathbb{R}$, $f(x) = O(g(x))$ denotes that $\limsup_{x\to\infty} \frac{|f(x)|}{|g(x)|} < \infty$. $f(x) = o(g(x))$ denotes that $\limsup_{x\to\infty} \frac{|f(x)|}{|g(x)|} = 0$. For any real number x, we define $\lfloor x \rfloor$ as the largest integer that is smaller than or equal to x when $x > 0$, the smallest integer that is larger than or equal to x when $x < 0$, and x itself when x is an integer. For any vector $\mathbf{Z} = \begin{bmatrix} Z_1 & Z_2 & \cdots Z_k \end{bmatrix}^T$ we define $\lfloor \mathbf{Z} \rfloor$ as $\begin{bmatrix} \lfloor Z_1 \rfloor & \lfloor Z_2 \rfloor & \cdots \lfloor Z_k \rfloor \end{bmatrix}^T$.

We denote the set $\{1, ..., K\}$ by $\mathcal{I}_K$ for a positive integer K. For a subset S of $\mathcal{I}_K$, $\mathcal{I}_K \backslash S$ denotes the set of elements that are in $\mathcal{I}_K$ but not in S, e.g., if $S = \{l\}, l \in \mathcal{I}_K$, then $\mathcal{I}_K \backslash l = \{1, ..., l-1, l+1, ..., K\}$. The cardinality of a set S is denoted as $|S|$. $\mathbf{I}_m$ denotes the $m \times m$ identity matrix and $\mathbf{0}_{m_1 \times m_2}$ denotes the $m_1 \times m_2$ matrix of zeros.

The following table lists the abbreviations used in this dissertation.

Table 1.1: Table of abbreviations

DoF	Degrees of Freedom
GDoF	Generalized Degrees of Freedom
SNR	Signal to Noise Ratio
CSIT	Channel State Information at Transmitters
MISO	Multiple Input Single Output
MIMO	Multiple Input Multiple Output
IC	Interference Channel
BC	Broadcast Channel
MAC	Multiple Access Channel
AIS	Aligned Image Set

Chapter 2

Replication-based Outer Bounds: On the Optimality of "Half the Cake" for Rank-Deficient MIMO Interference Networks

2.1 Problem Statement

Degrees of freedom (DoF) studies of wireless interference networks have produced a diverse array of new insights into the accessibility of signal dimensions under a variety of channel models. In order to consolidate these insights and to build upon them, it is important to make progress on unifying the underlying channel models. The motivation for this chapter, summarized in Figure 2.1, is to pursue such a generalization of the results from [2, 29, 30]. Specifically, in this chapter we start with the goal of consolidating the key insights regarding the optimality of half-the-cake (the "cake" refers to each user's interference-free DoF, cf.

[2]) for the K-user MIMO interference channel settings where the number of antennas at each receiver is equal to the number of antennas at the corresponding transmitter, i.e., all the *desired* channels are *square* matrices. The study of the unified setting leads us to a broader outcome – a novel class of replication-based outer bounds that are applicable not only to arbitrary rank-constrained MIMO interference networks but much more generally, even beyond Gaussian settings as well.

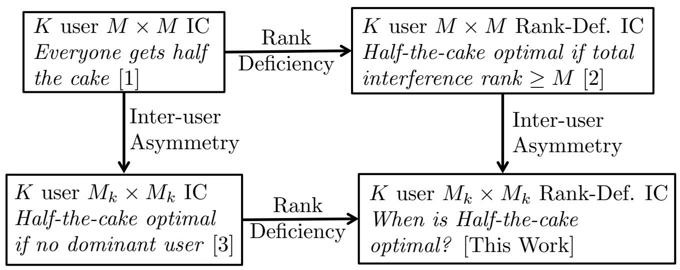

Figure 2.1: The motivation of this paper. Rank-Deficient is abbreviated as Rank-Def. .

2.1.1 Everyone Gets Half the Cake

It was shown by Cadambe and Jafar in [2] that in a K-user $M \times M$ MIMO interference channel where each node is equipped with M antennas, the optimal DoF value is $KM/2$. Since each user achieves half of his interference-free DoF, the result is often paraphrased as "everyone gets half the cake". Generalizations of this result have been explored in various directions, in particular to find out when the optimal solution may allow even more than half-the-cake. Indeed rectangular interference channels (cf. [15, 12, 40, 1, 41]), and multi-hop settings (cf. [17]) have shown that more than half-the-cake is possible. Of particular interest to us in this chapter are the generalizations in [29, 30].

2.1.2 Optimality of Half-the-cake: Key Insight from [29, 30]

The generalization in [29] concerns rank-deficient channels. Rank-deficient interference channels (cf. [4, 5, 45]) are frequently encountered due to poor scattering, keyhole effects, as well as underlying topological and structural concerns in single-hop abstractions of multihop networks with linear forwarding at intermediate nodes. Cross-channel rank-deficiencies have the potential to be helpful as the scope of zero forcing schemes is enhanced (although the scope of interference alignment schemes is limited by rank-deficiencies), opening the possibility that more than half-the-cake may be achievable. Exploring this possibility in [29], Krishnamurthy and Jafar establish that for the K-user $M \times M$ MIMO interference channel where all the cross channels are rank-deficient with the same rank $D \leq M$ and all the direct channels are full rank, $KM/2$ DoF (half-the-cake) are optimal if the sum of all interference ranks at each user, is greater than or equal to the number of antennas at the user, $(K-1)D \geq M$. In other words, *every signal dimension is accessible by at least one interfering user.* For $K = 3$ users, [29] considers a more general setting, so that at each receiver the interfering channel from the preceding transmitter is of rank D_1 and the interfering channel from the next transmitter (with wrap around) is of rank D_2. For $K = 2$ users the setting is fully general with all interfering channel ranks allowed to take arbitrary values. Remarkably, in all cases, the key insight remains the same:

Original Insight: "*Half-the-cake is optimal if at every transmitter and receiver, the sum of interfering channel ranks is greater than or equal to the number of antennas at that transmitter and receiver, respectively.*"

Finally, Liu, Tuninetti and Jafar in [30] consider a different generalization, to the K-user $M_k \times M_k$ MIMO interference channel with full rank generic channels, where the k^{th} user has M_k transmit and M_k receive antennas. For this setting [30] showed that half-the-cake is optimal provided there is no dominant user (a user with more antennas than all the rest of the users combined). Interestingly, this condition is also identical to the insight from [29] — once again, half-the-cake is optimal if the sum of interfering channel ranks is greater than or equal to the number of antennas at each user.

2.1.3 Overview

In order to further refine the key insight from [29, 30] and to identify its limitations, it is important to continue to test its validity under generalized settings. To this end, in this chapter we unify the channel models of [29] and [30] into the rank-deficient K-user $M_k \times M_k$ MIMO interference channel, and study the optimality of half-the-cake under *arbitrary* (no assumptions of symmetry) rank constraints on the cross-channels.

Surprisingly, we discover that the original insight fails in this generalized setting. Indeed, as a *counterexample* consider the 3-user MIMO interference channel with $M_1 = 10, M_2 = 8, M_3 = 6$, where the channel from Transmitter 1 to Receiver 2 has rank 5 and the channel from Transmitter 2 to Receiver 1 has rank 6. All other channels have full rank. Even though in this channel, the sum of interfering channel ranks at every user is greater than or equal to the number of antennas at that user, it is possible to achieve more than half-the-cake (half-the-cake is 12, but 12.5 DoF are achievable, as explained in Section 2.6). Therefore, a new outer bound is necessary for the K-user $M_k \times M_k$ MIMO interference channel.

Define $M_\Sigma = \sum_{k=1}^{K} M_k$. Define $\mathbf{H}_{ji}$ as the $M_j \times M_i$ channel matrix from Transmitter i to Receiver j, $i, j \in \mathcal{I}_K$. Define $\mathbf{H}$ as the overall $M_\Sigma \times M_\Sigma$ channel matrix from all K transmitters to all K receivers (i.e., $([\mathbf{H}_{ji}])$), and let $\bar{\mathbf{H}}$ be obtained from $\mathbf{H}$ by replacing all desired channels (i.e., channels between corresponding transmitter-receiver pairs, $\mathbf{H}_{kk}$) with zeros. Our new insight for the unified setting comes from a novel outer bound argument that shows that the DoF cannot exceed half-the-cake if $\bar{\mathbf{H}}$ has full rank. In light of our outer bound, the counterexample mentioned above implies that the 24×24 matrix

$$\bar{\mathbf{H}} = \begin{array}{c} \\ {\scriptstyle 10} \\ {\scriptstyle 8} \\ {\scriptstyle 6} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} {\scriptstyle 10} & {\scriptstyle 8} & {\scriptstyle 6} \end{array} \\ \begin{pmatrix} \mathbf{0} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \mathbf{H}_{21} & \mathbf{0} & \mathbf{H}_{23} \\ \mathbf{H}_{31} & \mathbf{H}_{32} & \mathbf{0} \end{pmatrix} \end{array}, \text{ with ranks } \begin{array}{c} \\ {\scriptstyle 10} \\ {\scriptstyle 8} \\ {\scriptstyle 6} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} {\scriptstyle 10} & {\scriptstyle 8} & {\scriptstyle 6} \end{array} \\ \begin{pmatrix} 0 & 6 & 6 \\ 5 & 0 & 6 \\ 6 & 6 & 0 \end{pmatrix} \end{array}$$

cannot have full rank for any possible realization. Indeed, this is the case because the 24×18 sub-matrix formed by its first 18 columns is rank-deficient (sum of row ranks cannot be more than $6 + 5 + 6 = 17$).

Stated in an equivalent form, the new outer bound leads us to a more precise understanding of the original insight, so that we are able to refine it to the following form for generic rank-deficient channels.

Refined Insight: "*Half-the-cake is optimal if at every transmitter and receiver, the sum of* **reduced**[1] *interfering channel ranks* **equals** *the number of antennas at that transmitter and receiver, respectively.*"

So according to the refined condition, we are allowed to reduce the ranks of the cross-channels, but the reduced interference channel ranks must then add up at each transmitter and receiver to precisely equal the number of antennas at that transmitter and receiver, respectively. The counterexample presented earlier does not satisfy the refined condition. Indeed, it is not possible to assign any (possibly reduced) rank values that add up to the row and column index for every row and every column.

[1] Consider arbitrary channel matrix $\mathbf{H}_{ji}$ with rank D_{ji}. By 'reduce the rank' we mean 'choose a number $\bar{D}_{ji} \leq D_{ji}$' instead of D_{ji}. The $\bar{D}_{ji}$ value is called the reduced rank.

On the other hand, consider a different $\bar{\mathbf{H}}$ with ranks

$$\begin{array}{c} \\ {\scriptstyle 10} \\ {\scriptstyle 8} \\ {\scriptstyle 6} \end{array} \begin{array}{c} \begin{array}{ccc} {\scriptstyle 10} & {\scriptstyle 8} & {\scriptstyle 6} \end{array} \\ \begin{pmatrix} 0 & 8 & 3 \\ 5 & 0 & 4 \\ 6 & 2 & 0 \end{pmatrix} \end{array} \text{ which can be reduced to } \begin{array}{c} \\ {\scriptstyle 10} \\ {\scriptstyle 8} \\ {\scriptstyle 6} \end{array} \begin{array}{c} \begin{array}{ccc} {\scriptstyle 10} & {\scriptstyle 8} & {\scriptstyle 6} \end{array} \\ \begin{pmatrix} 0 & 8 & 2 \\ 4 & 0 & 4 \\ 6 & 0 & 0 \end{pmatrix} \end{array}$$

so that the reduced ranks add up to the row and column index for every row and column. Therefore, any realization of $\bar{\mathbf{H}}$ channels with these (unreduced or reduced) ranks cannot achieve more than half-the-cake. Also, as we show, for generic channels half-the-cake is always achievable, so it is optimal.

As a "sufficient" condition for optimality of half-the-cake, the additional requirements in the refined condition may appear to weaken its impact. This is not the case, however, as we note that the refined condition still recovers all prior results on the optimality of half-the-cake from [2, 29, 30] as special cases of the K-user $M_k \times M_k$ rank-deficient MIMO channel model. For $K = 3$, we also show that if the rank of each interference link is symmetric, i.e., $\text{rank}\,(\mathbf{H}_{ji}) = \text{rank}\,(\mathbf{H}_{ij})$, then the condition is also necessary for half-the-cake DoF to be optimal.

The broader technical contribution of this chapter is a novel class of replication-based DoF outer bounds that are applicable to the general K-user $M_k \times N_k$ MIMO interference channel with arbitrary rank-constraints, where all the nodes can have different number of antennas. The DoF of general MIMO interference channels are of fundamental interest as they shed light into the accessibility of signal dimensions with local joint processing (MIMO) at each node within the globally distributed setting that is an interference network. In particular, information theoretic DoF outer bounds for MIMO interference channels offer a powerful tool beyond the cut-set bounds used extensively in the study of wireless and wired communication networks. As such, DoF outer bounds have been studied in [25, 15, 40, 37, 30, 41, 31], mostly for symmetric settings, leading to various approaches based on cooperation [15], change of basis operations [40, 37, 30] and genie-chains [41, 31]. However, in spite of much progress, the DoF of MIMO interference networks remain unknown in general, even in symmetric settings, but especially under asymmetric settings. Evidently, there is a need for new outer bounding arguments to extend, complement, and where possible, simplify the existing approaches. It is in these regards that the new DoF outer bounds developed in this chapter are significant.

The key step in our replication-based outer bounding approach is to include auxiliary users as copies of existing users with corresponding independent auxiliary messages, ensure the connectivity is such that any achievable scheme for the original K-user network continues to work in the new network, creating a new network where simple bounds (such as Carleial's bound in [3] and cooperation based bounds) can be applied to produce various weighted sum-rate bounds for the original network. This approach provides us a class of outer bounds for general K-user $M_k \times N_k$ MIMO interference channel with any given channel realization. While the new bounds are conceptually quite simple and easily extendable to weighted sum-rates, a challenging aspect of these information theoretic bounds is that there could be many valid connectivity patterns that produce distinct outer bounds so that finding the best bound may be computationally cumbersome. However, this aspect can be greatly simplified if the bounds are restricted to linear DoF, i.e., DoF achieved by *linear* precoding schemes. Remarkably, unlike prior works on feasibility of linear schemes [43, 1, 36, 14] which do not allow symbol extensions or asymmetric signaling and focus on generic channels, the resulting linear DoF outer bounds from our work allow all possible linear schemes (including symbol extensions, asymmetric signaling) and apply to arbitrary interfering channels (not only generic ones).

In Section 2.2, the system model is introduced. The main results are formally stated in

Section 2.3. Section 2.4 shows that prior results on optimality of half-the-cake can be recovered as special cases of our generalized result (a counterexample to the original insight we mentioned above is presented in Section 2.6). In Section 2.5, examples of applications of the new bound are presented. Indexing here is interpreted in a circular wrap-around manner, modulo the number of users, e.g., the K^{th} user is same as the 0^{th} user.

2.2 System Model

2.2.1 K-user Rank-Constrained MIMO Interference Channel

The general setting of interest is the K-user MIMO interference channel where there are M_k and N_k antennas at the k^{th} transmitter and receiver, respectively. Each transmitter sends an independent message to its corresponding receiver. We refer to this general setting as the $(M_k \times N_k)$ interference channel. At time slot $t \in \mathbb{Z}^+$, the received signal vector at Receiver j is given by

$$Y_j(t) = \sum_{i=1}^{K} \mathbf{H}_{ji}(t) X_i(t) + Z_j(t) \tag{2.1}$$

where $X_i(t) \in \mathbb{C}^{M_i \times 1}$ is the signal vector sent from Transmitter i which satisfies an average power constraint $\mathbb{E}(\|X_i(t)\|^2) \leq \rho$, $Z_j(t) \in \mathbb{C}^{N_j \times 1}$ is the i.i.d. circularly symmetric complex additive white Gaussian noise (AWGN) at Receiver j, each entry of which is an i.i.d. Gaussian random variable with zero-mean and unit-variance, and $\mathbf{H}_{ji}(t) \in \mathbb{C}^{N_j \times M_i}$ is the channel matrix from Transmitter i to Receiver j. We assume that perfect global channel knowledge is available at all nodes.

The desired channel matrices $\mathbf{H}_{ii}(t)$ are assumed to be full rank[2] while the cross channels

[2]Similar to [29], the extension to rank-deficient desired channels is straightforward. All outer bounds in this paper continue to hold, regardless of the ranks of the desired channels. Achievability may be influenced

$\mathbf{H}_{ji}$ are subject to rank constraint D_{ji}. By default the channels are assumed to be *generic* — by which we mean the channels are ergodically time-varying and drawn from continuous distributions subject to rank-constraints. Similar to [29], a rank-constrained generic $N_j \times M_i$ channel matrix of rank D_{ji} is modeled as a product of an $N_j \times D_{ji}$ matrix with a $D_{ji} \times M_i$ matrix, all of whose entries are drawn from a continuous distribution so that the $N_j \times D_{ji}$ matrix and the $D_{ji} \times M_i$ matrix both have rank D_{ji} (full rank) almost surely.

We note that our DoF outer bounds, which are the primary focus of this work, also hold for *arbitrary* channels, i.e., without the assumptions of generic and ergodically time-varying channels. Achievability results are included to highlight the quality of the bounds, which are shown to be tight for generic channels. While we expect the results to hold true (almost surely) even without ergodicity or time-variations, choosing ergodic time-varying channels allows us to simplify the achievability arguments as much as possible, so that the focus of this work remains on the outer bounds.

The achievable rates, capacity region and DoF region of this network are defined in the standard sense (see [2]). We define the sum-DoF value as $d_\Sigma = \lim_{\rho\to\infty} R_\Sigma(\rho)/\log(\rho)$, where $R_\Sigma(\rho)$ is the maximum sum rate at Signal-to-noise ratio, ρ. We also define $N_\Sigma = \Sigma_{k\in\mathcal{I}_K} N_k$, $M_\Sigma = \Sigma_{k\in\mathcal{I}_K} M_k$.

2.3 Results

In this section we state the main results of this work.

because we need the ranks of the desired channels to be large enough to support the DoF values that we wish to achieve.

2.3.1 The Rank-Constrained K-user $(M_k \times M_k)$ Interference Channel – Optimality of Half-the-Cake

In this section, we focus on the $(M_k \times M_k)$ setting, i.e., where $N_k = M_k$, so that the desired channel matrices are generic full rank square matrices, while the interference channel matrices are in general rectangular and subject to arbitrary rank-constraints. This setting unifies and generalizes the cases studied in [29] and [30], and forms our starting point. We start with the achievability result, which is a simple application of the ideas of ergodic interference alignment [35] and blind interference alignment [24], which says in this case, that for generic channels, "half-the-cake" is almost surely achievable.

Theorem 2.1. *For generic channels, regardless of interference rank-constraints*

$$d_\Sigma \geq M_\Sigma/2$$

Proof. Since the channels are ergodically time-varying and drawn from continuous distributions, we may partition the channels over all time slots to pairs of 2 channel uses, such that for each 2 channel uses, say at times t_1 and t_2, all channel matrices of interference links remain the same $\mathbf{H}_{ji}(t_1) = \mathbf{H}_{ji}(t_2)$, $i \neq j$, and all channel matrices of direct links change $\mathbf{H}_{ii}(t_1) \neq \mathbf{H}_{ii}(t_2)$ in a generic sense, i.e., their difference is also full rank. Then each transmission takes such 2 channel uses, and by letting each transmitter repeat its symbols over the 2 channel uses, each receiver can eliminate interference by subtracting the output at t_2 from the output at t_1, and obtain an $M_k \times M_k$ interference free channel, over which M_k DoF are obtained. Since, this requires two channel uses, effectively $\frac{M_k}{2}$ DoF are achieved for User k and in total $M_\Sigma/2$ sum-DoF are achieved. ■

The main question of interest is, when is half-the-cake *optimal*? To answer this we introduce a new replication-based outer bound argument that will turn out to be quite broadly applicable. Recall that $\bar{\mathbf{H}}$ is the overall $M_\Sigma \times M_\Sigma$ channel matrix where all desired channels $\mathbf{H}_{kk}$ have been set to zero. Specialized to our present purpose, the outer bound is presented below.

Theorem 2.2. *For arbitrary channel realizations, if*

$$rank(\bar{\mathbf{H}}) = M_\Sigma \quad then \quad d_\Sigma \leq M_\Sigma/2.$$

Note that the outer bound applies to arbitrary channels, i.e., without any requirements for time-varying, ergodic, or generic realizations. Remarkably, the proof is quite simple, based upon a replication argument.

Proof. Given the original K-user interference channel with channel matrices $\mathbf{H}_{ji}$, now create a $2K$-user interference channel by adding an auxiliary User k' for each Original User k. We denote the channels in the new $2K$-user network by notations with hat symbol, e.g., $\hat{\mathbf{H}}_{ji'}$ represents the channel matrix from Transmitter i' to Receiver j. The new channels are chosen so that $\forall i, j \in \mathcal{I}_K$, 1) $\hat{\mathbf{H}}_{j'i} = \hat{\mathbf{H}}_{ji'} = \mathbf{H}_{ji}$ whenever $i \neq j$, 2) $\hat{\mathbf{H}}_{i'i'} = \hat{\mathbf{H}}_{ii} = \mathbf{H}_{ii}$, 3) $\hat{\mathbf{H}}_{j'i'} = \hat{\mathbf{H}}_{ji}$ is the matrix of zeros whenever $i \neq j$, and 4) $\hat{\mathbf{H}}_{i'i} = \hat{\mathbf{H}}_{ii'}$ is the matrix of zeros. For a pictorial illustration of the case where $K = 3$, see Figure 2.2 and 2.3.

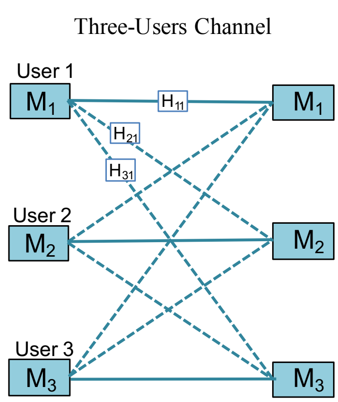

Figure 2.2: A 3-user interference channel.

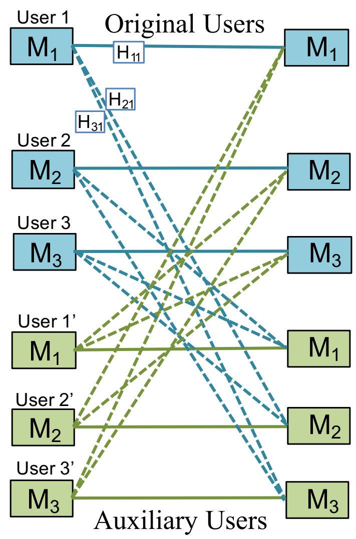

Figure 2.3: The 6-user interference channel created in Theorem 2.2.

Any coding scheme for the original channel still works if each auxiliary User i' uses the same codebook as User i. Since Users i and i' in the new network achieve the same rates as User i in the original network, the sum-DoF value for the new network is at least twice that of the original network. Now in the new network, allow all original transmitters to cooperate, all original receivers to cooperate, all auxiliary transmitters to cooperate and all auxiliary receivers to cooperate, which can only help. This creates a 2-user interference channel where everyone has M_Σ antennas, and where the interference matrix is $\bar{\mathbf{H}}$. If this interference matrix is full rank, then each user, after decoding its desired signal, can subtract it out and then proceed to decode the interfering signal as well (subject to noise distortion, inconsequential for DoF). Thus, the sum-DoF of the interference channel cannot be more than M_Σ, and therefore the sum-DoF of the original network cannot be more than $\frac{1}{2}M_\Sigma$. ■

For generic channels, the condition of Theorem 2.2 can be presented in a simpler alternative form, in terms of the ranks of the individual interfering channels, as follows.

Lemma 2.1. *For generic channel realizations, rank*$(\bar{\mathbf{H}}) = M_\Sigma$ *if and only if there exist* *reduced ranks* $\bar{D}_{ji} \leq D_{ji}$ *for each interference link, which satisfy the following condition,*

$$\sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ji} = \sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ij} = M_i, \forall i \in \mathcal{I}_K. \tag{2.2}$$

The proof of Lemma 2.1 is presented in Section 2.7.

Combined with Theorem 2.2, Lemma 2.1 directly proves the following theorem, which unifies and generalizes the results from [29] and [30].

Theorem 2.3. *For a* K*-user generic rank-deficient MIMO interference channel, if there exist reduced ranks* $\bar{D}_{ji} \leq D_{ji}$ *for each interference link, which satisfy the following condition,*

$$\sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ji} = \sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ij} = M_i, \forall i \in \mathcal{I}_K. \tag{2.3}$$

then almost surely half-the-cake is optimal, i.e., $d_\Sigma = \sum_{k=1}^{K} \frac{M_k}{2}$.

Corollary 2.1. *For 3 users, one can state Condition (2.3) more explicitly as follows.*

$$\begin{aligned} &\min\{M_1 + D_{32}, M_2 + D_{13}, M_3 + D_{21}\} + \\ &\min\{M_3 + D_{12}, M_1 + D_{23}, M_2 + D_{31}\} \geq M_1 + M_2 + M_3. \end{aligned} \tag{2.4}$$

The proof of Corollary 2.1 is presented in Section 2.8.

Theorem 2.3 presents a sufficient condition for the optimality of half-the-cake in generic settings. The condition is not a necessary condition for the optimality of half-the-cake. However, combined with the achievability result of Theorem 2.1, it recovers the corresponding results from [29] and [30]. Finding a condition that is both necessary and sufficient seems to be a difficult task in general, mainly due to the abundance of distinct parameter regimes. The following theorems offer interesting insights into this.

Theorem 2.4. *For a 3-user generic rank-deficient MIMO interference channel, if the rank of each interference link is symmetric, i.e.,* $D_{ji} = D_{ij}$*, then the condition in Theorem 2.3 is necessary and sufficient for half-the-cake to be optimal.*

The proof of Theorem 2.4 is presented in Section 2.9.

The following two theorems show that Condition (2.3) is not necessary for the optimality of half-the-cake. The proofs are presented in Section 2.10.

Theorem 2.5. *For a 3-user generic rank-deficient MIMO interference channel where* $M_1 = M_2 + M_3$*, half-the-cake is optimal, i.e., the sum-DoF value is* $\frac{1}{2}M_\Sigma$ *if the following condition is satisfied*

$$D_{12} = M_2, D_{13} = M_3 \quad or \quad D_{21} = M_2, D_{31} = M_3 \tag{2.5}$$

Remark: To see how the condition in Theorem 2.5 violates Condition (2.3), consider the example where $M_1 = 5$, $M_2 = 3$ and $M_3 = 2$, (i.e., $M_1 = M_2 + M_3$), $D_{21} = M_2 = 3$, $D_{31} = M_3 = 2$, and all other interference channel matrices are matrices of zeros. Note that this example satisfies the condition in Theorem 2.5. However, since $D_{12} = D_{13} = 0$, there are no reduced ranks $\bar{D}_{12}$ and $\bar{D}_{13}$ such that $\bar{D}_{12} + \bar{D}_{13} = M_1 = 5$, i.e., Condition (2.3) is violated.

Theorem 2.6. *For a 3-user generic rank-deficient MIMO interference channel where* $M_1 = M_2$*, half-the-cake is optimal, i.e., the sum-DoF value is* $\frac{1}{2}M_\Sigma$*, if the following condition is satisfied*

$$\begin{aligned} &D_{21} = M_1, D_{31} = D_{23} = M_3 \quad or \\ &\quad D_{12} = M_1, D_{13} = D_{32} = M_3 \end{aligned} \tag{2.6}$$

Remark: To see how the condition in Theorem 2.6 violates Condition (2.3), consider the example where $M_1 = M_2 = 5$ and $M_3 = 3$, $D_{21} = M_1 = 5$ and $D_{31} = D_{23} = M_3 = 3$, and all other interference channel matrices are matrices of zeros. Note that this example satisfies the condition in Theorem 2.6. However, since $D_{12} = D_{13} = 0$, there are no reduced ranks $\bar{D}_{12}$ and $\bar{D}_{13}$ such that $\bar{D}_{12} + \bar{D}_{13} = M_1 = 5$, i.e., Condition (2.3) is violated.

2.3.2 Replication-Based Bounds for General $(M_k \times N_k)$ Rank Constrained K-user Interference Channel

As discussed previously, the outer bound that we introduce in Theorem 2.2, is of particular interest in and of itself as it is based on a rather broadly applicable replication argument. The simplicity of this argument makes it easy to generalize the outer bounds. To emphasize this point, in this section we consider some generalizations of the outer bound to the $(M_k \times N_k)$ interference channel. For this, we first define a new $(\mu_1 + \mu_2 + \cdots + \mu_K)$-user "replicated" network as follows.

Definition 2.1. [Replicated Network] *For any given $(M_k \times N_k)$ interference channel described by channel matrices $\mathbf{H}_{ji}$, we create a new $(\mu_1 + \mu_2 + \cdots + \mu_K)$-user interference channel by replacing each User i with μ_i auxiliary users (replicas), and denoting them as User $i^{[1]}$, User $i^{[2]}$, $\cdots$, User $i^{[\mu_i]}$, each with its own independent message. In this replicated network, we denote the channel matrix from Transmitter $i^{[\alpha]}$ to Receiver $j^{[\beta]}$ as $\hat{\mathbf{H}}_{j^{[\beta]}i^{[\alpha]}}$, $\forall i, j \in \mathcal{I}_K, \alpha \in \mathcal{I}_{\mu_i}, \beta \in \mathcal{I}_{\mu_j}$. The channel matrices in the replicated network are chosen to satisfy the following constraints.*

1) $\forall i, \alpha, \hat{\mathbf{H}}_{i^{[\alpha]}i^{[\alpha]}} = \mathbf{H}_{ii}$ and $\forall \gamma \in \mathcal{I}_{\mu_i}, \gamma \neq \alpha, \hat{\mathbf{H}}_{i^{[\gamma]}i^{[\alpha]}}$ are matrices of zeros,

2) $\forall i \neq j$, $\forall \beta$, there exists an α such that $\hat{\mathbf{H}}_{j^{[\beta]}i^{[\alpha]}} = \mathbf{H}_{ji}$ and $\forall \gamma \in \mathcal{I}_{\mu_i}$, $\gamma \neq \alpha$, $\hat{\mathbf{H}}_{j^{[\beta]}i^{[\gamma]}}$ are matrices of zeros.

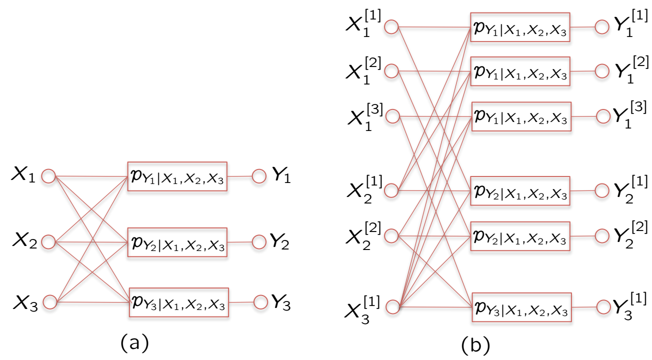

Figure 2.4: (a) A 3-user original IC. $p_{Y_i|X_1,X_2,X_3}$ denotes the conditional probability relating the output and input. (b) One possible replicated network when $(\mu_1, \mu_2, \mu_3) = (3, 2, 1)$.

In words, we require that in the replicated network, each desired link is the same as that of the original network, and each replicated receiver sees $K-1$ interferences, one from each interfering replicated transmitter. For a pictorial illustration of one replicated network for the case where $K = 3, (\mu_1, \mu_2, \mu_3) = (3, 2, 1)$, see Figure 2.4. Note that to highlight the generality of the replicated network, we draw the example in the discrete memoryless channel setting.

In the replicated network, each transmitter has the same power constraint as that of the original network and the Gaussian noise at each receiver has the same covariance matrix as that of the original network. Each transmitter has an independent message for its desired receiver. The replicated network is constructed so that its sum capacity is an outer bound to the weighted sum rate of the original network. We state this result in the following theorem.

Theorem 2.7. *For an integer weight vector $(\mu_1, \cdots, \mu_K)$, the weighted sum rate $\mu_1 R_1 + \cdots \mu_K R_k$ of the original network is bounded by the sum capacity $\hat{R}_\Sigma$ of the replicated network.*

Proof. We show that $\mu_1 R_1 + \cdots \mu_K R_k \leq \hat{R}_\Sigma$. It suffices to prove that if the rate tuple $(R_1, \cdots, R_K)$ is achievable over the original network, then the rate tuple

$$(\underbrace{R_1, \cdots, R_1}_{\mu_1 \text{ times}}, \underbrace{R_2, \cdots, R_2}_{\mu_2 \text{ times}}, \cdots, R_K)$$

is achievable over the replicated network. This is proved by using the encoding/decoding mappings of the original network in the replicated network. Suppose we are given a sequence of encoding and decoding mappings such that $(R_1, \cdots, R_K)$ is achievable over the original network. Then each Replicated Transmitter $i^{[\alpha]}$ encodes its desired message with the same encoding function as used by Transmitter i in the original network. As a result, from our construction of the replicated network, the received signal at each Replicated Receiver $j^{[\beta]}$ is statistically the same as the received signal at Receiver j in the original network, such that the same decoding mapping can be used to achieve the same rate R_j. Therefore the proof is complete. ■

Remark: It is not hard to see that the replicated network argument not only applies to Gaussian channels, but also to discrete memoryless channels. For example, the replicated network argument is used in the context of 2 and 3 user symmetric deterministic interference channels in an independent work [28] (see Lemma 1 in [28]). In this work we focus only on Gaussian channels and leave the extension to discrete memoryless channels as future work.

The above theorem is proved in terms of capacity, such that corresponding result on DoF is directly implied. Next we focus on sum-DoF (i.e., choose μ_k to be the same for all k) of the original $(M_k \times N_k)$ interference network, which can be bounded in terms of the sum-DoF of the replicated network.

Although the sum-DoF outer bound problem has been reduced to the sum-DoF outer bound problem of the replicated network, the latter is not available immediately. To obtain an explicit thus easily applicable bound on the sum-DoF of the replicated network, we turn to a simple cooperation based argument. Somewhat surprisingly, a simple cooperation argument for the replicated network can provide tighter bound than possible through the same simple cooperation argument for the original network. In this work, we only apply the simple cooperation argument to bound the replicated network and leave more sophisticated methods and full potential of using the replicated network as future work.

We use the cooperation argument in the following way. Assume that we replicate each user μ times. For the resulting $K\mu$-user interference channel, we divide the users into two groups and allow full cooperation between the transmitters/receivers in each group. Thus, we have a 2-user interference channel. For such a 2-user channel, we denote the number of antennas at Transmitter 1 and Receiver 1 as $\bar{M}_1$ and $\bar{N}_1$, respectively. Similarly, we denote the number of antennas at Transmitter 2 and Receiver 2 as $\bar{M}_2$ and $\bar{N}_2$, respectively. The $\bar{N}_2 \times \bar{M}_1$ channel matrix from Transmitter 1 to Receiver 2 is represented as $\bar{\mathbf{H}}^{\text{coop}}$. We are now ready to state the outer bound for the $(M_k \times N_k)$ interference channel, in the following theorem.

Theorem 2.8. *For arbitrary realizations of the rank-constrained K-user $(M_k \times N_k)$ MIMO interference channel, the sum-DoF value is outer bounded as follows.*

$$d_\Sigma \leq \frac{1}{\mu}\left[\bar{M}_1 + \bar{N}_2 - \text{rank}\left(\bar{\mathbf{H}}^{coop}\right)\right], \quad \forall \mu \in \mathbb{Z}^+. \tag{2.7}$$

where $\bar{\mathbf{H}}^{coop}$ is the interference channel in the replicated network after cooperation, as defined above.

Remark: For the same μ, there may be multiple possible replicated networks. For each possible replicated network, we also have multiple choices of forming groups (cooperation). $\bar{\mathbf{H}}^{\text{coop}}$ is defined according to one specific grouping of one specific replicated network. In this regard, Theorem 2.2 is a special case of Theorem 2.8 and it corresponds to the case where $\mu = 2$, the interference links in the replicated network are all between the two replicas of the original network, and cooperation is allowed within each replica of the original network.

Proof. By Theorem 2.7, the sum-DoF value $d_\Sigma = \frac{1}{\mu}(\mu d_1 + \cdots + \mu d_K)$ of the original network is bounded by $\frac{1}{\mu}$ of the sum-DoF of the replicated network, which is in turn bounded by $\frac{1}{\mu}$ of the sum-DoF of the 2-user interference channel after cooperation. Then the proof by Theorem 1 in [29] can be applied here. Specifically, we add $\bar{M}_1 - \text{rank}\left(\bar{\mathbf{H}}^{\text{coop}}\right)$ antennas at Receiver 2. This will not reduce the DoF. The channel coefficients corresponding to the new antennas are generic, so that the interference channel between Transmitter 1 and Receiver 2, now a matrix of size $[\bar{M}_1 + \bar{N}_2 - \text{rank}(\bar{\mathbf{H}}^{\text{coop}})] \times M_1$, will have full rank almost surely. Then Receiver 2, after decoding its desired signal, can subtract it out and then proceed to decode the interfering signal as well. Thus, the sum-DoF of this replicated network cannot be more than $\bar{M}_1 + \bar{N}_2 - \text{rank}(\bar{\mathbf{H}}^{\text{coop}})$ and the proof follows. ■

As there are multiple choices of replicated networks, it could be computationally cumbersome to find the one that would produce the tightest outer bound. Remarkably, if we relax our target from information theoretic DoF outer bounds to linear DoF outer bounds (i.e., the highest DoF achievable through linear precoding schemes), then a simpler alternative presents itself.

To present the result, we will need the following definition.

Definition 2.2. *Suppose we have an $(M_k \times N_k)$ interference channel with channel matrices $\mathbf{H}_{ji}$. Similar to Definition 2.1, we replicate User i μ_i times. We use notations with tilde symbol in this created network. The channels in the new network are designed as follows.*

1) $\forall i, \alpha, \tilde{\mathbf{H}}_{i^{[\alpha]} i^{[\alpha]}} = \mathbf{H}_{ii}$, *and* $\forall \gamma \in \mathcal{I}_{\mu_i}, \gamma \neq \alpha, \tilde{\mathbf{H}}_{i^{[\gamma]} i^{[\alpha]}}$ *are matrices of zeros,*

2) $\forall i \neq j$, $\forall \beta, \alpha$, $\tilde{\mathbf{H}}_{j^{[\beta]} i^{[\alpha]}} = a_{j^{[\beta]} i^{[\alpha]}} \mathbf{H}_{ji}$, *where* $a_{j^{[\beta]} i^{[\alpha]}}$ *is independently and uniformly drawn from the interval* $[0, 1]$.

In words, each replicated receiver here is connected to all interfering replicated transmitters, instead of seeing only one interference from each replicated transmitter, as in the replicated network. As such, in this new network, each receiver is connected to more transmitters than that of the original network, such that the decoding mapping used by the original network does not apply to the new network. Therefore, the new network does not serve as outer bound to the original network information theoretically, but we show that the outer bounding argument still holds in linear sense. We state the result in the following theorem.

Theorem 2.9. *For an integer weight vector $(\mu_1, \cdots, \mu_K)$, if the DoF tuple $(d_1, \cdots, d_K)$ is linearly achievable over the original network, then the DoF tuple*

$$(\underbrace{d_1, \cdots, d_1}_{\mu_1 \text{ times}}, \underbrace{d_2, \cdots, d_2}_{\mu_2 \text{ times}}, \cdots, d_K)$$

is also achievable linearly over the created network defined in Definition 2.2.

Proof. As the DoF tuple $(d_1, \cdots, d_K)$ is linearly achievable over the original network, we have integers m_k, n_k such that $d_k = \frac{m_k}{n}$ and User k is able to send m_k symbols with n channel uses through linear beamforming schemes. This means that there exist K beamforming matrices $\mathbf{V}_k \in \mathbb{C}^{M_k n \times m_k}$ used by each transmitter, respectively, and K filtering matrices $\mathbf{U}_k \in \mathbb{C}^{m_k \times N_k n}$ used by each receiver, respectively, such that

$$\operatorname{rank}(\mathbf{U}_k \mathbf{H}_{kk}^{\text{ex}} \mathbf{V}_k) = m_k, \quad \forall k \in \mathcal{I}_K \tag{2.8}$$

$$\mathbf{U}_j \mathbf{H}_{ji}^{\text{ex}} \mathbf{V}_i = \mathbf{0}, \quad \forall i, j \in \mathcal{I}_K, j \neq i \tag{2.9}$$

where $\mathbf{H}_{ji}^{\text{ex}}$ denotes the block diagonal channel matrix with n blocks and each block is $\mathbf{H}_{ji}$. We now proceed to show that User $k^{[\gamma]}, \gamma \in \mathcal{I}_{\mu_k}$ in the created network can also send m_k symbols over n_k channel uses, such that d_k DoF are achievable. For such a purpose, Transmitter $k^{[\gamma]}$ precodes its desired symbols through the beamforming matrix $\mathbf{V}_k$ and Receiver $k^{[\gamma]}$ decodes its desired symbols with the filtering matrix $\mathbf{U}_k$. As the desired channel matrices and interference channel matrices (although the number has increased) are the same as that of the original network, from (2.8) (2.9), all desired symbols can be decoded successfully. This completes the proof. ■

Remark: For a given weight vector, while the replicated network for information theoretic DoF bounds is not unique (i.e., there are multiple possible connectivities for the replicated network constructed in Definition 1 because each replicated receiver is connected to *one* arbitrary replicated transmitter among all replicated transmitters corresponding to the same original transmitter), the created network for linear DoF bounds is unique (i.e., there is only one possible connectivity for the network constructed in Definition 2 because each replicated receiver is connected to *all* interfering replicated transmitters). This makes it much easier to explore the linear DoF outer bound. To find an explicit DoF bound on the created network, which serves as outer bound to the linear DoF of the original network, we may also resort to simple cooperation arguments.

2.4 Recovering Prior Results as Special Cases

In this section we will show that the prior results in [29, 30], on the optimality of half-the-cake, can be recovered as special cases of Theorem 2.3.

2.4.1 Full rank case

In [30], half-the-cake DoF is shown to be optimal in a K-user $M_k \times M_k$ MIMO interference channel where there is no dominant user and all channels have full rank. To prove that full rank K-user $M_k \times M_k$ MIMO interference channels satisfy the condition in Theorem 2.3, it is sufficient to show that for any $M_1 \leq M_2 + \cdots + M_K$, we can always find a set of values for $\bar{D}_{ij} \leq \min(M_i, M_j)$ that satisfy the condition in Lemma 2.1.

To start, suppose $\forall k \in \mathcal{I}_K$, each Transmitter k has M_k chips and each Receiver k has an empty bin that can hold M_k chips. Transmitter 1 starts by dropping as many chips as possible into Receiver 2's bin, and then if the bin is full and he still has chips left over, he continues with Receiver 3's bin, and so on. After Transmitter 1 is done, Transmitter 2 does the same, starting with Receiver 3's bin. Transmitter 2 is followed by Transmitters

$3, 4, \cdots, K$, in that order. At the end, the number of chips in receiver bin i from Transmitter j is chosen to be the rank $\bar{D}_{ij}$. Since there is no dominant user, the total capacity of all bins is the same as the total number of chips, and users are arranged as $M_1 \geq M_2 \geq \cdots \geq M_K$, it is easy to see that this allocation works.

2.4.2 Symmetric case

In [29], it is shown that for a K-user rank deficient MIMO interference channel with M antennas at each node, if all the direct channels have full rank, and all cross channels have rank D, then half-the-cake DoF is optimal when $(K-1)D \geq M$. We now show that this result is also a special case of Theorem 2.3.

Note that if $\frac{M}{K-1}$ is an integer, then we just need to reduce D to the value $\frac{M}{K-1}$. When $\frac{M}{K-1}$ is not an integer, we can write $M = \left\lfloor \frac{M}{K-1} \right\rfloor (K-1) + \Delta$ for some positive integer $\Delta < K-1$. Now, assign reduced interference ranks as follows.

$$\begin{aligned}
\bar{D}_{ji} &= \left\lfloor \tfrac{M}{K-1} \right\rfloor + 1 \leq D, \text{ if } j \in \{i+1, i+2, \cdots, i+\Delta\}, \\
\bar{D}_{ji} &= \left\lfloor \tfrac{M}{K-1} \right\rfloor \leq D, \text{ otherwise.}
\end{aligned}$$

With these reduced ranks, the condition in Lemma 2.1 is always satisfied. Thus, Theorem 2.3 applies and half-the-cake DoF is optimal.

2.5 Examples of Applications of New Outer Bounds

As an example of the broader applicability of the new DoF outer bounds, we next recover a known DoF result in $(M \times N)$ setting with our new bound. After that, we will apply the new bound to the generalized $(M_k \times N_k)$ interference channel.

2.5.1 Example 1: $(M \times N)$ Interference Channel

We consider a 3-user (2×3) generic full rank MIMO interference channel. It is shown in [40] that the sum-DoF value of this channel is $\frac{18}{5}$. We will show that the sum-DoF outer bound can be obtained in a simple manner by using a replicated network and cooperation based bound.

The replicated network is described as follows. We set $\mu_1 = \mu_2 = \mu_3 = \mu = 5$, i.e., we replicate each user $i \in \mathcal{I}_3$ 5 times. The desired channels in the replicated network are the same as that in the original network. The interference channels are chosen as, $\forall \alpha, \beta \in \mathcal{I}_5, \hat{\mathbf{H}}_{i^{[\beta]}(i+1)^{[\alpha]}}$ equals $\mathbf{H}_{i(i+1)}$ if $\alpha = \beta - 3$, and $\mathbf{0}_{3\times2}$ otherwise, $\hat{\mathbf{H}}_{i^{[\beta]}(i+2)^{[\alpha]}}$ equals $\mathbf{H}_{i(i+2)}$ if $\alpha = \beta - 2$, and $\mathbf{0}_{3\times2}$ otherwise. It can be verified that the replicated network satisfies Definition 2.1.

Next we allow the first three replicas of the original network (i.e., Users $1^{[l]}, 2^{[l]}, 3^{[l]}, l = 1, 2, 3$) to cooperate, and the remaining users to cooperate. This creates a 2-user interference channel where Transmitter 1 has $\bar{M}_1 = 18$ antennas, Receiver 1 has $\bar{N}_1 = 27$ antennas, Transmitter 2 has $\bar{M}_2 = 12$ antennas and Receiver 2 has $\bar{N}_2 = 18$ antennas. $\bar{\mathbf{H}}^{\text{coop}}$ is the 18×18 interference matrix from Transmitter 1 to Receiver 2. By Theorem 2.8, we have $d_\Sigma \leq \frac{1}{\mu}\left[\bar{M}_1 + \bar{N}_2 - \text{rank}\left(\bar{\mathbf{H}}^{\text{coop}}\right)\right] = \frac{1}{5}[18 + 18 - \text{rank}\left(\bar{\mathbf{H}}^{\text{coop}}\right)]$. In order to prove $\frac{18}{5}$ is a valid outer bound, we are left to prove that $\text{rank}\left(\bar{\mathbf{H}}^{\text{coop}}\right) = 18$, i.e., $\bar{\mathbf{H}}^{\text{coop}}$ has full rank almost surely.

To show this, it suffices to prove that the determinant polynomial of $\bar{\mathbf{H}}^{\text{coop}}$ is not identically zero, which can be proved by constructing a specific channel such that this is true. One such channel may be

$$\mathbf{H}_{i(i+1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{H}_{i(i+2)} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.10}$$

It is readily verifiable that for such a channel, the determinant of $\bar{\mathbf{H}}^{\text{coop}}$ is non-zero. Therefore, $\bar{\mathbf{H}}^{\text{coop}}$ has full rank and the proof is complete.

2.5.2 Example 2: $(M_k \times N_k)$ Interference Channel

We now consider a 3-user $(10\times10)(8\times10)(6\times3)$ MIMO interference channel. It is assumed that $\mathbf{H}_{31}$ is the matrix of zeros, i.e., $D_{31} = 0$, and all other interference matrices are generic full rank. This channel setting has not been considered in the literature and its sum-DoF value is not known. We show that the sum-DoF value is 12, with the help of the insights from our general outer bound (Theorem 2.8).

We start with the outer bound. The replicated network is described as follows. We set $\mu_1 = \mu_2 = \mu = 2$, i.e., we replicate each user 2 times. The channels in the replicated network are chosen so that $\forall i, j \in \mathcal{I}_3$, 1) $\hat{\mathbf{H}}_{j^{[1]}i^{[2]}} = \hat{\mathbf{H}}_{j^{[2]}i^{[1]}} = \mathbf{H}_{ji}$ whenever $i \neq j$, 2) $\hat{\mathbf{H}}_{i^{[1]}i^{[1]}} = \hat{\mathbf{H}}_{i^{[2]}i^{[2]}} = \mathbf{H}_{ii}$, 3) $\hat{\mathbf{H}}_{j^{[1]}i^{[1]}} = \hat{\mathbf{H}}_{j^{[2]}i^{[2]}}$ is the matrix of zeros whenever $i \neq j$, and 4) $\hat{\mathbf{H}}_{i^{[1]}i^{[2]}} = \hat{\mathbf{H}}_{i^{[2]}i^{[1]}}$ is the matrix of zeros. It can be verified that this replicated network satisfies Definition 2.1.

Next we allow users $1^{[1]}$, $2^{[1]}$ and $3^{[1]}$ to cooperate, and users $1^{[2]}$, $2^{[2]}$ and $3^{[2]}$ to cooperate. This creates a 2-user interference channel where Transmitter 1 has $\bar{M}_1 = 24$ antennas, Receiver 1 has $\bar{N}_1 = 23$ antennas, Transmitter 2 has $\bar{M}_2 = 24$ antennas and Receiver 2 has $\bar{N}_2 = 23$ antennas. $\bar{\mathbf{H}}^{\text{coop}}$ is the 23×24 interference matrix from Transmitter 1 to Receiver 2. If $\bar{\mathbf{H}}^{\text{coop}}$ has full rank, then by Theorem 2.8, we have the desired result,

$d_\Sigma \leq \frac{1}{\mu}\left[\bar{M}_1 + \bar{N}_2 - \text{rank}\left(\bar{\mathbf{H}}^{\text{coop}}\right)\right] = \frac{1}{2}\left[24 + 23 - 23\right] = 12.$

To show that $\bar{\mathbf{H}}^{\text{coop}}$ has full rank almost surely, it suffices to prove that in $\bar{\mathbf{H}}^{\text{coop}}$, there exists a 23×23 sub-matrix where its determinant polynomial is not identically zero. This can be proved by constructing a specific channel such that this is true. One such channel may be

$$\begin{aligned}
\mathbf{H}_{21} =& \mathbf{I}_{N_2}, \quad \mathbf{H}_{32} = \begin{bmatrix} \mathbf{I}_{N_3} & \mathbf{0}_{N_3 \times (M_2 - N_3)} \end{bmatrix}, \\
\mathbf{H}_{13} =& \mathbf{H}_{23} = \begin{bmatrix} \mathbf{I}_{M_3} \\ \mathbf{0}_{(N_1 - M_3) \times M_3} \end{bmatrix}, \\
\mathbf{H}_{12} =& \begin{bmatrix} \mathbf{0}_{(N_1 - M_2) \times M_2} \\ \mathbf{I}_{M_2} \end{bmatrix}.
\end{aligned}$$

It is readily verifiable that for such a channel, the determinant of the sub-matrix consist by the first 23 columns of $\bar{\mathbf{H}}^{\text{coop}}$ is non-zero. Therefore, $\bar{\mathbf{H}}^{\text{coop}}$ has full rank and the outer bound proof is complete.

We next proceed to the achievability. We show that the DoF tuple $(d_1, d_2, d_3) = (7, 3, 2)$ can be achieved, such that the sum-DoF bound of 12 is tight. We use $\mathbf{v}_{i1}$, $\mathbf{v}_{i2}$, ..., $\mathbf{v}_{id_i}$ to denote the beamforming vectors at Transmitter i. We first choose $\mathbf{v}_{21}$ and $\mathbf{v}_{31}$ so that

$$\mathbf{H}_{32}\mathbf{v}_{21} = \mathbf{0}, \mathbf{H}_{12}\mathbf{v}_{21} = \mathbf{H}_{13}\mathbf{v}_{31} \quad \Leftrightarrow$$

$$\underbrace{\begin{bmatrix} \mathbf{H}_{32} & \mathbf{0} \\ \mathbf{H}_{12} & -\mathbf{H}_{13} \end{bmatrix}_{13 \times 14}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{v}_{21} \\ \mathbf{v}_{31} \end{bmatrix}}_{\mathbf{v}} = \mathbf{0}. \tag{2.11}$$

Note that $\mathbf{v}$ can be chosen from the right null space of $\mathbf{A}$.

Next we choose $\mathbf{v}_{22}$ so that

$$\mathbf{H}_{32}\mathbf{v}_{22} = 0, \text{ and } \mathbf{v}_{22} \text{ is linearly independent of } \mathbf{v}_{21}, \tag{2.12}$$

and $\mathbf{v}_{23}, \mathbf{v}_{32}$ so that

$$\begin{aligned} &\mathbf{H}_{12}\mathbf{v}_{23} = \mathbf{H}_{13}\mathbf{v}_{32}, \text{ where } \mathbf{v}_{23} \text{ is linearly independent of} \\ &\mathbf{v}_{21}, \mathbf{v}_{22}, \text{ and } \mathbf{v}_{32} \text{ is independent of } \mathbf{v}_{31}. \end{aligned} \tag{2.13}$$

The existence of $\mathbf{v}_{22}$ is guaranteed as $\mathbf{H}_{32}$ is a 3×8 generic matrix, whose right null space has 5 dimensions. The existence of $\mathbf{v}_{23}, \mathbf{v}_{32}$ is guaranteed as $\mathbf{H}_{12}$ has dimension 10×8 and $\mathbf{H}_{13}$ has dimension 10×6, such that the two overlap in a 4 dimensional subspace.

Then we choose $\mathbf{v}_{11}, \mathbf{v}_{12}$ so that

$$\begin{aligned} &\mathbf{H}_{21}\begin{bmatrix}\mathbf{v}_{11} & \mathbf{v}_{12}\end{bmatrix} = \mathbf{H}_{23}\begin{bmatrix}\mathbf{v}_{31} & \mathbf{v}_{32}\end{bmatrix} \Rightarrow \\ &\begin{bmatrix}\mathbf{v}_{11} & \mathbf{v}_{12}\end{bmatrix} = \mathbf{H}_{21}^{-1}\mathbf{H}_{23}\begin{bmatrix}\mathbf{v}_{31} & \mathbf{v}_{32}\end{bmatrix} \end{aligned} \tag{2.14}$$

At the last step, $\mathbf{v}_{13}$, ..., $\mathbf{v}_{17}$ are chosen as generic vectors. Thus, we have allocated all the beamforming vectors.

We are left to show that at each receiver, the interferences are aligned to a subspace that is independent of the desired signal space. First, we consider Receiver 3. Note that $\mathbf{H}_{31} = 0$. From (2.11) (2.12), the interference space is $\mathbf{H}_{32}[\mathbf{v}_{21}, \mathbf{v}_{22}, \mathbf{v}_{23}] = \mathbf{H}_{32}\mathbf{v}_{23}$, which has dimension $1 = N_3 - d_3$. Next, we consider Receiver 2. From (2.14), the interference from Transmitter 3 lies in the span of the interference from Transmitter 1, so that the total interference occupies $d_1 = 7$ dimensions, leaving $10 - 7 = 3 = d_2$ dimensions for the desired signal, as desired.

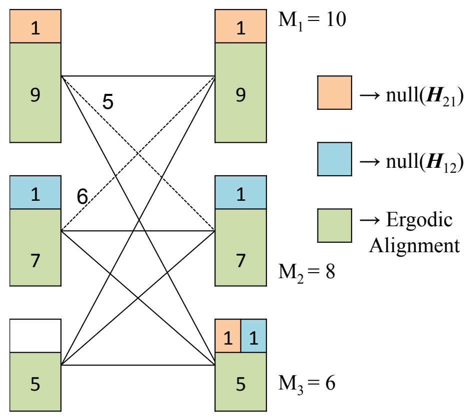

Figure 2.5: Example for achieving more than half-the-cake DoF.

We now consider Receiver 1. From (2.12) (2.13), the interference from Transmitter 3 lies in the span of the interference from Transmitter 2, so that the total interference has $d_3 = 3$ dimensions. The desired signal is left with $10 - 3 = 7 = d_1$ dimensions. Finally, as desired channels do not appear when we design the beamforming vectors, the independence of the aligned interference and desired signal is guaranteed. This completes the proof.

2.6 Counterexample to Original Insight

Here we briefly summarize how more than half-the-cake DoF can be achieved in the 3-user setting shown in Figure 2.5, where $D_{12} = 6$, $D_{21} = 5$ and all other links have full rank.

The transmission takes place over 2 channel uses, where all cross channels remain the same, and all direct channels change to different generic values [24]. We use $\mathbf{v}_1^z$ and $\mathbf{v}_2^z$ to denote the beamforming vectors at Transmitters 1 and 2 that need to be aligned at Receiver 3 after being chosen from the null space they see at each other. The symbols carried by $\mathbf{v}_1^z$ and $\mathbf{v}_2^z$

are different over two channel uses. Mathematically, we have

$$\begin{aligned} \mathbf{H}_{21}\mathbf{v}_1^z &= \mathbf{0}, \\ \mathbf{H}_{12}\mathbf{v}_2^z &= \mathbf{0}, \\ \mathbf{H}_{31}\mathbf{v}_1^z &= \mathbf{H}_{32}\mathbf{v}_2^z. \end{aligned} \quad \Rightarrow \quad \underbrace{\begin{bmatrix} \mathbf{H}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{12} \\ \mathbf{H}_{31} & -\mathbf{H}_{32} \end{bmatrix}_{24\times 18}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{v}_1^z \\ \mathbf{v}_2^z \end{bmatrix}}_{\mathbf{v}} = \mathbf{0}.$$

Note that matrix $\mathbf{A}$ has rank 17, thus $\mathbf{v}$ can be chosen from the right null space of $\mathbf{A}$. In the same manner, we choose the receive combining vectors $\mathbf{u}_1^z$ and $\mathbf{u}_2^z$ at Receivers 1 and 2 satisfying the following equations

$$\mathbf{u}_1^z\mathbf{H}_{12} = \mathbf{0}, \quad \mathbf{u}_2^z\mathbf{H}_{21} = \mathbf{0}, \quad \mathbf{u}_1^z\mathbf{H}_{13} = \mathbf{u}_2^z\mathbf{H}_{23}.$$

Next, we use $\mathbf{V}_k^e$ and $\mathbf{U}_k^e$ to denote the $M_k \times (M_k - 1)$ and $(M_k - 1) \times M_k$ matrices at each transmitter and receiver, respectively. These matrices carry the signals for ergodic alignment (green area in Figure 2.5), i.e., signals repeated over the two channel uses. User 3 needs to choose its beamforming/combining matrices to satisfy $\mathbf{V}_3^e = \text{span}(\text{null}(\mathbf{u}_2^z\mathbf{H}_{23}))$ and $\mathbf{U}_3^e = \text{span}(\text{null}(\mathbf{H}_{32}\mathbf{v}_2^z))$. As a result, each receiver can eliminate interference by only subtracting the part of received signals corresponding to $\mathbf{U}_k^e$ of two time slots. Thus, a total of 25 DoF are achieved over the two channel uses, or equivalently, 12.5 DoF per channel use (half-the-cake is 12 DoF per channel use).

2.7 Proof of Lemma 2.1

We prove Lemma 2.1 by first showing that Condition (2.2) is sufficient for $\bar{\mathbf{H}}$ to have full rank, and then showing that this condition is also necessary.

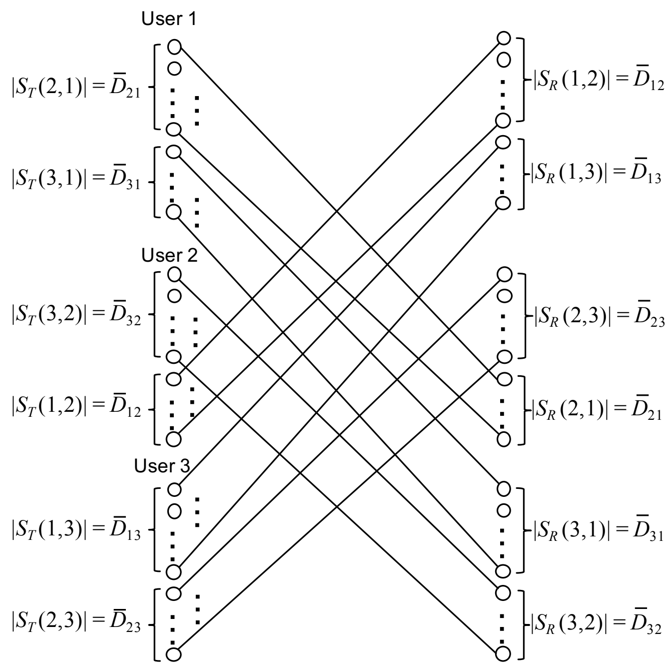

Figure 2.6: Illustration of the sufficiency proof for Lemma 2.1 when $K = 3$.

The determinant polynomial of the matrix $\bar{\mathbf{H}}$ is the polynomial expression obtained as the determinant of $\bar{\mathbf{H}}$ when the elements of $\bar{\mathbf{H}}$ are viewed as variables. For example the determinant polynomial of the 2×2 matrix $(x_1, x_2; x_3, x_4)$ is $x_1x_4 - x_2x_3$.

2.7.1 Sufficiency

To prove that $\bar{\mathbf{H}}$ is full-rank almost surely for generic rank-deficient channels with given ranks, it suffices to show that its determinant polynomial is not identically zero. To show this, it suffices to find one realization of $\bar{\mathbf{H}}$ for which the determinant is not zero. Such a realization is constructed as follows. At Receiver i, starting from the first antenna, label the first set of $\bar{D}_{i,i+1}$ antennas as $S_R(i, i+1)$, the next $\bar{D}_{i,i+2}$ as $S_R(i, i+2)$, and so on, until the final set of $\bar{D}_{i,i+K-1}$ antennas is labeled as $S_R(i, i+K-1)$. Similarly, at Transmitter j, starting from the first antenna, label the first set of $\bar{D}_{j+1,j}$ antennas as $S_T(j+1, j)$, the next set of $\bar{D}_{j+2,j}$ antennas as the set $S_T(j+2, j)$, and so on until the last set of $\bar{D}_{j+K-1,j}$

antennas is labeled as $S_T(j+K-1,j)$. Now connect transmit antennas in $S_T(i,j)$ with the receive antennas in $S_R(i,j)$ through identity matrices. For a pictorial illustration of such channel realization for the case where $K=3$, see Figure 2.6. With this channel realization, each transmit antenna is connected to exactly one undesired receive antenna, so that $\bar{\mathbf{H}}$ has exactly one 1 in each row and each column, and is therefore full rank. Increasing any of the ranks only introduces additional variables into the polynomial which can be set to zero to return to the same realization described above, thus proving that the polynomial is not identically zero.

2.7.2 Necessity

If the following partitioned matrix $\bar{\mathbf{H}}$ has full rank,

$$\bar{\mathbf{H}} = \begin{array}{c} \\ {\scriptstyle M_1} \\ {\scriptstyle M_2} \\ \vdots \\ {\scriptstyle M_K} \end{array} \begin{array}{c} \begin{array}{ccccc} {\scriptstyle M_1} & {\scriptstyle M_2} & {\scriptstyle \cdots} & {\scriptstyle M_{K-1}} & {\scriptstyle M_K} \end{array} \\ \begin{pmatrix} \mathbf{0} & \mathbf{H}_{12} & \cdots & \mathbf{H}_{1(K-1)} & \mathbf{H}_{1K} \\ \mathbf{H}_{21} & \mathbf{0} & \cdots & \mathbf{H}_{2(K-1)} & \mathbf{H}_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{H}_{K1} & \mathbf{H}_{K2} & \cdots & \mathbf{H}_{K(K-1)} & \mathbf{0} \end{pmatrix} \end{array} \tag{2.15}$$

then the first observation is that the following sub-matrices of $\bar{\mathbf{H}}$,

$$\begin{bmatrix} \mathbf{H}_{1i}^T & \cdots & \mathbf{H}_{(i-1)i}^T & \mathbf{0} & \mathbf{H}_{(i+1)i}^T & \cdots & \mathbf{H}_{Ki}^T \end{bmatrix}^T, \ \forall i \in \mathcal{I}_K$$

$$\begin{bmatrix} \mathbf{H}_{i1} & \cdots & \mathbf{H}_{i(i-1)} & \mathbf{0} & \mathbf{H}_{i(i+1)} & \cdots & \mathbf{H}_{iK} \end{bmatrix}, \ \forall i \in \mathcal{I}_K$$

must have full rank, i.e., the ranks of the sub-matrices must satisfy the following conditions.

$$\sum_{j\in\mathcal{I}_K\backslash i} D_{ji} \geq M_i, \ \sum_{j\in\mathcal{I}_K\backslash i} D_{ij} \geq M_i, \ \forall i \in \mathcal{I}_K. \tag{2.16}$$

With the help of this observation, the necessity of Condition (2.2) can be proved as follows. Any $\mathbf{H}_{ji}$ of rank D_{ji} can be represented as a sum of D_{ji} matrices, each of which has rank 1, i.e.,

$$\mathbf{H}_{ji} = a_{ji}^{[1]}\mathbf{v}_{ji}^{[1]}\mathbf{u}_{ji}^{[1]} + a_{ji}^{[2]}\mathbf{v}_{ji}^{[2]}\mathbf{u}_{ji}^{[2]} + \cdots + a_{ji}^{[D_{ji}]}\mathbf{v}_{ji}^{[D_{ji}]}\mathbf{u}_{ji}^{[D_{ji}]} \tag{2.17}$$

where $\mathbf{v}_{ji}^{[m]}$ and $\mathbf{u}_{ji}^{[m]}$ are $M_j \times 1$ and $1 \times M_i$ unit vectors, respectively. Now let us consider the $a_{ji}^{[m]}$ as variables while $\mathbf{v}_{ji}^{[m]}$ and $\mathbf{u}_{ji}^{[m]}$ are treated as constants. Each entry of $\mathbf{H}_{ji}$ can be represented as a linear combination of $a_{ji}^{[m]}$, i.e., $L(a_{ji}^{[1]}, a_{ji}^{[2]}, \cdots, a_{ji}^{[D_{ji}]})$. The determinant of $\bar{\mathbf{H}}_{ji}$ is a polynomial in the $a_{ji}^{[m]}$ variables, i.e.,

$$\mathcal{A} \triangleq \{a_{ji}^{[m]} : j \in \mathcal{I}_K, i \in \mathcal{I}_K, m \in \mathcal{I}_{D_{ji}}]\} \tag{2.18}$$

$$\det(\bar{\mathbf{H}}) \triangleq p(\mathcal{A}) \tag{2.19}$$

Note that since $\bar{\mathbf{H}}$ has full rank, $p(\mathcal{A})$ cannot be the zero polynomial, i.e., there exists some realization of the variables in $\mathcal{A}$ for which $p(\mathcal{A}) \neq 0$.

Next, we go through the following procedure. Initialize $\mathcal{A}' = \{\}$.

Step 1 Choose any $a_{ji}^{[m]} \in \mathcal{A}$. Add this variable to $\mathcal{A}'$, i.e., $\mathcal{A}' \longleftarrow \mathcal{A}' \cup \{a_{ji}^{[m]}\}$.

Step 2 If substituting $a_{ji}^{[m]} = 0$ makes $p(\mathcal{A})$ the zero polynomial, then do nothing. Otherwise, i.e., if substituting $a_{ji}^{[m]} = 0$ does not make $p(\mathcal{A})$ the zero polynomial, then fix $a_{ji}^{[m]} = 0$

as a constant, and remove $a_{ji}^{[m]}$ from $\mathcal{A}$, i.e., $\mathcal{A} \longleftarrow \mathcal{A}/\{a_{ji}^{[m]}\}$. $p(\mathcal{A})$ now denotes the polynomial in the remaining $\mathcal{A}$ variables.

Step 3 If there remain $a_{ji}^{[m]}$ terms that have not yet been chosen, i.e., if $\mathcal{A} \not\subset \mathcal{A}'$, then go back to Step 1. If $\mathcal{A} \subset \mathcal{A}'$, i.e., all $a_{ji}^{[m]}$ have been tested, then exit.

At this stage, the number of remaining $a_{ji}^{[m]}$ variables for each sub-matrix defines the reduced rank value $\bar{D}_{ji}$ for that matrix.

$$\bar{D}_{ji} \quad = \quad |\{a_{kl}^{[m]} : a_{kl}^{[m]} \in \mathcal{A}, k = j, l = i, m \in \mathcal{I}_{D_{ji}}\}| \tag{2.20}$$

Note the following two facts.

Fact 1 Each remaining $a_{ji}^{[m]} \in \mathcal{A}$ is a factor of the polynomial $p(\mathcal{A})$.

Fact 2 The total degree of the determinant polynomial $p(\mathcal{A})$ is less than or equal to M_Σ.

Fact 1 is true because for any remaining $a_{ji}^{[m]} \in \mathcal{A}$, setting $a_{ji}^{[m]} = 0$ makes $p(\mathcal{A})$ identically 0. Fact 2 is true because the maximum degree of any term in the $\bar{\mathbf{H}}$ matrix is 1, and $\bar{\mathbf{H}}$ is an $M_\Sigma \times M_\Sigma$ matrix. Since $p(\mathcal{A})$ cannot have more factors than its total degree, it follows that the number of remaining variables, $|\mathcal{A}| \leq M_\Sigma$, i.e., $\sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ji} \leq M_i$, $\sum_{j \in \mathcal{I}_K \setminus i} \bar{D}_{ij} \leq M_i$. Since all the $\bar{D}_{ji}$ must also satisfy Condition (2.16) in order for $\bar{\mathbf{H}}$ to have full rank, all the inequalities in (2.16) must take equality. In other words, for any full rank matrix $\bar{\mathbf{H}}$, there always exist reduced ranks $\bar{D}_{ji} \leq D_{ji}$ which satisfy Condition (2.2). This completes the proof.

2.8 Proof of Corollary 2.1

We want to prove that the following two polytopes are equivalent.

The polytope (denoted as $\bar{\mathcal{D}}^*$) given by Condition (2.3) (when $K = 3$) is the set of tuples $(D_{12}, D_{21}, D_{23}, D_{32}, D_{31}, D_{13}) \in \mathbb{Z}_+^6$ such that there exist $\bar{D}_{ji} \leq D_{ji}$, which satisfy the following constraints.

$$\bar{D}_{12} + \bar{D}_{13} = M_1 \tag{2.21}$$

$$\bar{D}_{21} + \bar{D}_{23} = M_2 \tag{2.22}$$

$$\bar{D}_{31} + \bar{D}_{32} = M_3 \tag{2.23}$$

$$\bar{D}_{21} + \bar{D}_{31} = M_1 \tag{2.24}$$

$$\bar{D}_{12} + \bar{D}_{32} = M_2 \tag{2.25}$$

$$\bar{D}_{13} + \bar{D}_{23} = M_3 \tag{2.26}$$

.

The polytope (denoted as $\mathcal{D}^*$) given by Condition (2.4) is the set of tuples $(D_{12}, D_{21}, D_{23}, D_{32}, D_{31}, D_{13}) \in \mathbb{Z}_+^6$ defined by the following constraints.

$$D_{12} + D_{13} \geq M_1 \tag{2.27}$$

$$D_{21} + D_{23} \geq M_2 \tag{2.28}$$

$$D_{31} + D_{32} \geq M_3 \tag{2.29}$$

$$D_{21} + D_{31} \geq M_1 \tag{2.30}$$

$$D_{12} + D_{32} \geq M_2 \tag{2.31}$$

$$D_{13} + D_{23} \geq M_3 \tag{2.32}$$

$$D_{12} + D_{21} \geq M_1 + M_2 - M_3 \tag{2.33}$$

$$D_{23} + D_{32} \geq M_2 + M_3 - M_1 \tag{2.34}$$

$$D_{13} + D_{31} \geq M_1 + M_3 - M_2 \tag{2.35}$$

The above 9 linear inequalities are obtained by expanding each term in the min expression of (2.4) and rearranging.

Next we prove $\mathcal{D}^* = \bar{\mathcal{D}}^*$ by proving that $\mathcal{D}^* \subseteq \bar{\mathcal{D}}^*$ and $\bar{\mathcal{D}}^* \subseteq \mathcal{D}^*$.

$\mathcal{D}^* \subseteq \bar{\mathcal{D}}^*$: We need to show that if D_{ji} satisfy (2.27) to (2.35), then we can find $\bar{D}_{ji} \leq D_{ji}$ that satisfy (2.21) to (2.26). Without loss of generality, we assume

$$\min(M_3 + D_{12}, M_1 + D_{23}, M_2 + D_{31}) = M_1 + D_{23} \tag{2.36}$$

We set $\bar{D}_{ji}$ as follows.

$$\bar{D}_{12}=D_{12}-(D_{12}+M_3-M_1-D_{23})=M_1+D_{23}-M_3 \tag{2.37}$$

$$\bar{D}_{13}=D_{13}-(D_{13}+D_{23}-M_3)=M_3-D_{23} \tag{2.38}$$

$$\bar{D}_{21}=D_{21}-(D_{21}+D_{23}-M_2)=M_2-D_{23} \tag{2.39}$$

$$\bar{D}_{23}=D_{23} \tag{2.40}$$

$$\bar{D}_{31}=D_{31}-(D_{31}+M_2-M_1-D_{23})=M_1+D_{23}-M_2 \tag{2.41}$$

$$\bar{D}_{32}=D_{32}-(D_{32}+D_{23}+M_1-M_2-M_3)=M_2+M_3-M_1-D_{23} \tag{2.42}$$

It is easy to verify that (2.21) to (2.26) are satisfied by above assignment. We are left to prove that each difference term is valid, i.e., $0 \leq D_{ji} - \bar{D}_{ji} \leq D_{ji}$. This proof is a simple manipulation of the inequalities (2.27) to (2.36) and the rank property $0 \leq D_{ji} \leq$

$\min(M_i, M_j)$, thus we omit it. Therefore this direction is proved.

$\bar{\mathcal{D}}^* \subseteq \mathcal{D}^*$: We need to show that if there exist $\bar{D}_{ji} \leq D_{ji}$ that satisfy (2.21) to (2.26), then D_{ji} must satisfy (2.27) to (2.35). To see this, note that we have

$$(2.21)+(2.22)-(2.26) \Rightarrow \bar{D}_{12}+\bar{D}_{21} = M_1+M_2-M_3 \tag{2.43}$$

$$(2.22)+(2.23)-(2.24) \Rightarrow \bar{D}_{23}+\bar{D}_{32} = M_2+M_3-M_1 \tag{2.44}$$

$$(2.21)+(2.23)-(2.25) \Rightarrow \bar{D}_{13}+\bar{D}_{31} = M_1+M_3-M_2 \tag{2.45}$$

Combining with (2.21) to (2.26), we have the exact same form of the inequalities in (2.27) to (2.35). As $\bar{D}_{ji} \leq D_{ji}$, (2.21) to (2.26) and (2.43) to (2.45) imply (2.27) to (2.35). This direction is proved.

2.9 Proof of Theorem 2.4

We want to prove that for a 3-user interference channel, if the rank of each interference link is symmetric, i.e., $D_{ji} = D_{ij}$, then Condition (2.3) (equivalently Condition (2.4)) is necessary for half-the-cake optimality. To prove this, it suffices to prove that when Condition (2.4) (inequalities (2.27) to (2.35)) does not hold, we can always achieve more than half-the-cake DoF. We consider two cases, one when (2.33) - (2.35) is violated and the other when (2.27) to (2.32) is violated. We start with the first case.

2.9.1 More than Half-the-cake when Inequalities (2.33) - (2.35) are violated

As inequalities (2.33) - (2.35) are symmetric, without loss of generality, we assume (2.33) is violated, i.e.,

$$D_{12} + D_{21} < M_1 + M_2 - M_3 \tag{2.46}$$

Note that there is no assumption on (2.27) - (2.32), (2.34) and (2.35), they can be either violated or not. We will show that $\frac{M_1+M_2+M_3+1}{2}$ DoF can be achieved, by generalizing the scheme of the counterexample in Section 2.6.

The high level idea is the following. There exists a beamforming vector at Transmitter 1 and 2, respectively, that can align at Receiver 3 after being chosen from the null space they see at each other, as $M_1 - D_{21} + M_2 - D_{12} > M_3$ (refer to (2.46)). So these two symbols occupy only 3 dimensions in total at all receivers (see Figure 2.7 for an illustration). For the remaining $M_1 + M_2 + M_3 - 3$ dimensions, we apply ergodic alignment to achieve the DoF tuple $(\frac{M_1-1}{2}, \frac{M_2-1}{2}, \frac{M_3-1}{2})$ (green area in Figure 2.7). Added with the DoF tuple $(1, 1, 0)$ achieved as mentioned before, DoF tuple $(\frac{M_1+1}{2}, \frac{M_2+1}{2}, \frac{M_3-1}{2})$ is achieved in total. Thus, the sum-DoF value is more than half-the-cake.

Next we describe how to choose the beamforming vectors. Specifically, we operate over 2 channel uses, where all cross channels remain the same, and all direct channels are generically different. We use $\mathbf{v}_1^z$ and $\mathbf{v}_2^z$ to denote the beamforming vectors of the signal at Transmitter 1 and Transmitter 2 that need to be aligned after zero-forcing. These signals are different over two channel uses. Mathematically, we have

$$\mathbf{H}_{21}\mathbf{v}_1^z = \mathbf{0}, \mathbf{H}_{12}\mathbf{v}_2^z = \mathbf{0}, \mathbf{H}_{31}\mathbf{v}_1^z = \mathbf{H}_{32}\mathbf{v}_2^z \tag{2.47}$$

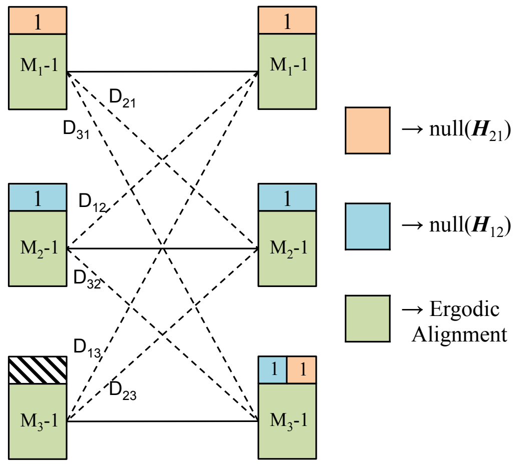

Figure 2.7: Illustration of the scheme that achieves more than half-the-cake when $D_{12}+D_{21} < M_1 + M_2 - M_3$.

$$\Rightarrow \underbrace{\begin{bmatrix} \mathbf{H}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{12} \\ \mathbf{H}_{31} & -\mathbf{H}_{32} \end{bmatrix}_{(M_1+M_2+M_3)\times(M_1+M_2)}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{v}_1^z \\ \mathbf{v}_2^z \end{bmatrix}}_{\mathbf{v}} = \mathbf{0}. \tag{2.48}$$

Note that matrix $\mathbf{A}$ is rank-deficient (sum of row ranks cannot be more than $D_{12}+D_{21}+M_3 < M_1 + M_2$, refer to (2.46)), thus $\mathbf{v}$ can be determined as one basis vector of the right null space of $\mathbf{A}$. In the same manner, we can choose the received beamforming vectors $\mathbf{u}_1^z$ and $\mathbf{u}_2^z$ at Receiver 1 and Receiver 2 satisfying the following equations

$$\mathbf{u}_1^z\mathbf{H}_{12} = \mathbf{0}, \mathbf{u}_2^z\mathbf{H}_{21} = \mathbf{0}, \mathbf{u}_1^z\mathbf{H}_{13} = \mathbf{u}_2^z\mathbf{H}_{23}. \tag{2.49}$$

Next, we use $\mathbf{V}_k^e$ and $\mathbf{U}_k^e$ to denote the $M_k \times (M_k - 1)$ and $(M_k - 1) \times M_k$ beamforming and filtering matrices at each transmitter and receiver, respectively. These matrices carry the signals for ergodic alignment, i.e., signals repeated by each user over two channel uses. User 1 and User 2 can choose $\mathbf{V}_1^e, \mathbf{V}_2^e$ and $\mathbf{U}_1^e, \mathbf{U}_2^e$ generically. User 3 chooses its beamforming matrix as follows

$$\mathbf{V}_3^e = \text{span}(\text{null}(\mathbf{u}_2^z \mathbf{H}_{23})), \mathbf{U}_3^e = \text{span}(\text{null}(\mathbf{H}_{32}\mathbf{v}_2^z)). \tag{2.50}$$

As a result, each receiver can eliminate interference by only subtracting the part of received signals corresponding to $\mathbf{U}_k^e$ over two channel uses. Thus, a total of $M_1 + M_2 + M_3 + 1$ DoF are achieved over two channel uses, which is more than half-the-cake DoF. The proof is complete. Remarkably, note that this proof does not require the assumption of symmetry, $D_{ij} \neq D_{ji}$, so it works for asymmetric settings as well.

2.9.2 More than Half-the-cake when Inequalities (2.27) - (2.32) are violated

We now consider the case where (2.27) - (2.32) are violated. Without loss of generality, we assume (2.27) is violated, i.e.,

$$D_{12} + D_{13} < M_1. \tag{2.51}$$

Since the ranks of the interference channels are symmetric, we have $D_{12} = D_{21}$ and $D_{13} = D_{31}$. Thus

$$D_{21} + D_{31} < M_1, \tag{2.52}$$

i.e., (2.30) is violated as well. Note that there is no assumption on (2.28), (2.29), (2.31) - (2.35), they can be either violated or not.

We will show that $\frac{M_1+M_2+M_3+1}{2}$ DoF can be achieved, by combining zero-forcing and ergodic alignment.

This case turns out to be quite simple. The high level idea is the following. There exists a beamforming vector at Transmitter 1 that cannot be seen by both Receivers 2 and 3. The symbol carried by this vector occupies only 1 dimension in total at all receivers. For the remaining $M_1+M_2+M_3-1$ dimensions, we apply ergodic alignment to achieve the DoF tuple $(\frac{M_1-1}{2}, \frac{M_2}{2}, \frac{M_3}{2})$. Added with the DoF tuple $(1, 0, 0)$ achieved as mentioned above, DoF tuple $(\frac{M_1+1}{2}, \frac{M_2}{2}, \frac{M_3}{2})$ is achieved in total. Thus, the sum-DoF value is more than half-the-cake.

Next we proceed to describe the scheme. Specifically, we operate over 2 channel uses, where all cross channels remain the same, and all direct channels are generically different. We use $\mathbf{v}_1^z$ to denote the beamforming vector of the signal at Transmitter 1 that is zero-forced at Receivers 2 and 3. This signal is different over two channel uses. Mathematically, we have

$$\begin{bmatrix}\mathbf{H}_{21} & \mathbf{H}_{31}\end{bmatrix} \mathbf{v}_1^z = 0. \tag{2.53}$$

Note that matrix $\begin{bmatrix}\mathbf{H}_{21} & \mathbf{H}_{31}\end{bmatrix}$ is rank-deficient (the rank cannot be more than $D_{21} + D_{31} < M_1$, refer to (2.52)), thus $\mathbf{v}_1^z$ can be determined as one basis vector of the right null space of $\begin{bmatrix}\mathbf{H}_{21} & \mathbf{H}_{31}\end{bmatrix}$. In the same manner, we can choose the received beamforming vectors $\mathbf{u}_1^z$ at Receiver 1 such that $\mathbf{u}_1^z \begin{bmatrix}\mathbf{H}_{12} & \mathbf{H}_{13}\end{bmatrix} = 0$.

Next, we use $\mathbf{V}_1^e$ and $\mathbf{U}_1^e$ to denote the $M_k \times (M_k - 1)$ and $(M_k - 1) \times M_k$ beamforming and filtering matrices at Transmitter 1 and Receiver 1, respectively. For $k \in \{1, 2\}$, we use $\mathbf{V}_k^e$ and $\mathbf{U}_k^e$ to denote the $M_k \times M_k$ beamforming and filtering matrices at Transmitter k and Receiver k, respectively. These matrices carry the signals for ergodic alignment, i.e., signals repeated by each user over two channel uses. Each user can choose its beamforming and filtering matrices generically.

As a result, each receiver can eliminate interference by only subtracting the part of received signals corresponding to $\mathbf{U}_k^e$ over two channel uses. Thus, a total of $M_1 + M_2 + M_3 + 1$ DoF are achieved over two channel uses, which is more than half-the-cake DoF. The proof is complete.

2.10 Non-necessity of Condition (2.3) at Boundary Cases

2.10.1 Proof of Theorem 2.5: $M_1 = M_2 + M_3$

Consider a 3-user interference channel where $M_1 = M_2 + M_3$, We want to show that if $D_{12} = M_2, D_{13} = M_3$ or $D_{21} = M_2, D_{31} = M_3$, then $d_\Sigma = \frac{1}{2}M_\Sigma$. Achievability is implied by Theorem 2.1, so we proceed to the outer bound. Since we are considering the outer bound, cooperation between the users will not hurt. Therefore, we allow Transmitter 2 and Transmitter 3 to cooperate and they form a new Transmitter 2′. Similarly, we allow Receiver 2 and Receiver 3 to cooperate and they form a new Receiver 2′. We now arrive at a 2-user interference channel, where Transmitter/Receiver 1 has M_1 antennas and Transmitter/Receiver 2′ has $M_{2'} = M_2 + M_3$ antennas. The desired channels have full rank, the interference channel from Transmitter 1 to Receiver 2′ has rank $D_{2'1} = D_{21} + D_{31}$, and the interference channel from Transmitter 2′ to Receiver 1 has rank $D_{12'} = D_{12} + D_{13}$. For such a rank-deficient 2-user MIMO interference channel, we invoke Theorem 1 in [29] to obtain the following outer bound which also serves as outer bound for the original 3-user interference channel, $d_\Sigma \leq M_1 + M_{2'} - \max(D_{2'1}, D_{12'}) = M_1 + M_2 + M_3 - \max(D_{21} + D_{31}, D_{12} + D_{13})$. Therefore, if $D_{12} = M_2, D_{13} = M_3$ or $D_{21} = M_2, D_{31} = M_3$, the outer bound becomes $d_\Sigma \leq M_1 = \frac{1}{2}M_\Sigma$. This completes the proof.

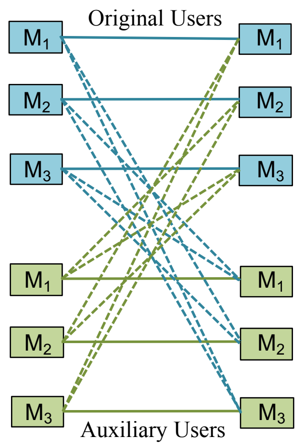

Figure 2.8: A 6-user IC created by adding an auxiliary user for each user in the original 3-user channel.

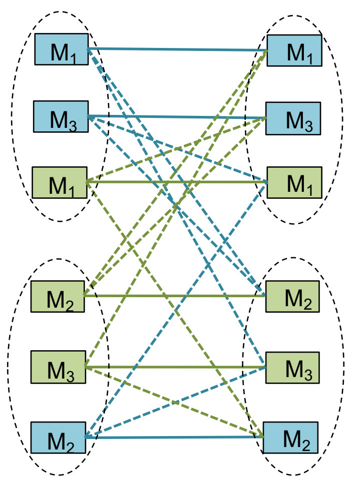

Figure 2.9: Illustration of users' cooperation in this new channel.

2.10.2 Proof of Theorem 2.6: $M_1 = M_2$

Consider a 3-user interference channel where $M_1 = M_2$. We want to show that if $D_{21} = M_1, D_{31} = D_{23} = M_3$ or $D_{12} = M_1, D_{13} = D_{32} = M_3$, then $d_\Sigma = \frac{1}{2}M_\Sigma$. Achievability is implied by Theorem 2.1, so we proceed to the outer bound. For such a purpose, we create a 6-user interference channel by adding an auxiliary User k' for each Original User k. We denote the channels in the new network by notations with hat symbol, e.g., $\hat{\mathbf{H}}_{ji'}$, and the channels in the original network by notations with no hat symbol, e.g., $\mathbf{H}_{ji}$. The channels in the new network are chosen in the same manner as in Theorem 2.2, i.e., $\forall i, j \in \{1, 2, 3\}$,

1) $\hat{\mathbf{H}}_{j'i} = \hat{\mathbf{H}}_{ji'} = \mathbf{H}_{ji}$ whenever $i \neq j$, 2) $\hat{\mathbf{H}}_{i'i'} = \hat{\mathbf{H}}_{ii} = \mathbf{H}_{ii}$, 3) $\hat{\mathbf{H}}_{j'i'} = \hat{\mathbf{H}}_{ji}$ is the matrix of zeros whenever $i \neq j$, and 4) $\hat{\mathbf{H}}_{i'i} = \hat{\mathbf{H}}_{ii'}$ is the matrix of zeros. See Figure 2.8 and 2.9 for a pictorial illustration. By this construction, any coding scheme for the original channel still works if each auxiliary User i' uses the same codebook as User i. Therefore the sum-DoF value of this new network is at least twice that of the original network. Now in this new network, we allow User 1, User 3 and User 1′ to cooperate, and User 2′, User 3′ and User 2 to cooperate, which can only help. This creates a 2-user interference channel where the first transmitter/receiver has $2M_1 + M_3$ antennas, the second transmitter/receiver has $2M_2 + M_3$ antennas. We denote the interference channel between the first/second transmitter and the second/first receiver as $\bar{\mathbf{H}}_{21}$ and $\bar{\mathbf{H}}_{12}$, respectively. Note that as $M_1 = M_2$, both $\bar{\mathbf{H}}_{21}$ and $\bar{\mathbf{H}}_{12}$ are square matrices. They may be written as

$$\bar{\mathbf{H}}_{21} = \begin{array}{c} \scriptstyle M_2 \\ \scriptstyle M_3 \\ \scriptstyle M_2 \end{array} \overset{\begin{array}{ccc} \scriptstyle M_1 & \scriptstyle M_3 & \scriptstyle M_1 \end{array}}{\begin{pmatrix} \mathbf{H}_{21} & \mathbf{H}_{23} & \mathbf{0} \\ \mathbf{H}_{31} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{21} \end{pmatrix}} \tag{2.54}$$

$$\bar{\mathbf{H}}_{12} = \begin{array}{c} \scriptstyle M_1 \\ \scriptstyle M_3 \\ \scriptstyle M_1 \end{array} \overset{\begin{array}{ccc} \scriptstyle M_2 & \scriptstyle M_3 & \scriptstyle M_2 \end{array}}{\begin{pmatrix} \mathbf{H}_{12} & \mathbf{H}_{13} & \mathbf{0} \\ \mathbf{H}_{32} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{12} \end{pmatrix}} \tag{2.55}$$

If $\bar{\mathbf{H}}_{21}$ has full rank, then the first receiver, after decoding its desired signal, can subtract it out and then proceed to decode the interfering signal as well (subject to noise distortion, inconsequential for DoF). Thus, the sum-DoF of the interference channel cannot be more than $2M_1 + M_3 = M_\Sigma$, and therefore the sum-DoF of the original network cannot be more than $\frac{1}{2}M_\Sigma$. Similarly, if $\bar{\mathbf{H}}_{12}$ has full rank, then the second receiver can decode both messages such that $d_\Sigma \leq \frac{1}{2}(2M_2+M_3) = \frac{1}{2}M_\Sigma$. We are left to prove that if $D_{21} = M_1, D_{31} = D_{23} = M_3$, then $\bar{\mathbf{H}}_{21}$ has full rank and symmetrically, if $D_{12} = M_1, D_{13} = D_{32} = M_3$, then $\bar{\mathbf{H}}_{12}$ has full rank. We prove the first statement and the second follows similarly. We prove that when $D_{21} = M_1, D_{31} = D_{23} = M_3$, the determinant polynomial of $\bar{\mathbf{H}}_{21}$ is not identically zero. It suffices to find one channel realization such that the determinant polynomial is not zero. The channels we construct are as follows.

$$\mathbf{H}_{21} = \mathbf{I}_{M_1},$$
$$\mathbf{H}_{31} = \begin{bmatrix} \mathbf{I}_{M_3} & \mathbf{0}_{M_3\times(M_1-M_3)} \end{bmatrix},$$
$$\mathbf{H}_{23} = \begin{bmatrix} \mathbf{I}_{M_3} \\ \mathbf{0}_{(M_2-M_3)\times M_3} \end{bmatrix}.$$

Note that the rank constraints are satisfied and it is easily seen that the determinant of $\bar{\mathbf{H}}_{21}$ is non-zero. Therefore, $\bar{\mathbf{H}}_{21}$ has full rank almost surely. We now finish the proof of the outer bound. Note that the procedure is a specific realization of Theorem 2.8. Combined with the achievability, the proof is complete.

2.11 Summary

The motivation for this chapter was to explore the sharper insights, especially into information theoretic DoF outer bounds, that might emerge from the study of rank-deficient MIMO interference channels under a model that unifies and generalizes prior works. For a K-user MIMO interference channel with arbitrarily rank-deficient cross-channels, where there are M_k antennas at the k^{th} user pair, it was shown that the sum-DoF cannot exceed half-the-cake if the overall $M_\Sigma \times M_\Sigma$ channel matrix $\bar{\mathbf{H}}$ where all desired channels have been set to zero, has full rank. This was accomplished through a new outer bound based on the idea of creating a replicated-network, i.e., creating copies (replicas) of certain users and choosing the connectivity of the replicated network in such a way that any achievable scheme in the original network translates into an achievable scheme for the replicated network. Depending on the number of replicas created for each user, the sum rate of the replicated network bounds the corresponding weighted sum of rates from the original network. What is remarkable about the replicated network is that it creates a new perspective of the problem, so that even simple arguments such as user cooperation become quite powerful when applied in the replicated network, giving rise to stronger outer bounds, than when applied directly in the original network. The replication argument is applicable not only to arbitrary MIMO interference channels with arbitrary rank-constraints, but much more broadly, even beyond Gaussian interference channels. The conceptual simplicity and apparent breadth of replication based bounds calls for future work into understanding their full potential, especially for MIMO interference channels where the DoF remain open in general.

Chapter 3

Elevated Multiplexing and Signal Space Partitioning in the 2 User MIMO IC with Partial CSIT

Degrees of freedom studies of wireless networks have contributed many fundamental insights into their capacity limits [22]. One of the most critical determinants of these capacity limits is the amount of CSIT. In this chapter, we will study the MIMO IC with partial CSIT. The channel model for partial CSIT is formulated under the GDoF metric. Therefore, the GDoF metric will be first introduced though an interesting problem, i.e., IC with shared message set.

3.1 GDoF of IC with Shared Message Set and No CSIT

The interference channel with shared message set can be treated as a combination of IC and BC. Specifically, there are two transmitter-receiver pairs. Rx1 want to decode message $W_1 = \{W_{1c}, W_{1p}\}$, Rx2 want to decode message $W_2 = \{W_{2c}, W_{2p}\}$. W_{ic} and W_{ip} are all independent sub-messages with the difference that W_{ic} is shared among two transmitters while W_{ip} is only know by Txi.

One may notice that the capacity of this channel is outer bounded by the capacity of BC. As shown in [6], the DoF of this channel collapse to 1 with no CSIT. However, this DoF metric suffers from one limitation, i.e., it essentially treats all non-zero channels as equally strong (i.e., each link is capable of carrying exactly one DoF) in the high SNR limit, and thus totally ignores the strength distinctions of various signals, which are critical for interference management in practice. Therefore, in order to explore the channel settings with both weak and strong interference and offers insights into optimal schemes for those channels. We need to expend the coarse DoF metric to a more general metric – GDoF.

Under the GDoF framework, the channel model for the 2-user IC at time slot t is defined by the following input-output equations.

$$Y_1(t) = \sqrt{P}G_{11}(t)X_1(t) + \sqrt{P^{\alpha_{12}}}G_{12}(t)X_2(t) + Z_1(t), \tag{3.1}$$

$$Y_2(t) = \sqrt{P^{\alpha_{21}}}G_{21}(t)X_1(t) + \sqrt{P}G_{22}(t)X_2(t) + Z_2(t), \tag{3.2}$$

Here, $X_k(t)$ is the symbol sent from Transmitter k, $k \in \{1, 2\}$, which is subject to a unit power constraint. $Y_k(t)$ is the received signal at Receiver k. $Z_k(t)$ is additive white Gaussian noise (AWGN) at Receiver k with zero-mean and unit-variance. $G_{kl}(t)$ are the channel fading coefficients between Transmitter l and Receiver k and are not known to the Transmitters. The channel strengths are represented in α_{kl} parameters. Without loss of generality, it is assumed $\alpha_{12} \geq \alpha_{21}$.

The definitions of achievable rates $R_i(P)$ and capacity region $\mathcal{C}(P)$ are standard. The GDoF region is defined as

$$\begin{aligned}\mathcal{D} =& \{(d_1, d_2) : \exists(R_1(P), R_2(P)) \in \mathcal{C}(P),\\ &\text{s.t. } d_k = \lim_{P\to\infty} \frac{R_k(P)}{\frac{1}{2}\log(P)}, \forall k \in \{1, 2\}\}\end{aligned} \tag{3.3}$$

Define d_{ic} and d_{ip} as the DoF for the message W_{ic} and W_{ip}, respectivelly. Also define $\gamma = \frac{d_{1c}+d_{2c}}{d_{1p}+d_{2p}}$, then the sum GDoF of this channel can be presented as a function of γ, $d_\Sigma(\gamma) = d_{1c} + d_{1p} + d_{2c} + d_{2p} = (1+\gamma)(d_{1p} + d_{2p})$. If $\gamma = 0$, this channel becomes IC since no messages are shared. If $\gamma = +\infty$, this channel becomes BC since all messages are shared between the two transmitter.

Now let us use d_{IC} and d_{BC} to denote the sum GDoF of IC and BC with the same channel strengths, then the sum GDoF of IC with shared message set can be outer bounded by the following theorem.

Theorem 3.1. *For the weak and mixed interference cases, i.e.,* $\alpha_{21} \leq 1$, $d_\Sigma(\gamma) \leq \min[d_{BC}, (1+\gamma)d_{IC}]$.

For the strong interference cases, i.e., $\alpha_{21} > 1$, $d_\Sigma(\gamma) \leq \min[d_{BC}, (1+\gamma)d_{IC}, \frac{1+\gamma}{2+\gamma}(\alpha_{12}+\alpha_{21})]$.

Proof

The first two bounds in Theorem 3.1 are trivial. $d_\Sigma(\gamma) \leq d_{BC}$ is because GDoF for BC is always an outer bound here. $d_\Sigma(\gamma) \leq (1+\gamma)d_{IC}$ is because $d_\Sigma(\gamma) \leq (1+\gamma)(d_{1p} + d_{2p})$ and the amount of private messages, i.e., W_{1p} and W_{2p}, send in this channel cannot exceed that can be send in IC. In other words, $d_{1p} + d_{2p} \leq d_{IC}$. We now prove that for $\alpha_{21} > 1$, $d_\Sigma(\gamma) \leq \frac{1+\gamma}{2+\gamma}(\alpha_{12} + \alpha_{21})$.

We use the AIS approach here, i.e., instead of no CSIT now consider the channel coefficients $G_{ij}(t)$ are available to Txs up to finite precision. Then we obtain the following deterministic channel model.

$$\bar{Y}_1(t) = \lfloor \sqrt{P^{1-\alpha_{21}}} G_{11}(t)\bar{X}_1(t) \rfloor + \lfloor G_{12}(t)\bar{X}_2(t) \rfloor, \tag{3.4}$$

$$\bar{Y}_2(t) = \lfloor G_{21}(t)\bar{X}_1(t) \rfloor + \lfloor \sqrt{P^{1-\alpha_{12}}} G_{22}(t)\bar{X}_2(t) \rfloor, \tag{3.5}$$

Here the definition of $\bar{X}_i(t)$ and the finite precision channel coefficients $G_{ij}(t)$ is the same as

[9]. Now for Rx1, we have

$$n(R_1-\epsilon) \leq I(W_{1c}, W_{1p}; \bar{Y}_1^n) \tag{3.6}$$

$$\leq I(W_{1c}, W_{1p}; \bar{Y}_1^n | W_{2c}) \tag{3.7}$$

$$\leq H(\bar{Y}_1^n | W_{2c}) - H(\bar{Y}_1^n | W_{1c}, W_{1p}, W_{2c}) \tag{3.8}$$

$$\leq \alpha_{12} n \log \bar{P} - H(\bar{X}_2^n | W_{1c}, W_{1p}, W_{2c}) + n\ o(\log \bar{P}) \tag{3.9}$$

$$\leq \alpha_{12} n \log \bar{P} - H(W_{2p}) + n\ o(\log \bar{P}) \tag{3.10}$$

Where (3.7) is because W_2c is independent with other sub messages. (3.9) is due to the fact that given W_{1c}, W_{1p}, W_{2c}, Rx1 can reconstruct X_1^n and subtract it from its received signal. Then the remaining received signal at Rx1 is $\lfloor G_{12}^n \bar{X}_2^n \rfloor$. (3.10) is according to the AIS argument, we have $H(\lfloor G_{12}^n \bar{X}_2^n \rfloor) \geq H(\lfloor \sqrt{P^{1-\alpha_{12}}} G_{22}^n \bar{X}_2^n \rfloor)$. Then one can decode W_{2p} from $\lfloor \sqrt{P^{1-\alpha_{12}}} G_{22}^n \bar{X}_2^n \rfloor$.

Thus we have $nR_1 + H(W_{2p}) \leq \alpha_{12} n \log \bar{P}$.

In a same manner, we also have $nR_2 + H(W_{1p}) \leq \alpha_{21} n \log \bar{P}$.

Therefore, by combining these two inequalities, we have $d_{1c}+d_{2c}+2(d_{1p}+d_{2p}) = (2+\gamma)(d_{1p}+d_{2p}) \leq \alpha_{12} + \alpha_{21}$.

Thus $d_\Sigma(\gamma) \leq (1+\gamma)(d_{1p}+d_{2p}) \leq \frac{1+\gamma}{2+\gamma}(\alpha_{12}+\alpha_{21})$. This concludes the proof.

3.2 From GDoF to Partial CSIT

3.2.1 Sharp Contrast between Perfect and Finite Precision CSIT

At one extreme, if the CSIT is perfect, i.e., available with infinite precision, then tremendous DoF gains are possible, mainly through zero-forcing and interference alignment [22]. At the other extreme, if the CSIT is absent then the DoF collapse [21, 38, 46, 32]. In fact, even if partial CSIT is present, as long as it is limited to finite precision, then the DoF still collapse [6]. For example, consider an arbitrary channel coefficient H_{ij}, which is modeled under partial CSIT as

$$H_{ij} = \hat{H}_{ij} + \sqrt{\epsilon}\tilde{H}_{ij}$$

so that $\hat{H}_{ij}$ is the channel estimate known to the transmitter, while $\tilde{H}_{ij}$ is the normalized estimation error, with mean squared error $\epsilon > 0$. Even if ϵ is very small, as long as ϵ does not diminish with SNR (P),[1] the DoF collapse. Since in practice, CSIT can only be obtained to finite precision, at first sight the collapse of DoF under finite precision CSIT seems to suggest that there is no benefit of zero-forcing or interference alignment techniques in practice.

3.2.2 Expanding the DoF Formulation to Capture Partial CSIT

Upon careful assessment it becomes evident that the collapse of DoF under finite precision CSIT is primarily due to the limitation of the traditional DoF formulation which cannot distinguish between the relative strengths of constants (e.g., any non-zero channel, regardless of its strength, carries 1 DoF). Intuitively, from a DoF perspective a small estimation error ϵ is no different than a large estimation error. Since the quality of channel estimates is a such a crucial factor, it is important to expand the DoF formulation to be non-trivially responsive to this parameter. Motivated by the generalized degrees of freedom (GDoF) framework, the channel estimation error strength is captured in the parameter β, so that $\epsilon = P^{-\beta}$ [42, 16, 6, 26, 9, 19]. With this formulation, it turns out that in the DoF sense, $\beta = 1$ corresponds to perfect CSIT while $\beta = 0$ corresponds to no CSIT (also finite precision CSIT). As β spans the range of values between 0 and 1 it captures all intermediate levels of partial CSIT. *While the scaling of estimation error with SNR may seem unnatural for a given channel, the interpretation consistent with the GDoF framework, is not that the SNR is increasing for a given channel, but rather that a given channel is only associated with a given SNR.* As SNR value is allowed to increase, each new value of SNR defines a new channel.

[1] Following convention, we use P to represent the nominal SNR variable.

The reason this class of channels is studied together is because, normalized by log(SNR), they have the same approximate capacity. Indeed this is precisely how the GDoF metric has been used to find the approximate capacity of several wireless networks of interest including, most prominently, the capacity characterization of the 2 user interference channels to an accuracy of within 1 bit for all choices of channel parameters [10].

3.2.3 DoF under Partial CSIT: Signal Space Partitioning

DoF under partial CSIT have been studied under a variety of settings [42, 16, 6]. A common observation repeatedly encountered in these studies is the idea of signal space partitioning in accordance with partial CSIT. Starting from the earliest instances in [16, 42], essentially the same phenomenon has been recognized independently as interference enhancement[2] [7] and topological rate-splitting [26]. The broad implications of signal space partitioning are most recently highlighted in [9] as follows. Essentially, the signal space is partitioned according to the partial CSIT level β, so that the bottom β power levels correspond to perfect CSIT, while the remaining top $1-\beta$ power levels correspond to no CSIT. To understand the idea of signal space partitioning intuitively, consider a wireless network where all channels are subject to CSIT level β. For each transmitter in this network, the transmit signal X (subject to transmit power P) is decomposed into two parts $\hat{X}$ and $\tilde{X}$ corresponding to perfect and no CSIT respectively, each normalized to unit power, and each encoded independently from Gaussian codebooks, so that

$$\begin{aligned} X &= \sqrt{P^\beta}\hat{X} + \sqrt{P - P^\beta}\tilde{X} \\ &= \sqrt{P^\beta}\hat{X} + \sqrt{P}\tilde{X} + O(1) \end{aligned}$$

[2] As explicitly shown in [16] and also observed recently in [9] for the vector broadcast setting, the partial CSIT setting framework where estimation error decays with a constant negative exponent of SNR, translates into the GDoF framework where channels have strengths that scale with different SNR exponents and only finite precision CSIT is available. This is because without loss of generality, the transmitter can rotate its signal space to map estimated zero-forcing directions directly to specific transmit antennas. As such GDoF studies under finite precision CSIT translate into DoF studies under partial CSIT.

where $O(1)$ is a negligible term for DoF purposes whose power is bounded by a constant. Note that as this transmitted signal goes through a channel H, its contribution to the received signal is

$$\begin{aligned} HX &= (\hat{H} + \sqrt{P^{-\beta}}\tilde{H})(\sqrt{P^{\beta}}\hat{X} + \sqrt{P}\tilde{X} + O(1)) \\ &= \sqrt{P^{\beta}}\hat{H}\hat{X} + \sqrt{P}H\tilde{X} + O(1) \end{aligned}$$

Thus, at each receiver, all the different $\tilde{X}$ signals from every transmitter are received at power $\sim P$, while the $\hat{X}$ signals from every transmitter are received at power $\sim P^{\beta}$. The $\tilde{X}$ signals go through the partially known channel H, and carry only common message(s) which are decoded by every receiver (e.g., as a multiple access channel) while treating the interference from the $\hat{X}$ parts as noise. Since this decoding has an SINR value $P/P^{\beta} = P^{1-\beta}$, the common messages achieve a total of $1 - \beta$ DoF. Once the $\tilde{X}$ terms are decoded and subtracted out, only the $\hat{X}$ terms are left. For these terms note that the SNR is P^{β} and very importantly, these terms only go through the channel estimate $\hat{H}$ which is perfectly known to the transmitter. Therefore, the $\hat{X}$ signals are able to achieve β times the DoF value under perfect CSIT.

This achievability argument based on signal space partitioning is broadly applicable. For example, consider the K user interference channel, which has $\hat{D} = K/2$ DoF under perfect CSIT [2] and only $\tilde{D} = 1$ DoF under finite-precision CSIT [6]. If all channels have channel uncertainty level β, then the K user interference channel achieves $\beta\hat{D}$ DoF from the $\hat{X}$ codewords and $(1-\beta)\tilde{D}$ DoF from the $\tilde{X}$ codewords, for a total of $\frac{K}{2}\beta + 1 - \beta$ DoF. Similarly, the X channel which has K_1 transmitters and K_2 receivers, and achieves $\hat{D} = K_1K_2/(K_1 + K_2 - 1)$ DoF under perfect CSIT, and only $\tilde{D} = 1$ DoF under finite precision CSIT, achieves $\beta\hat{D} + (1-\beta)\tilde{D}$ with partial CSIT level β. As the final example, consider the MISO BC with K transmit antennas and K single antenna users which has $\hat{D} = K$ DoF with perfect CSIT and only $\tilde{D} = 1$ DoF under finite precision CSIT. With partial CSIT level β, this channel has exactly $\beta\hat{D} + (1-\beta)\tilde{D}$ DoF as shown in [9] where both achievability and outer bound are shown to prove the optimality of this DoF value, and therefore also the optimality of signal space partitioning under partial CSIT. Thus, note that as the partial CSIT level β spans the range between 0 and 1, it bridges the contrasting extremes of DoF under perfect CSIT and finite precision CSIT.

The idea of signal space partitioning for partial CSIT allows generalizations to settings with asymmetric β parameters through multilevel hierarchical partitions, with each power level (measured in terms of the exponent of P) allowing perfect CSIT for those links whose CSIT parameters β are at that level or higher. Progress along these lines is reported in [19]. Another generalization, reported in [9], explores the role of partial CSIT is conjunction with the diversity of channel strengths as measured through power exponents in the GDoF framework. However, a most interesting direction that remains unexplored is the role of signal space partitioning in MIMO interference channels, especially with arbitrary antenna configurations and arbitrary partial CSIT levels. This is the direction that we wish to explore in this work.

We explore the DoF of a 2 user MIMO interference channel with arbitrary antenna configurations (M_1, M_2 antennas at transmitters $1, 2$ and N_1, N_2 antennas at receivers $1, 2$, respectively) and arbitrary partial CSIT levels. Specifically, we ask for the DoF that can be achieved by one user while the other user achieves his maximum possible interference-free DoF. The focus here is on achievable schemes, leaving the outer bounds for future work. As one might expect, signal space partitioning becomes a much more sophisticated in a MIMO setting. As a highlight, we note the need for "elevated multiplexing", i.e., spreading of signals across transmit antennas at elevated power levels. A remarkable consequence of elevated multiplexing is that there is a DoF benefit from increasing the number of antennas at a transmitter even when it already has more antennas than its desired receiver and no CSIT is available to the transmitter.

3.3 System Model

Consider a 2-user Gaussian MIMO interference channel, where transmitters 1, 2 are equipped with M_1, M_2 antennas, respectively, and receivers 1, 2 are equipped with N_1, N_2 antennas, respectively. Each transmitter wishes to send an independent message to its corresponding receiver. At time slot $t \in \mathbb{N}$, the channel input-output equations are given by

$$Y_1(t) = \mathbf{H}_{11}(t)X_1(t) + \mathbf{H}_{12}(t)X_2(t) + Z_1(t), \tag{3.11}$$

$$Y_2(t) = \mathbf{H}_{21}(t)X_1(t) + \mathbf{H}_{22}(t)X_2(t) + Z_2(t), \tag{3.12}$$

Here, $X_k(t)$ is the $M_k \times 1$ signal vector sent from Transmitter k, $k \in \{1, 2\}$, which is subject to the power constraint P. $Y_k(t)$ is the $N_k \times 1$ received signal vector at Receiver k. $Z_k(t)$ is the $N_k \times 1$ i.i.d. additive white Gaussian noise (AWGN) vector at Receiver k, each entry of which is an i.i.d. Gaussian random variable with zero-mean and unit-variance. $\mathbf{H}_{ji}(t)$ is the $N_j \times M_i$ channel matrix from Transmitter i to Receiver j. Under partial CSIT, channel matrices $\mathbf{H}_{ji}(t)$, $\forall i, j \in \{1, 2\}$, are represented as

$$\mathbf{H}_{ji}(t) = \hat{\mathbf{H}}_{ji}(t) + \sqrt{P^{-\beta_{ji}}}\tilde{\mathbf{H}}_{ji}(t) \tag{3.13}$$

where $\hat{\mathbf{H}}_{ji}(t)$ is the $N_j \times M_i$ estimated channel matrix while $\tilde{\mathbf{H}}_{ji}(t)$ is the $N_j \times M_i$ estimation error matrix. We assume that the entries of $\hat{\mathbf{H}}_{ji}(t)$ and $\tilde{\mathbf{H}}_{ji}(t)$ are drawn from continuous joint distributions with bounded densities, with the difference that the actual realizations of $\hat{\mathbf{H}}_{ji}(t)$ are revealed to the transmitters, but the realizations of $\tilde{\mathbf{H}}_{ji}(t)$ are not available to the transmitter. To avoid degenerate conditions, the ranges of values of all channel coefficients are bounded away from infinity. The parameter β_{ji} measures the quality of the channel estimate. If $\beta_{ji} = 0$, then it corresponds to the case when there is no current CSIT. If $\beta_{ji} \geq 1$, then it corresponds to the case that the current CSIT is as good as perfect (for DoF). Throughout this paper, we assume that $\beta \in [0, 1]$.

Since codebooks, probability of error, achievable rates $(R_1,\ R_2)$ and capacity region $\mathcal{C}(P)$ are all defined in the standard Shannon theoretic sense, their definitions will not be repeated here. The DoF tuple $(d_1,\ d_2)$ is said to be achievable if there exist $(R_1(P), R_2(P)) \in \mathcal{C}(P)$ such that

$$d_1 = \lim_{p\to\infty} \frac{R_1(P)}{\log(P)}, \tag{3.14}$$

$$d_2 = \lim_{p\to\infty} \frac{R_2(P)}{\log(P)}. \tag{3.15}$$

We are interested in the DoF achievable by a user while the other user is achieving his interference-free maximum DoF. To this end, without loss of generality, we will assume, that User 1 achieves $d_1 = \min(M_1, N_1)$ DoF, and explore the DoF that are simultaneously achievable by User 2.

3.4 Examples

Before stating the general result, we present a few examples that highlight key ideas, in particular what we mean by "elevated multiplexing". Consider a transmitter that has no CSIT, and has as many antennas as its desired receiver. Is there a DoF benefit from further increasing the number of antennas at such a transmitter? Additional antennas are typically useful for zero-forcing or interference alignment. Since the absence of CSIT makes both zero-forcing and interference alignment impossible for this transmitter, one might expect that additional transmit antennas bring no DoF benefit. The following examples shows that indeed there is a DoF benefit from additional antennas, and the key to this counterintuitive outcome is the idea of elevated multiplexing.

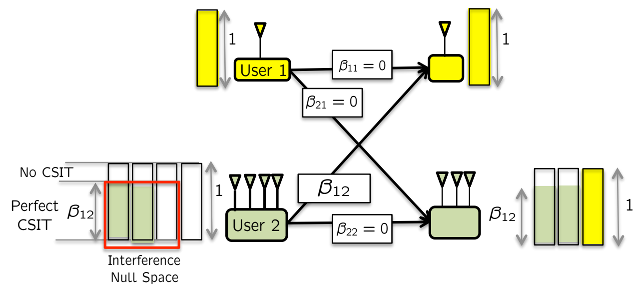

Figure 3.1: $(M_1, M_2, N_1, N_2) = (1, 4, 1, 3)$, $(d_1, d_2) = (1, 2\beta_{12})$.

3.4.1 $(M_1, M_2, N_1, N_2) = (1, 4, 1, 3)$

Let us start with the setting $(M_1, M_2, N_1, N_2) = (1, 4, 1, 3)$, where User 1 achieves $d_1 = 1$, i.e., his maximum DoF. Suppose Transmitter 2 has partial CSIT level β_{12} for his interference carrying link to Receiver 1, but no other CSIT is available, i.e., all other $\beta_{ij} = 0$. Since $d_1 = 1$, the signal from User 1 occupies one full spatial dimension at both receivers. At Receiver 1, this exhausts the desired signal space, so any interference from User 2 should not rise above the noise floor (in the DoF sense). The only signal space this leaves for Transmitter 2 is the null-space of the estimated channel $\hat{\mathbf{H}}_{12}$, within which Transmitter 2 must not exceed the power level $P^{\beta_{12}}$. Only 2 dimensions are left free from interference at Receiver 2, and the desired signal power in each dimension is $P^{\beta_{12}}$. So User 2 achieves $d_2 = 2\beta_{12}$.

Mathematically, the transmitted signals are,

$$X_1 = \sqrt{P}\tilde{X}_1$$
$$X_2 = \sqrt{P^{\beta_{12}}}(V_{21}\hat{X}_{21} + V_{22}\hat{X}_{22})$$

Here $\tilde{X}_1, \hat{X}_{21}, \hat{X}_{22}$ are independent Gaussian codewords from unit power codebooks which carry $1, \beta_{12}, \beta_{12}$ DoF, respectively. V_{21}, V_{22} are 4×1 unit vectors in the null space of $\hat{\mathbf{H}}_{12}$, i.e.,

$$\hat{\mathbf{H}}_{12}[V_{21} \quad V_{22}] = [0 \;\; 0]$$

The received signals are

$$
\begin{aligned}
Y_1 &= \sqrt{P}\mathbf{H}_{11}\tilde{X}_1 + \sqrt{P^{\beta_{12}}}(\hat{\mathbf{H}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{H}}_{12})(V_{21}\hat{X}_{21} + V_{22}\hat{X}_{22}) + Z_1 \\
&= \sqrt{P}\mathbf{H}_{11}\tilde{X}_1 + O(1) + Z_1 \\
Y_2 &= \sqrt{P}\mathbf{H}_{21}\tilde{X}_1 + \sqrt{P^{\beta_{12}}}\mathbf{H}_{22}(V_{21}\hat{X}_{21} + V_{22}\hat{X}_{22}) + Z_2
\end{aligned}
$$

Thus, $(d_1, d_2) = (1, 2\beta_{12})$ is achieved. Incidentally, in this channel if $d_1 = 1$, then the maximum possible DoF for User 2 with no CSIT ($\beta_{12} = 0$) is $\tilde{d}_2 = 0$, and with perfect CSIT ($\beta_{12} = 1$) is $\hat{d}_2 = 2$. Therefore, the subspace partition scheme presented above achieves $d_2 = \beta_{12}\hat{d}_2 + (1 - \beta_{12})\tilde{d}_2$ DoF, which can be shown to be optimal.

3.4.2 $(M_1, M_2, N_1, N_2) = (3, 4, 1, 3)$

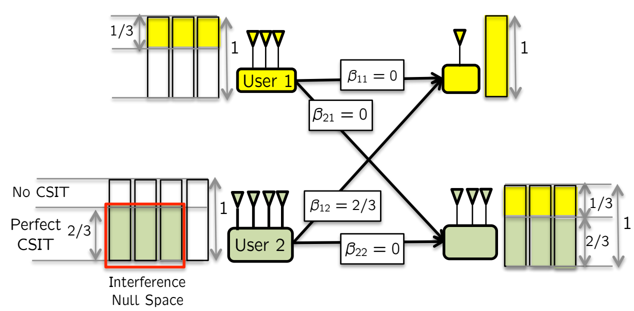

Figure 3.2: $(M_1, M_2, N_1, N_2) = (3, 4, 1, 3)$. Elevated multiplexing at Transmitter 1 helps achieve $(d_1, d_2) = (1, \min(2, 3\beta_{12}))$.

Even though in the previous example Transmitter 1 already has as many antennas as Receiver 1, let us further increase the number of transmit antennas to $M_1 = 3$, while keeping everything else the same, so Transmitter 1 still has no CSIT and $d_1 = 1$. To further simplify the exposition, let us consider specifically $\beta_{12} = 2/3$. Remarkably, as shown in Figure 3.2, it is now possible for User 2 to achieve 2 DoF (same as with perfect CSIT). To accomplish this, User 1 multiplexes his 1 DoF into three streams, each carrying 1/3 DoF and transmits them from its three antennas, each with elevated power level $\sim P$. At the same time, User 2 transmits three streams, each with power $P^{2/3}$ along the three dimensions that are in the null space of his estimated channel to Receiver 1. As before this signal space partitioning ensures that the interference caused at Receiver 1 from Transmitter 2 remains at the noise floor level. In the absence of interference, Receiver 1 *jointly* decodes the three desired streams from Transmitter 1 as a multiple access channel (MAC). Receiver 2 first decodes the interfering signal from Transmitter 1 by treating its own desired signals as noise. Each of the three desired streams is received at power level $\sim P^{2/3}$ while each of the undesired streams is received at power level P, so the SINR for each stream is $P/P^{2/3} = P^{1/3}$. Since each interfering stream carries only 1/3 DoF, and Receiver 2 has 3 antennas to separate the streams, it is able to decode and subsequently remove all interference. This leaves only the desired signal streams, which are then decoded to achieve $d_2 = 2/3 \times 3 = 2$ DoF for User 2. Note that this is clearly optimal, in fact it is also the best possible DoF for User 2 even if perfect CSIT was available to both transmitters. Also note the role of elevated multiplexing at Transmitter 1 which has no CSIT. Because of this elevated multiplexing, Receiver 1 is able to resolve the three streams jointly in its one interference-free received dimension, while Receiver 2 is able to simultaneously resolve the three streams separately in its 3 received dimensions, each of which sees an elevated noise floor (due to his desired signals) of $P^{2/3}$. Generalization to other values of β_{12} is straightforward. If $\beta_{12} > 2/3$ then $(d_1, d_2) = (1, 2)$ is still trivially achievable because improved CSIT cannot hurt. If $\beta_{12} < 2/3$ then Transmitter 2 sets the power level and the DoF of each of his 3 streams as β_{12} to keep the interference at Receiver 1 below the noise floor. The signals are decoded as before at each receiver, to achieve the DoF tuple $(d_1, d_2) = (1, 3\beta_{12})$.

3.4.3 $(M_1, M_2, N_1, N_2) = (2, 4, 1, 3)$.

Consider now that Transmitter 1 has $M_1 = 2$ antennas, while all other assumptions remain the same. In this case, if User 1 achieves $d_1 = 1$ DoF, User 2 can simultaneously achieve $d_2 = \min(1 + \beta_{12}, 3\beta_{12})$ DoF as shown in Figure 3.3. To accomplish this, User 1 multiplexes his 1 DoF into 2 streams, each carrying 1/2 DoF and transmits them from

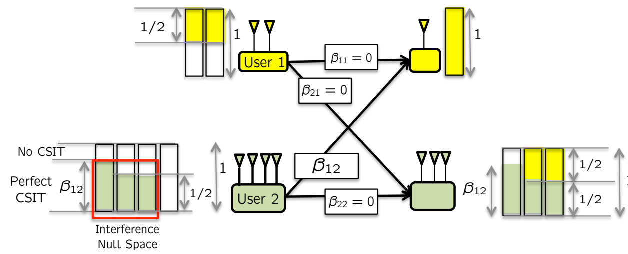

Figure 3.3: $(M_1, M_2, N_1, N_2) = (2, 4, 1, 3)$. Elevated multiplexing at Transmitter 1 helps achieve $(d_1, d_2) = (1, \min(1 + \beta_{12}, 3\beta_{12}))$.

its two antennas, each with elevated power level $\sim P$. At the same time, User 2 transmits three streams, the first with power $P^{\beta_{12}}$ and the next two with power $P^{\min(1/2,\beta_{12})}$, along the three dimensions that are in the null space of his estimated channel to Receiver 1 ($\beta_{12} > 1/2$ in Figure 3.3). As before this signal space partitioning ensures that the interference caused at Receiver 1 from Transmitter 2 remains at the noise floor level. In the absence of interference, Receiver 1 *jointly* decodes the two desired streams from Transmitter 1 as a multiple access channel (MAC). Receiver 2 first decodes the interfering signal from Transmitter 1 in the two dimensional space orthogonal to its own first desired stream, by treating its remaining desired signals as noise. Each of the two remaining desired streams is received at power level $\sim P^{\min(1/2,\beta_{12})}$ while each of the undesired streams is received at power level P, so the SINR for each stream is $P^{1-\min(1/2,\beta_{12})} \geq P^{1/2}$. Since each interfering stream carries only 1/2 DoF, Receiver 2 is able to decode and subsequently remove all interference. This leaves only the desired signal streams, which are then decoded to achieve $d_2 = \beta_{12} + \min(1/2, \beta_{12}) + \min(1/2, \beta_{12})$ DoF for User 2.

The three examples discussed so far are summarized in Figure 3.4.

3.4.4 $(M_1, M_2, N_1, N_2) = (1, 4, 2, 3)$.

For the next example, we consider the setting $(M_1, M_2, N_1, N_2) = (1, 4, 2, 3)$ where User 1 has more receive antennas than transmit antennas. We consider $\beta_{12} = 1/2$. Here $(d_1, d_2) = (1, 2)$ is achieved as shown in Figure 3.5. This example shows how elevated multiplexing is useful at

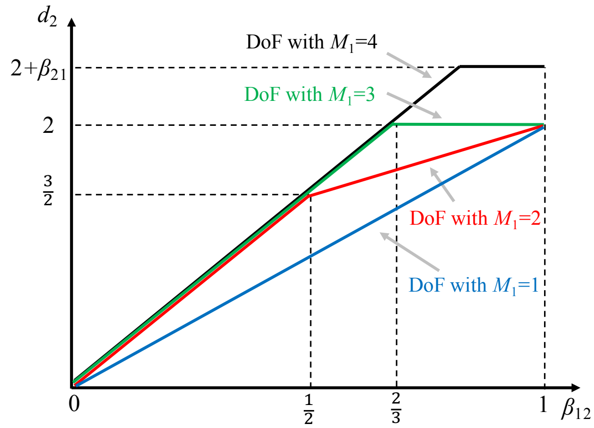

Figure 3.4: $(M_2, N_1, N_2) = (4, 1, 3)$. DoF achieved by User 2 when User 1 achieves his maximum DoF, $d_1 = 1$.

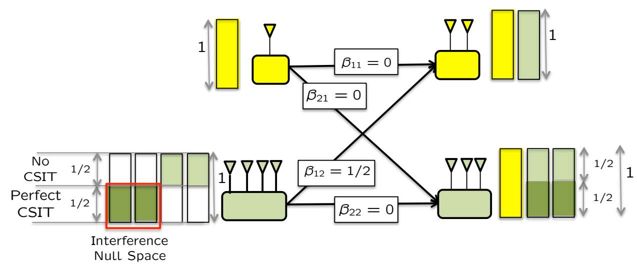

Figure 3.5: $(M_1, M_2, N_1, N_2) = (1, 4, 2, 3)$. Elevated multiplexing at Transmitter 2 helps achieve $(d_1, d_2) = (1, 2)$.

User 2, in a way that the multiplexed signals are decoded separately in space by the desired receiver and jointly in signal levels by the undesired receiver. User 1 simply sends his 1 DoF carrying stream from his single transmit antenna at power level $\sim P$. Transmitter 2 fully occupies the two dimensions in the null space of $\hat{\mathbf{H}}_{12}$, along which it can send at power levels up to $P^{1/2}$ without exceeding the noise floor at Receiver 1. Since Receiver 1 also has an extra dimension, Transmitter 2 uses elevated multiplexing to send two more streams, each carrying 1/2 DoF at elevated power levels of $\sim P$, along generic directions. At Receiver 1, first the desired signal is zero forced and the two elevated interference streams are jointly decoded in the remaining dimension. After these interfering streams are removed, the Receiver is able to decode its desired signal to recover the desired $d_1 = 1$ DoF. Receiver 2 on the other hand, first zero forces the interference and in the remaining 2 dimensions, first decodes the elevated

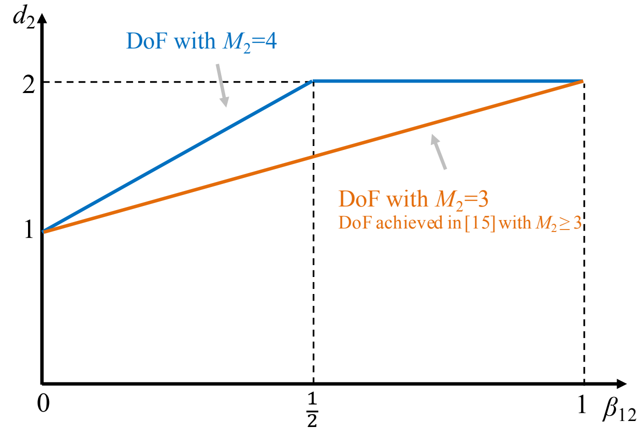

Figure 3.6: The achievable DoF for User 2 as a function of β_{12}, when User 1 achieves its maximum DoF with $(M_1, N_1, N_2) = (1, 2, 3)$.

streams while treating the other streams as noise. Since the elevated streams have power $\sim P$ while the other 2 desired streams have power $\sim P^{1/2}$, the SINR for this decoding is $P^{1/2}$ per dimension, which allows Receiver 2 to decode and subtract both elevated streams. The remaining streams are then decoded separately along the two interference-free dimensions. The scheme is easily generalized to arbitrary β_{12} values. The results for this example (and the case $M_2 = 3$ for comparison) are shown in Figure 3.6.

3.4.5 $(M_1, M_2, N_1, N_2) = (4, 4, 1, 3)$.

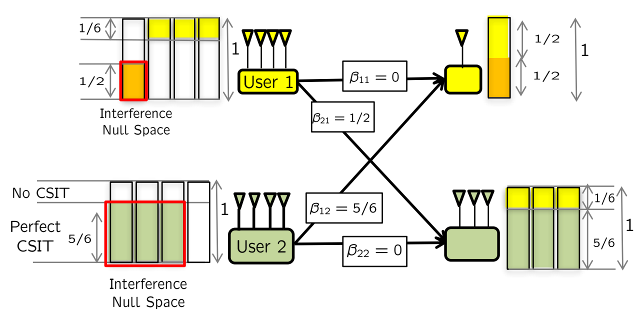

Figure 3.7: $(M_1, M_2, N_1, N_2) = (4, 4, 1, 3)$. Elevated multiplexing at Transmitter 2 helps achieve $(d_1, d_2) = (1, 2.5)$.

While it turns out that the CSIT of desired channels is irrelevant for DoF, in general the DoF may depend on the partial CSIT level at *both* cross channels, i.e., β_{12} and β_{21}. Our final example, illustrates such a setting. Specifically we consider $(M_1, M_2, N_1, N_2) = (4, 4, 1, 3)$ with $\beta_{12} = 5/6$ and $\beta_{21} = 1/2$, where $(d_1, d_2) = (1, 5/2)$ is achieved as shown in Figure 3.7. Transmitters 1 and 2 fully exploit the null-space of the estimated channel to the undesired receiver, up to a power levels $P^{\beta_{21}}$ and $P^{\beta_{12}}$, respectively, which keeps this interference below the noise floor at the undesired receiver. Transmitter 1 uses elevated multiplexing to obtain the remaining $1 - \beta_{21} = 1/2$ of its DoF by splitting into three streams which carry 1/6 DoF each. At Receiver 1 the multiplexed streams are jointly decoded to achieve $d_1 = 1$ DoF. At Receiver 2 the interfering multiplexed streams are separated in spatial dimensions and decoded while treating its own desired signal as noise. Removing the decoded interference then allows Receiver 2 to decode its desired signal to achieve $d_2 = 5/2$ DoF. For arbitrary values of β_{12}, β_{21}, the scheme generalizes to achieve $(d_1, d_2) = (1, \min(3\beta_{12}, 2 + \beta_{21}))$.

3.5 MIMO IC: General setting

With the key ideas highlighted in the previous section through various examples, we are now ready to consider the MIMO interference channel with arbitrary number of antennas at each node and arbitrary levels of partial CSIT. The achievable DoF for this general setting are stated in the following theorem.

Theorem 3.2. *For the 2-user MIMO interference channel with partial CSIT, if User 1 achieves its interference-free DoF, i.e.,* $d_1 = \min\{M_1, N_1\}$*, the DoF value achieved by User 2 is*

$$d_2 = \min{}^{+}\{A, B, C\}, \tag{3.16}$$

where

$$A = \min[\max(M_1, N_2), \max(M_2, N_1), M_1{+}M_2, N_1{+}N_2] - \min(M_1, N_1), \tag{3.17}$$

$$
\begin{aligned}
B =& \min{}^+(N_1 - M_1, M_2) + \min(M_1, N_2) - \min(M_1, N_1) \\
& + \beta_{12}\min{}^+(N_2 - M_1, M_2) + \beta_{21}\min{}^+(M_1 - N_2, N_1), \qquad (3.18)\\
C =& \min{}^+(N_1 - M_1, M_2) + \beta_{12}\min{}^+(M_2 - N_1, N_2). \qquad (3.19)
\end{aligned}
$$

Note that in a same channel, if $d_2 = \min(M_2, N_2)$, then the DoF value achieved by User 1 can be obtained by just switching the indices in Theorem 3.2.

Let us conclude with some high level insights into the theorem. First, we note that in the theorem, A corresponds to the DoF achieved by User 2 with perfect CSIT, B corresponds to the restrictions at Receiver 2 needed to decode all the desired messages. C corresponds to the maximum DoF that can be sent by Transmitter 2 without hurting User 1.

Note also that the result matches the known DoF for two extreme cases, i.e., MIMO IC with perfect CSIT [23] when all $\beta_{ji} = 1$ and MIMO IC with no CSIT [46] when all $\beta_{ji} = 0$. However, more significantly, the result is not a simple extension of the two extreme cases.

Next we note that the DoF do not depend on β_{11}, β_{22}, i.e., the channel knowledge of desired links is not critical. This observation is consistent with the understanding of interference alignment and zero forcing schemes based on all previous studies.

Another remarkable observation is how the CSIT requirement changes with the null space of cross-links. It is clear that if there is no null space for the channel from Transmitter i to Receiver j, i.e., $M_i \geq N_j$, then the achievable DoF do not depend on β_{ji}, i.e., CSIT for this cross-link is not needed.

3.5.1 Proof for Achievability

We will consider the case where $d_1 = \min(M_1, N_1)$ and also the case where $d_2 = \min(M_2, N_2)$. In each case we will determine the achievable DoF of the other user. With this approach we can assume with no loss of generality that $N_1 \leq N_2$. Then the parameter space of Theorem 3.2 can be divided into the following four cases.

Case 1: $M_2 \leq N_2$

$M_2 \leq N_1$ is trivial because the DoF is the same as that with both perfect and no CSIT. Therefore, let us consider $M_2 > N_1$, so that (3.16) becomes

$$d_2 = (N_1 - M_1)^+ + \beta_{12}(M_2 - N_1). \tag{3.20}$$

With $d_1 = \min(M_1, N_1)$, (3.20) can be achieved with only partial zero-forcing precoding. In each channel use, User 1 sends $\min(M_1, N_1)$ streams, each carrying 1 DoF, and each with power level $\sim P$. Transmitter 2 fully occupies the $M_2 - N_1$ dimensions in the null space of $\hat{\mathbf{H}}_{12}$, along which it can send at power levels up to $P^{\beta_{12}}$ without exceeding the noise floor at Receiver 1. Since Receiver 1 also has $(N_1 - M_1)^+$ extra dimensions, Transmitter 2 sends $(N_1 - M_1)^+$ additional streams, each carrying 1 DoF at power levels of P, along generic directions.

Mathematically, the transmitted signals are,

$$X_1 = c_o\sqrt{P} \sum_{l=1}^{\min(M_1,N_1)} V_{1l}^c X_{1l}^c \tag{3.21}$$

$$X_2 = c_1\sqrt{P} \sum_{i=1}^{(N_1-M_1)^+} V_{2i}^c X_{2i}^c + c_2\sqrt{P^{\beta_{12}}} \sum_{j=1}^{M_2-N_1} V_{2j}^p X_{2j}^p \tag{3.22}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1\min(M_1,N_1)}^c$ and X_{21}^c, $\cdots$, $X_{2(N_1-M_1)^+}^c$ are independent Gaussian codewords from unit power codebooks, each carries 1 DoF, and the superscript 'c' is used to indicate that these codewords can be decoded by both receivers (common). X_{21}^p, $\cdots$, $X_{2(M_2-N_1)}^p$ are independent Gaussian codewords from unit power codebooks, each carries β_{12} DoF, and the superscript p is used to indicate that these are 'private', i.e., only decoded by the intended receiver, in this case User 2. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied. V_{1l}^c and V_{2i}^c are $M_1 \times 1$ and $M_2 \times 1$

generic unit vectors, respectively. V^p_{2j} are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{H}}_{12}\begin{bmatrix} V^p_{21} & V^p_{22} & \cdots & V^p_{2(M_2-N_1)} \end{bmatrix} = \mathbf{0} \tag{3.23}$$

The received signals are

$$\begin{aligned} Y_1 =& c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{\min(M_1,N_1)} V^c_{1l}X^c_{1l} + c_1\sqrt{P}\mathbf{H}_{12}\sum_{i=1}^{(N_1-M_1)^+} V^c_{2i}X^c_{2i} \\ &+ c_2\sqrt{P^{\beta_{12}}}(\hat{\mathbf{H}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{H}}_{12})\sum_{j=1}^{M_2-N_1} V^p_{2j}X^p_{2j} + Z_1 \qquad (3.24) \\ =& c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{\min(M_1,N_1)} V^c_{1l}X^c_{1l} + c_1\sqrt{P}\mathbf{H}_{12}\sum_{i=1}^{(N_1-M_1)^+} V^c_{2i}X^c_{2i} + O(1) + Z_1 \qquad (3.25) \\ Y_2 =& c_o\sqrt{P}\mathbf{H}_{21}\sum_{l=1}^{\min(M_1,N_1)} V^c_{1l}X^c_{1l} + c_1\sqrt{P}\mathbf{H}_{22}\sum_{i=1}^{(N_1-M_1)^+} V^c_{2i}X^c_{2i} \\ &+ c_2\sqrt{P^{\beta_{12}}}\mathbf{H}_{22}\sum_{j=1}^{M_2-N_1} V^p_{2j}X^p_{2j} + Z_2 \qquad (3.26) \end{aligned}$$

Since $\min(M_1, N_1) + (N_1-M_1)^+ = N_1$, Receiver 1 has enough antennas to decode all the streams carrying common messages by treating other signals as white noise. Similarly, Receiver 2 has enough antennas to decode all the streams separately due to $\min(M_1, N_1) + (N_1-M_1)^+ + M_2-N_1 \le N_2$.

On the other hand, if $d_2 = \min(M_2, N_2) = M_2$, then $d_1 = 0$ is trivially achieved.

Case 2: $M_1<N_1\le N_2<M_2$

In this case if $d_1 = M_1$ then (3.16) becomes

$$d_2 = \min[N_2 - M_1, N_1 - M_1 + \beta_{12}\min(M_2 - N_1, N_2 - M_1)]. \tag{3.27}$$

The examples in Section 3.4.4 correspond to this case. To achieve (3.27), not only partial zero-forcing precoding, but also the elevated multiplexing is required at Transmitter 2. In each channel use, User 1 simply sends M_1 streams from his M_1 transmit antennas, each carrying 1 DoF, and each at power level $\sim P$. Transmitter 2 occupies $\min(M_2 - N_1, N_2 - M_1)$ dimensions in the null space of $\hat{\mathbf{H}}_{12}$, along which it can send at power levels up to $P^{\bar{\beta}_{12}}$ without exceeding the noise floor at Receiver 1, where $\bar{\beta}_{12}$ is define as $\bar{\beta}_{12} = \min(\beta_{12}, \frac{N_2-N_1}{\min(M_2-N_1,N_2-M_1)})$. Since Receiver 1 also has N_1-M_1 extra dimensions, Transmitter 2 uses elevated multiplexing to send M_2 more streams, each carrying $\frac{N_1-M_1}{M_2}$ DoF at power levels of P, along generic directions.

Mathematically, the transmitted signals are,

$$X_1 = c_o\sqrt{P}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c \tag{3.28}$$

$$X_2 = c_1\sqrt{P}\sum_{i=1}^{M_2} V_{2i}^c X_{2i}^c + c_2\sqrt{P^{\bar{\beta}_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2-M_1)} V_{2j}^p X_{2j}^p \tag{3.29}$$

Here $X_{11}^c, X_{12}^c, \cdots, X_{1M_1}^c$ and $X_{21}^c, \cdots, X_{2M_2}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries 1 DoF while each X_{2i}^c carries $\frac{N_1-M_1}{M_2}$ DoF. $X_{21}^p, \cdots, X_{2\min(M_2-N_1,N_2-M_1)}^p$ are independent Gaussian codewords from unit power codebooks, each carries $\bar{\beta}_{12}$ DoF that to be decoded only by User 2. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here V_{1l}^c and V_{2i}^c are $M_1 \times 1$ and $M_2 \times 1$ generic unit vectors, respectively. V_{2j}^p is a $M_2 \times 1$ unit vector chosen so that

$$\hat{\mathbf{H}}_{12}\begin{bmatrix} V_{21}^p & V_{22}^p & \cdots & V_{2\min(M_2-N_1,N_2-M_1)}^p \end{bmatrix} = \mathbf{0} \tag{3.30}$$

The received signals are

$$\begin{aligned} Y_1 = & c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c + c_1\sqrt{P}\mathbf{H}_{12}\sum_{i=1}^{M_2} V_{2i}^c X_{2i}^c \\ & + c_2\sqrt{P^{\bar{\beta}_{12}}}(\hat{\mathbf{H}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{H}}_{12})\sum_{j=1}^{\min(M_2-N_1,N_2-M_1)} V_{2j}^p X_{2j}^p + Z_1 \end{aligned} \tag{3.31}$$

$$=c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c+c_1\sqrt{P}\mathbf{H}_{12}\sum_{i=1}^{M_2}V_{2i}^cX_{2i}^c+O(1)+Z_1 \tag{3.32}$$

$$\begin{aligned}Y_2=&c_o\sqrt{P}\mathbf{H}_{21}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c+c_1\sqrt{P}\mathbf{H}_{22}\sum_{i=1}^{M_2}V_{2i}^cX_{2i}^c\\&+c_2\sqrt{P^{\bar{\beta}_{12}}}\mathbf{H}_{22}\sum_{j=1}^{\min(M_2-N_1,N_2-M_1)}V_{2j}^pX_{2j}^p+Z_2\end{aligned} \tag{3.33}$$

At Receiver 1, first the signals from Transmitter 1 are zero forced and the M_2 elevated streams are jointly decoded in the remaining N_1-M_1 dimensions as a MAC channel. After all the X_{2i}^c are removed, Receiver 1 is able to decode the M_1 signals from Transmitter 1, i.e., X_{1l}^c. Receiver 2 on the other hand, first zero forced the signals from Transmitter 1 and in the remaining N_2-M_1 dimensions, first jointly decode M_2 elevated streams as a MAC channel while treating other $\min(M_2-N_1,N_2-M_1)$ streams carrying X_{2j}^p as noise. After X_{2i}^c are decoded and removed, X_{2j}^p are then decoded separately along $\min(M_2-N_1,N_2-M_1)$ interference-free dimensions.

On the other hand, in this setting, if $d_2=N_2$ then $d_1=0$ can be trivially achieved.

Case 3: $N_1{\le}M_1{\le}N_2{<}M_2$

In this case if $d_1=N_1$ then (3.16) becomes

$$d_2{=}\min[\beta_{12}\min(M_2-N_1,N_2),M_1-N_1+\beta_{12}(N_2-M_1)]. \tag{3.34}$$

The examples in Section 3.4.1, 3.4.2 and 3.4.3 correspond to this case. Partial zero-forcing precoding at User 2 is required. What's more, to help User 2 to achieve (3.34), User 1 needs to use elevated multiplexing. Specifically, Transmitter 1 multiplexes his N_1 DoF into M_1 streams, each carrying $\frac{N_1}{M_1}$ DoF with elevated power level $\sim P$. At the same time, User 2 occupies $\min(M_2-N_1,N_2)$ dimensions in the null space of $\hat{\mathbf{H}}_{12}$, the first N_2-M_1 streams are sent with power levels up to $P^{\beta_{12}}$ and the rest $M_1-N_2+\min(M_2-N_1,N_2)$ streams are sent with power levels up to $P^{\bar{\beta}_{12}}$ without exceeding the noise floor at Receiver 1, where $\bar{\beta}_{12}$

is defined as $\bar{\beta}_{12} = \min(\beta_{12}, \frac{M_1-N_1}{M_1-N_2+\min(M_2-N_1,N_2)})$.

Mathematically, the transmitted signals are,

$$X_1 = c_o\sqrt{P}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c \tag{3.35}$$

$$X_2 = c_1\sqrt{P^{\beta_{12}}}\sum_{j=1}^{N_2-M_1} V_{2j}^p X_{2j}^p + c_2\sqrt{P^{\bar{\beta}_{12}}}\sum_{i=N_2-M_1+1}^{\min(M_2-N_1,N_2)} V_{2i}^p X_{2i}^p \tag{3.36}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\frac{N_1}{M_1}$ DoF. X_{21}^p, $\cdots$, $X_{2\min(M_2-N_1,N_2)}^p$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X_{2j}^p and X_{2i}^p carries β_{12} and $\bar{\beta}_{12}$ DoF, respectively. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here V_{1l}^c are $M_1 \times 1$ generic unit vectors. V_{2j}^p and V_{2i}^p are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{H}}_{12}\begin{bmatrix} V_{21}^p & V_{22}^p & \cdots & V_{2\min(M_2-N_1,N_2)}^p \end{bmatrix} = \mathbf{0} \tag{3.37}$$

The received signals are

$$Y_1 = c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c + (\hat{\mathbf{H}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{H}}_{12})X_2 + Z_1 \tag{3.38}$$

$$= c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c + O(1) + Z_1 \tag{3.39}$$

$$\begin{aligned} Y_2 =& c_o\sqrt{P}\mathbf{H}_{21}\sum_{l=1}^{M_1} V_{1l}^c X_{1l}^c + c_1\sqrt{P^{\beta_{12}}}\mathbf{H}_{22}\sum_{j=1}^{N_2-M_1} V_{2j}^p X_{2j}^p \\ &+ c_2\sqrt{P^{\bar{\beta}_{12}}}\mathbf{H}_{22}\sum_{i=N_2-M_1+1}^{\min(M_2-N_1,N_2)} V_{2i}^p X_{2i}^p + Z_2 \end{aligned} \tag{3.40}$$

As before the signal space partitioning ensures that the interference caused at Receiver 1 from Transmitter 2 remains at the noise floor level. In the absence of interference, Receiver 1 can jointly decode the M_1 desired streams from Transmitter 1 as a MAC channel.

At the same time, Receiver 2 zero forces the first $N_2 - M_1$ signals from Transmitter 2, i.e., X^p_{2j}. In the remaining M_1 dimensions, M_1 elevated streams from Transmitter 1 can be decoded (see Lemma 3.1 in the Section 3.6) by treating the rest $M_1 - N_2 + \min(M_2 - N_1, N_2)$ streams carrying X^p_{2i} as white noise. After X^c_{1l} are decoded and removed, all the remaining signals can then be decoded separately along $\min(M_2 - N_1, N_2)$ interference-free dimensions. On the other hand, in this setting, if $d_2 = N_2$ then $d_1 = 0$ can be trivially achieved.

Case 4: $N_1 \leq N_2 < \min(M_1, M_2)$

In this case if $d_1 = N_1$ then (3.16) becomes

$$d_2 = \min[\beta_{12} \min(M_2 - N_1, N_2), N_2 - N_1 + \beta_{21} \min(M_1 - N_2, N_1)]. \tag{3.41}$$

This case can be seen as an extension of Case 3 where there is null space for the channel from Transmitter 1 to Receiver 2, The examples in Section 3.4.5 correspond to this case. Thus d_2 depends on both β_{12} and β_{21}. To achieve (3.41), the only difference is that User 1 needs both partial zero-forcing precoding and elevated multiplexing to help User 2. Specifically, Transmitter 1 occupies $\min(M_1 - N_2, N_1)$ dimensions in the null space of $\hat{\mathbf{H}}_{21}$, along which it can send at power levels up to $P^{\beta_{21}}$ without exceeding the noise floor at Receiver 2. Then Transmitter 1 multiplexes his remaining $N_1 - \beta_{21} \min(M_1 - N_2, N_1)$ DoF into M_1 streams, each carrying $\frac{N_1 - \beta_{21} \min(M_1 - N_2, N_1)}{M_1}$ DoF with elevated power level $\sim P$. At the same time, Transmitter 2 occupies $\min(M_2 - N_1, N_2)$ dimensions in the null space of $\hat{\mathbf{H}}_{12}$, along which it can send at power levels up to $P^{\bar{\beta}_{12}}$ without exceeding the noise floor at Receiver 2, where $\bar{\beta}_{12}$ is defined as $\bar{\beta}_{12} = \min(\beta_{12}, \frac{N_2 - N_1 + \beta_{21} \min(M_1 - N_2, N_1)}{\min(M_2 - N_1, N_2)})$.
Mathematically, the transmitted signals are,

$$X_1 = c_o \sqrt{P} \sum_{l=1}^{M_1} V^c_{1l} X^c_{1l} + c_1 \sqrt{P^{\beta_{21}}} \sum_{k=1}^{\min(M_1 - N_2, N_1)} V^p_{1k} X^p_{1k} \tag{3.42}$$

$$X_2 = c_2 \sqrt{P^{\bar{\beta}_{12}}} \sum_{i=1}^{\min(M_2 - N_1, N_2)} V^p_{2i} X^p_{2i} \tag{3.43}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\frac{N_1-\beta_{21}\min(M_1-N_2,N_1)}{M_1}$ DoF. X_{11}^p, $\cdots$, $X_{1\min(M_1-N_2,N_1)}^p$ and X_{21}^p, $\cdots$, $X_{2\min(M_2-N_1,N_2)}^p$ are independent Gaussian codewords from unit power codebooks that are intended to be decoded only by their desired receiver. Each X_{1k}^p and X_{2i}^p carries β_{21} and $\bar{\beta}_{12}$ DoF, respectively. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here V_{1l}^c are $M_1 \times 1$ generic unit vectors. V_{1k}^p and V_{2i}^p are $M_1 \times 1$ and $M_2 \times 1$ unit vectors, respectively, chosen so that

$$\hat{\mathbf{H}}_{21}\begin{bmatrix}V_{11}^p & V_{12}^p & \cdots & V_{1\min(M_1-N_2,N_1)}^p\end{bmatrix} = \mathbf{0} \tag{3.44}$$

$$\hat{\mathbf{H}}_{12}\begin{bmatrix}V_{21}^p & V_{22}^p & \cdots & V_{2\min(M_2-N_1,N_2)}^p\end{bmatrix} = \mathbf{0} \tag{3.45}$$

The received signals are

$$\begin{aligned}
Y_1 =& c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c + c_1\sqrt{P^{\beta_{21}}}\mathbf{H}_{11}\sum_{k=1}^{\min(M_1-N_2,N_1)}V_{1k}^pX_{1k}^p \\
&+ c_2\sqrt{P^{\bar{\beta}_{12}}}(\hat{\mathbf{H}}_{12}+\sqrt{P^{-\beta_{12}}}\tilde{\mathbf{H}}_{12})\sum_{i=1}^{\min(M_2-N_1,N_2)}V_{2i}^pX_{2i}^p + Z_1 && (3.46)\\
=& c_o\sqrt{P}\mathbf{H}_{11}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c + c_1\sqrt{P^{\beta_{21}}}\mathbf{H}_{11}\sum_{k=1}^{\min(M_1-N_2,N_1)}V_{1k}^pX_{1k}^p + O(1) + Z_1 && (3.47)\\
Y_2 =& c_o\sqrt{P}\mathbf{H}_{21}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c + c_2\sqrt{P^{\bar{\beta}_{12}}}\mathbf{H}_{22}\sum_{i=1}^{\min(M_2-N_1,N_2)}V_{2i}^pX_{2i}^p \\
&+ c_1\sqrt{P^{\beta_{21}}}(\hat{\mathbf{H}}_{21}+\sqrt{P^{-\beta_{21}}}\tilde{\mathbf{H}}_{21})\sum_{k=1}^{\min(M_1-N_2,N_1)}V_{1k}^pX_{1k}^p + Z_2 && (3.48)\\
=& c_o\sqrt{P}\mathbf{H}_{21}\sum_{l=1}^{M_1}V_{1l}^cX_{1l}^c + c_2\sqrt{P^{\bar{\beta}_{12}}}\mathbf{H}_{22}\sum_{i=1}^{\min(M_2-N_1,N_2)}V_{2i}^pX_{2i}^p + O(1) + Z_2 && (3.49)
\end{aligned}$$

At Receiver 1 the multiplexed streams from Transmitter 1 are jointly decoded as a MAC channel. At the same time, at Receiver 2, M_1 elevated streams from Transmitter 1 can be decoded as a MAC channel (see Lemma 3.1 in the Section 3.6) by treating the remaining $\min(M_2 - N_1, N_2)$ streams carrying X_{2i}^p as white noise. After X_{1l}^c are decoded and removed, all the rest of the signals can then be decoded separately along $\min(M_2-N_1, N_2)$ interference-free dimensions.

On the other hand, if $d_2 = \min(M_2, N_2) = N_2$, then the achievability of

$$d_1 = \overset{+}{\min}[\beta_{21} \min(M_1-N_2, N_1), N_1 - N_2 + \beta_{12} \min(M_2-N_1, N_2)]$$

can be shown similarly by simply switching the indices of the scheme in this subsection. The only difference is that when $N_1 - N_2 + \beta_{12} \min(M_2-N_1, N_2) \leq 0$, $\bar{\beta}_{21} \leq 0$, then in this case $d_1 = 0$.

3.6 Proof for Lemma 3.1

Consider a multiple access channel with K signal antenna transmitters. The receiver has M antennas. The $M \times 1$ received signal vector Y is represented as follows

$$\mathbf{Y} = \sqrt{P} \sum_{k=1}^{K} \mathbf{H}_k X_k + \sum_{m=1}^{M} \sqrt{P^{\alpha_m}} \mathbf{G}_m Z_m \tag{3.50}$$

Here, $X_1, X_2, \cdots, X_K$ the transmitted symbols, normalized to unit transmit power constraint. Z_m are i.i.d. Gaussian zero mean unit variance terms. The $\mathbf{H}_k, \mathbf{G}_n$ are $M \times 1$ generic vectors, i.e., generated from continuous distributions with bounded density, so that any M of them are linearly independent almost surely. All $\alpha_m \in [0, 1]$.

Lemma 3.1. *The DoF tuple $(d_1, d_2, \cdots, d_K)$ is achievable in the multiple access channel described above, if*

$$\sum_{i \in \max,k} d_i + \sum_{j \in \min,\min(k,M)} \alpha_j \leq \min(k, M), \qquad \forall k \in [1, 2, \cdots, K] \tag{3.51}$$

where $\sum_{i \in \max,k} d_i$ is the sum of the k largest terms in $\{d_1, d_2, \cdots, d_K\}$ and $\sum_{j \in \min,\min(k,M)} \alpha_j$ is the sum of the $\min(k, M)$ *smallest terms in $\{\alpha_1, \alpha_2, \cdots, \alpha_M\}$.*

Proof. Choose all X_i as zero mean unit variance i.i.d. Gaussians. A rate tuple $(R_1, R_2, \cdots, R_K)$ is achievable if the following inequalities are satisfied.

$$\sum_{i\in\mathcal{U}} R_i \leq I(\{X_i, \forall i \in \mathcal{U}\}; \mathbf{Y}|\{X_j, \forall j \in \mathcal{U}^c\}), \quad \forall\, \mathcal{U} \subseteq \mathcal{I}_K, \tag{3.52}$$

where $\mathcal{U}$ can be any subset of $\mathcal{I}_K$.

$$\begin{aligned} &I(\{X_i, \forall i \in \mathcal{U}\}; \mathbf{Y}|\{X_j, \forall j \in \mathcal{U}^c\}) \\ &\quad = h(\mathbf{Y}|\{X_j, \forall j \in \mathcal{U}^c\}) - h(\mathbf{Y}|\{X_j, \forall j \in \mathcal{I}_K) && (3.53) \\ &\quad = \min(|\mathcal{U}|, M) \log P + \sum_{j\in\max, M-\min(|\mathcal{U}|,M)} \alpha_j \log P - \sum_{j=1}^{M} \alpha_j \log P + o(\log P) && (3.54) \\ &\quad = \min(|\mathcal{U}|, M) \log P - \sum_{j\in\min,\min(|\mathcal{U}|,M)} \alpha_j \log P + o(\log P) && (3.55) \end{aligned}$$

$|\mathcal{U}|$ is the cardinality of $\mathcal{U}$. $\sum_{j\in\max,M-\min(|\mathcal{U}|,M)} \alpha_j$ in (3.54) is the sum of the $M - \min(|\mathcal{U}|, M)$ largest terms in $\{\alpha_1, \alpha_2, \cdots, \alpha_M\}$. $\sum_{j\in\min,\min(|\mathcal{U}|,M)} \alpha_j$ in (3.55) is the sum of the $\min(|\mathcal{U}|, M)$ smallest terms in $\{\alpha_1, \alpha_2, \cdots, \alpha_M\}$. (3.54) follows from Lemma 3 in [11].
From (3.55), we obtain the achievable DoF region,

$$\sum_{i\in\mathcal{U}} d_i \leq \min(|\mathcal{U}|, M) - \sum_{j\in\min,\min(|\mathcal{U}|,M)} \alpha_j, \quad \forall\, \mathcal{U} \subseteq \mathcal{K}. \tag{3.56}$$

This concludes the proof. ■

3.7 Summary

In this chapter, we studied the two-user MIMO interference channel with partial CSIT and arbitrary antenna configuration at each node. Through various examples we introduced the ideas of signal space partitioning and elevated multiplexing, and how they work together. Remarkably, we found that there is a DoF benefit from increasing the number of antennas at a transmitter even if it has no CSIT and it already has more antennas than its desired receiver. Building upon these insights, a general achievable DoF result with partial CSIT was presented. Generalizations of this work to the DoF region, developing new tight outer bounds are presented in next chapter.

Chapter 4

Degrees of Freedom Region for MIMO IC with Partial CSIT

In this chapter we characterize the DoF region of a MIMO IC with arbitrary antenna configurations and arbitrary levels of partial CSIT. For the MIMO IC previous works have found DoF characterizations under perfect CSIT [23] and with no CSIT [21, 46, 38]. Achievable DoF regions under partial CSIT have also been found in [44, 20] with arbitrary antenna configurations (M_1, M_2 antennas at transmitters $1, 2$ and N_1, N_2 antennas at receivers $1, 2$, respectively) and arbitrary partial CSIT levels. Signal space partitioning becomes much more sophisticated in this MIMO setting. The authors in [44] studied the corner points of DoF region where one user achieves his maximum DoF. An achievability scheme based on "elevated multiplexing", i.e., spreading of signals across transmit antennas at elevated power levels, was proposed. It was shown that there is a DoF benefit from increasing the number of antennas at a transmitter even if that transmitter already has more antennas than its desired receiver and has no CSIT. Independently, Hao, Rasouli and Clerckx in [20] employed a space-time transmission scheme to obtain an achievable DoF region. Interestingly, a direct comparison of these two results shows that for the case $M_1 < N_1 \leq N_2 < M_2$, the approach in [44] based on elevated multiplexing is stronger. On the other hand, for the case

$N_1 \leq N_2 < \min(M_1, M_2)$, the approach in [20] based on space-time transmission is stronger. However, the optimal DoF for this channel remains unknown, mainly due to the difficulty of obtaining DoF outer bounds that are tight under partial CSIT. This is the direction that we wish to explore in this chapter. The DoF region of a MIMO IC with arbitrary antenna configurations and arbitrary levels of (partial) CSIT is fully characterized. The achievability results follows from the ideas presented in [44, 20], while the outer bounds are derived with the aid of the AIS approach, sum-set inequalities, and sub-modularity properties of entropy function.

For $n \in \mathbb{N}$, define the notation $[n] = \{1, 2, \cdots, n\}$. The notation $X^{[n]}$ stands for

$$\{X(1), X(2), \cdots X(n)\}$$

Moreover, $X_i^{[n]}$ also stands for $\{X_i(t) : \forall t \in [n]\}$. The support of a random variable X is denoted as $\text{supp}(X)$. We use $\mathbb{P}(\cdot)$ to denote the probability function $\text{Prob}(\cdot)$.

4.1 Definitions

Definition 4.1 (Bounded Density Channel Coefficients). *Define a set of real-valued random variables, $\mathcal{G}$ such that the magnitude of each random variable $g \in \mathcal{G}$ is bounded away from infinity, $|g| \leq \Delta_2 < \infty$, for some positive constant $\Delta_2 \geq 1$, and there exists a finite positive constant $f_{\max} \geq 1$, such that for all finite cardinality disjoint subsets $\mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{G}$, the joint probability density function of all random variables in $\mathcal{G}_1$, conditioned on all random variables in $\mathcal{G}_2$, exists and is bounded above by $f_{\max}^{|\mathcal{G}_1|}$.*

Definition 4.2 (Arbitrary Channel Coefficients). *Let $\mathcal{H}$ be a set of arbitrary constant values that are bounded above by Δ_2, i.e., if $h \in \mathcal{H}$ then $|h| \leq \Delta_2 < \infty$.*

Definition 4.3 (Power Levels). *Consider integer valued random variables X_i over alphabet*

$\mathcal{X}_{\alpha_i}$,

$$\mathcal{X}_{\alpha_i} \triangleq \{0, 1, 2, \cdots, \bar{P}^{\alpha_i}\} \tag{4.1}$$

where $\bar{P}^{\alpha_i}$ *is a compact notation for* $\lfloor \sqrt{P^{\alpha_i}} \rfloor$. *We refer to* $P \in \mathbb{R}_+$ *as* power, *and are primarily interested in limits as* $P \to \infty$. *Quantities that do not depend on* P *will be referred to as constants. The constant* $\alpha_i \in \mathbb{R}_+$ *denotes the* power level *of* X_i.

Definition 4.4. *For* $X \in \mathcal{X}_\alpha$, *and* $0 \le \alpha_1 \le \alpha_2 \le \alpha$, *define the random variables* $(X)^{\alpha}_{\alpha-\alpha_2}$, $(X)_{\alpha_1}$ *and* $(X)^{\alpha_2}_{\alpha_1}$ *as,*

$$(X)^{\alpha}_{\alpha-\alpha_2} \triangleq \left\lfloor \frac{X}{\bar{P}^{\alpha-\alpha_2}} \right\rfloor \tag{4.2}$$

$$(X)_{\alpha_1} \triangleq X - \bar{P}^{\alpha_1} \left\lfloor \frac{X}{\bar{P}^{\alpha_1}} \right\rfloor \tag{4.3}$$

$$(X)^{\alpha_2}_{\alpha_1} \triangleq \left\lfloor \frac{X - \bar{P}^{\alpha_2} \left\lfloor \frac{X}{\bar{P}^{\alpha_2}} \right\rfloor}{\bar{P}^{\alpha_1}} \right\rfloor \tag{4.4}$$

We may show $(X)^{\alpha}_{\alpha-\alpha_2}$ as $(X)^{\alpha_2}$ if there is no cause for ambiguity, i.e.,

$$(X)^{\alpha_2} \triangleq (X)^{\alpha}_{\alpha-\alpha_2} \tag{4.5}$$

In words, $(X)^{\alpha_1}$ retrieves the top α_1 power levels of X, while $(X)_{\alpha_1}$ retrieves the bottom α_1 levels of X. $(X)^{\alpha_2}_{\alpha_1}$ retrieves only the partition of X that lies between power levels α_1 and α_2. Note that $(X)^{\alpha_2}_{\alpha_1} = ((X)^{\alpha-\alpha_1})_{\alpha_2-\alpha_1}$. Also note that $X \in \mathcal{X}_\alpha$ can be expressed as $X = \bar{P}^{\alpha-\alpha_1}(X)^{\alpha_1} + (X)_{\alpha-\alpha_1}$ for $0 \le \alpha_1 \le \alpha$. Equivalently, suppose $X_1 \in \mathcal{X}_{\alpha_1}$, $X_2 \in \mathcal{X}_{\alpha_2}$, and $X = X_1 + X_2\bar{P}^{\alpha_1}$. Then $X_1 = (X)_{\alpha_1}$, $X_2 = (X)^{\alpha_2}$. Moreover, for $0 \le \alpha \le \alpha_1, \alpha_2$, define $(X)^{\alpha_2} \triangleq X$, $(X)_{\alpha_1} \triangleq X$. For the vector $\mathbf{V} = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix}^T$, we define $(\mathbf{V})^{\alpha_2}$, $(\mathbf{V})_{\alpha_1}$ and $(\mathbf{V})^{\alpha_2}_{\alpha_1}$ as,

$$(\mathbf{V})^{\alpha_2} \triangleq \begin{bmatrix} w_1 & w_2 & \cdots & w_k \end{bmatrix}^T, \text{ where, } w_r = \left\lfloor \frac{v_r}{\bar{P}^{\alpha-\alpha_2}} \right\rfloor, \forall r \in [k] \tag{4.6}$$

$$(\mathbf{V})_{\alpha_1} \triangleq \begin{bmatrix} w'_1 & w'_2 & \cdots & w'_k \end{bmatrix}^T, \text{ where, } w'_r = v_r - \bar{P}^{\alpha_1} \left\lfloor \frac{v_r}{\bar{P}^{\alpha_1}} \right\rfloor, \forall r \in [k] \tag{4.7}$$

$$(\mathbf{V})^{\alpha_2}_{\alpha_1} \triangleq \begin{bmatrix} w''_1 & w''_2 & \cdots & w''_k \end{bmatrix}^T, \text{ where, } w''_r = \left\lfloor \frac{v_r - \bar{P}^{\alpha_2} \left\lfloor \frac{v_r}{\bar{P}^{\alpha_2}} \right\rfloor}{\bar{P}^{\alpha_1}} \right\rfloor, \forall r \in [k] \tag{4.8}$$

Definition 4.5. *For any vector* $V = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}^T$ *and non-negative integer numbers* m *and* n *less than* k*, define*

$$V_{m,n} \triangleq \begin{cases} \begin{bmatrix} v_{m+1} & \cdots & v_{m+n} \end{bmatrix}^T, & m+n \leq k \\ \begin{bmatrix} v_{m+1} & \cdots & v_k & v_1 & \cdots & v_{m+n-k} \end{bmatrix}^T, & k < m+n \end{cases} \tag{4.9}$$

Moreover, for the two vectors $V = \begin{bmatrix} v_1 & \cdots & v_{k_1} \end{bmatrix}^T$ *and* $W = \begin{bmatrix} w_1 & \cdots & w_{k_2} \end{bmatrix}^T$ *define* $V; W$ *as* $\begin{bmatrix} v_1 & \cdots & v_{k_1} & w_1 & \cdots & w_{k_2} \end{bmatrix}^T$.

4.2 System Model

4.2.1 The Channel

The channel model for the two user MIMO IC is defined by the following input-output equations.

$$\mathbf{Y}_1(t) = \mathbf{G}_{11}(t)\mathbf{X}_1(t) + \mathbf{G}_{12}(t)\mathbf{X}_2(t) + \mathbf{\Gamma}_1(t), \tag{4.10}$$

$$\mathbf{Y}_2(t) = \mathbf{G}_{21}(t)\mathbf{X}_1(t) + \mathbf{G}_{22}(t)\mathbf{X}_2(t) + \mathbf{\Gamma}_2(t), \tag{4.11}$$

Here, $\mathbf{X}_k(t) = [X_k^1(t)\ X_k^2(t)\ \cdots\ X_k^{M_k}(t)]^T$ is the $M_k \times 1$ signal vector sent from Transmitter k, $k \in \{1, 2\}$, which is subject to the power constraint P. $\mathbf{Y}_k(t) = [Y_k^1(t)\ Y_k^2(t)\ \cdots\ Y_k^{N_k}(t)]^T$ is the $N_k \times 1$ the received signal vector at Receiver k. $\mathbf{\Gamma}_k(t)$ is the $N_k \times 1$ i.i.d. additive white Gaussian noise (AWGN) vector at Receiver k, each entry of which is an i.i.d. Gaussian random variable with zero-mean and unit-variance. $\mathbf{G}_{ji}(t)$ is the $N_j \times M_i$ channel matrix from Transmitter i to Receiver j. $G_{ji}^{[nm]}(t)$ is the element of $\mathbf{G}_{ji}(t)$ on n-th row and m-th column. Note that while our results extend to complex channels as well, for ease of exposition we consider only the real setting here.

4.2.2 Bounded Density Assumption

An important definition for this chapter is the notion of a "bounded density" assumption.

4.2.3 Partial CSIT

Under partial CSIT, the channel coefficients may be represented as

$$G_{ji}^{[nm]}(t) \quad = \quad \hat{G}_{ji}^{[nm]}(t) + \sqrt{P^{-\beta_{ji}}}\tilde{G}_{ji}^{[nm]}(t)$$

where $\hat{G}_{ji}^{[nm]}(t)$ are the channel estimate terms and $\tilde{G}_{ji}^{[nm]}(t)$ are the estimation error terms. To avoid degenerate conditions, the ranges of values are bounded away from zero and infinity as follows, i.e., there exist constants Δ_1, Δ_2 such that

$$0<\Delta_1\leq|G_{ji}^{[nm]}(t)|,\ 0<|\hat{G}_{ji}^{[nm]}(t)|,\ |\tilde{G}_{ji}^{[nm]}(t)|<\Delta_2<\infty$$

. In addition, for each $N_j \times M_i$ channel matrix $\mathbf{G}_{ji}(t)$, we require that all its $\min(N_j, M_i) \times \min(N_j, M_i)$ submatrices are non-singular, i.e., their determinants are bound away from zero. To this end, if $N_j \leq M_i$, then for all $i, j \in \{1, 2\}$, and for all choices of N_j transmit antenna indices $\{m_1, m_2, \cdots, m_{N_j} : m_k \in [1, 2, \cdots, M_i]\}$ define the determinant $D(t)$ as

$$D(t) \triangleq \begin{vmatrix} G_{ji}^{[1m_1]}(t) & G_{ji}^{[1m_2]}(t) & \cdots & G_{ji}^{[1m_{N_j}]}(t) \\ \vdots & \vdots & \ddots & \vdots \\ G_{ji}^{[N_jm_1]}(t) & G_{ji}^{[N_jm_2]}(t) & \cdots & G_{ji}^{[N_jm_{N_j}]}(t) \end{vmatrix} \tag{4.12}$$

If $N_j > M_i$, then for all $i, j \in \{1, 2\}$, and for all choices of M_i receive antenna indices $\{n_1, n_2, \cdots, n_{M_i} : n_k \in [1, 2, \cdots, N_j]\}$ define the determinant $D(t)$ as

$$D(t) \triangleq \begin{vmatrix} G_{ji}^{[n_11]}(t) & G_{ji}^{[n_12]}(t) & \cdots & G_{ji}^{[n_1M_i]}(t) \\ \vdots & \vdots & \ddots & \vdots \\ G_{ji}^{[n_{M_i}1]}(t) & G_{ji}^{[n_{M_i}2]}(t) & \cdots & G_{ji}^{[n_{M_i}M_i]}(t) \end{vmatrix} \tag{4.13}$$

Then we require that there exists a positive constant $\Delta_3 > 0$, such that $|D(t)| \geq \Delta_3$, for all $i, j \in \{1, 2\}$, $\{m_1, m_2, \cdots, m_{N_j} : m_k \in [1, 2, \cdots, M_i]\}$, $\{n_1, n_2, \cdots, n_{M_i} : n_k \in [1, 2, \cdots, N_j]\}$. The channel variables $\hat{G}_{ji}^{[nm]}(t), \tilde{G}_{ji}^{[nm]}(t)$, $\forall i, j \in \{1, 2\}, t \in \mathbb{N}$, are subject to the bounded density assumption with the difference that the actual realizations of $\hat{G}_{ji}^{[nm]}(t)$ are revealed to the transmitter, but the realizations of $\tilde{G}_{ji}^{[nm]}(t)$ are not available to the transmitter.

Note that under the partial CSIT model, the variance of the channel coefficients $G_{ji}^{[nm]}(t)$ behaves as $\sim P^{-\beta_{ji}}$ and the peak of the probability density function behaves as $\sim\sqrt{P^{\beta_{ji}}}$. In order to span the full range of partial channel knowledge at the transmitters, the corresponding range of β_{ji} parameters, assumed throughout this chapter, is $0 \leq \beta_{ji} \leq 1$. Note that $\beta_{ji}=0$ and $\beta_{ji}=1$ correspond to the two extremes where the channel knowledge is essentially absent and perfect, respectively.

4.3 Main Results

Without loss of generality, let us assume that $N_1 \leq N_2$. Further, according to [21, 38, 46, 23], when $M_2 \leq N_1$, the DoF region of two-user MIMO IC with perfect CSIT is identical to the DoF region with no CSIT. Therefore, without loss of generality, we only need to consider $N_1 \leq N_2$ and $N_1 < M_2$.

For the two-user MIMO IC, allow arbitrary channel uncertainty parameters β_{ji} for each channel coefficient. Since $N_1 \leq N_2$ and $N_1 < M_2$, we can categorize the antenna configurations into three cases as follows. The DoF region of each case is characterized in the following theorems.

Theorem 4.1. *If $N_1 < M_2 \leq N_2$, then the DoF region $\mathcal{D}_1$ is as follows*

$$\{ (d_1, d_2) \in \mathbb{R}^{2+},$$

$$L_1: \quad d_1 \leq \min(M_1, N_1), \tag{4.14}$$

$$L_2: \quad d_1 + d_2 \leq M_2, \tag{4.15}$$

$$L_3: \quad \frac{d_1}{\min(M_1,N_1)} + \frac{d_2}{M_2-(N_1-M_1)^+} \leq \frac{M_2+(M_2-N_1)\beta_{12}}{M_2-(N_1-M_1)^+} \} \tag{4.16}$$

Theorem 4.2. *If* $\max(M_1,N_1) \leq N_2 < M_2$*, then the DoF region* $\mathcal{D}_2$ *is as follows*

$$\{ (d_1,d_2) \in \mathbb{R}^{2+},$$

$$L_1: \quad d_1 \leq \min(M_1,N_1), \tag{4.17}$$

$$L_2: \quad d_1+d_2 \leq N_2, \tag{4.18}$$

$$L_3: \quad \frac{d_1}{\min(M_1,N_1)} + \frac{d_2}{N_2-(N_1-M_1)^+} \leq \frac{N_2+\min(M_2-N_1,N_2)\beta_{12}}{N_2-(N_1-M_1)^+}, \tag{4.19}$$

if $N_1+M_1<N_2$*,* (d_1,d_2) *subject to* L_4 *which is defined as*

$$L_4: \quad \frac{d_1}{\min(M_1,N_1)} + \frac{d_2}{N_2-\max(M_1,N_1)+\min(M_1,N_1)} \leq \frac{N_2+(N_2-M_1)\beta_{12}}{N_2-\max(M_1,N_1)+\min(M_1,N_1)}, \tag{4.20}$$

if $N_1+M_1 \geq N_2$*,* (d_1,d_2) *subject to* L_5 *and* L_6 *which are defined as*

$$L_5: \quad d_1+\frac{d_2}{2} \leq \frac{1}{2}[M_1+N_1+(N_2-M_1)\beta_{12}], \tag{4.21}$$

$$L_6: \quad \frac{d_1}{M_1}+\frac{d_2}{N_2+M_1-N_1} \leq \frac{N_2}{N_2+M_1-N_1} + \left[\frac{N_2-M_1}{N_2+M_1-N_1}+\frac{(M_1+N_1-N_2)\min(M_1,M_1+M_2-N_1-N_2)}{M_1(N_2+M_1-N_1)}\right]\beta_{12} \} \tag{4.22}$$

For Theorem 4.1, the achievable DoF in [44] and [20] both match our outer bounds. Theorem 4.2 can be achieved by the scheme based on elevated multiplexing [44], the achievability for Theorem 4.2 is presented in Section 4.11.

Theorem 4.3. *If* $N_1 \leq N_2 < \min(M_1,M_2)$*, then the DoF region* $\mathcal{D}_3$ *is as follows*

$$\{ (d_1,d_2) \in \mathbb{R}^{2+},$$

$$d_1 \leq N_1, \tag{4.23}$$

$$
\begin{aligned}
d_2 &\leq N_2, & (4.24)\\
\frac{d_1}{N_1}+\frac{d_2}{N_2} &\leq \frac{N_2+\min(M_2-N_1,N_2)\beta_{12}}{N_2}, & (4.25)\\
d_1+d_2 &\leq N_2+\min(M_1-N_2,N_1)\beta_{21}, & (4.26)\\
d_1+d_2 &\leq N_2+\min(M_2-N_2,N_1)\beta_{12}, & (4.27)\\
d_1+d_2 &\leq N_2+\min(M_1-N_2,N_1)B\ \} & (4.28)
\end{aligned}
$$

where B is defined as

$$
\begin{cases}
\frac{N_1-N_2+(N_2-N_1)\beta_{21}+\min(M_2-N_1,N_2)\beta_{12}}{\min(M_1-N_1,N_2)}, & \text{if } \beta_{12}+\beta_{21}\geq 1\\
\frac{\min(M_2-N_2,N_1)\beta_{12}\beta_{21}}{(N_2-N_1)(1-\beta_{12})+\min(M_1-N_2,N_1)\beta_{21}}, & \text{if } \beta_{12}+\beta_{21}<1
\end{cases}
$$

Remark 4.1. *For sum-DoF results in Theorem 4.3, one can state* (4.26) (4.27) *and* (4.28) *more compactly as follows*

$$
\begin{aligned}
d_1+d_2 \leq\ & N_2+\min(M_1-N_2,N_1)\beta_{21}\min\left[1,\frac{\min(M_2-N_2,N_1)\beta_{12}}{\min(M_1-N_2,N_1)\beta_{21}},\right.\\
&\left.\frac{(N_2-N_1)(\beta_{21}+\beta_{12}-1)^{+}+\min(M_2-N_2,N_1)\beta_{12}}{(N_2-N_1)(1-\beta_{21}-\beta_{12})^{+}+\min(M_1-N_1,N_2)\beta_{21}}\right] \qquad (4.29)
\end{aligned}
$$

For Theorem 4.3, one can verify that the corner points of the DoF region are identical with the conditions in Proposition 2 in [20], thus it can be achieved by the scheme in [20] which jointly uses the signal levels and time slots.

An interesting observation from Theorem 4.3 is that β_{ji} appears in the denominator of (4.28) when $\beta_{12}+\beta_{21}<1$. This implies that the DoF is not linear with β_{21} and β_{12} in this case. To the best of our knowledge this is the first nonlinear dependence of DoF on SNR exponents. Thus the conventional AIS approach [6, 7, 9]. is not sufficient to obtain this outer bound. This provides us the ideal setting for the first application of the new sum-set inequalities in [8].

In addition, this case is also special from the achievability aspect. As shown in [44, 20], all the results in Theorem 4.1, 4.2 and 4.3 can be described by appropriate signal space partitioning over a single time slot, except when (4.28) is tight and $\beta_{12} + \beta_{21} < 1$. In this case, the new scheme spanning multiple time slots in [20] is necessary.

Since the achievability results follow from the ideas presented in [44, 20], the focus of this chapter is mainly on the DoF outer bounds. We will first provide the proof of the outer bound (4.28) for a representative setting $(M_1, N_1, M_2, N_2, \beta_{12}, \beta_{21}) = (3, 1, 3, 2, \frac{1}{3}, \frac{1}{2})$ which exemplifies all the key arguments of the general proof. The general outer bound proof itself is presented in Section 4.5, 4.6 and 4.7.

Remark 4.2. *Note that our results stay consistent with previous observations in [44]. First, the channel knowledge of desired links is not critical, i.e., DoF do not depend on* β_{11} *and* β_{22}*. Second, only the cross-links with right null space matter, i.e., DoF only depend on* β_{ji}*, if* $M_i > N_j, \forall i, j \in \{1, 2\}, i \neq j$*.*

DoF vs CSIT Budget

Let us use $\mathcal{D}_\Sigma$ to denote the sum-DoF value. Since $\mathcal{D}_\Sigma$ in Theorem 4.3 depends on both β_{12} and β_{21}, it offers insights into the optimal allocation of CSIT resources to maximize the sum-DoF. The CSIT budget formulation depends on the relative costs of acquiring CSIT for each link, which may depend on the feedback mechanism employed. Since β_{11} and β_{22} are irrelevant, as a simple example, suppose the total CSIT budget is

$$\beta = \beta_{21} + \beta_{12}.$$

Then, given the value of β, $0 \leq \beta \leq 2$, it should be optimally allocated among β_{21} and β_{12} in order to maximize the sum-DoF value. This can be done based on (4.26) (4.27) and (4.28). For example, consider the setting where $(M_1, N_1, M_2, N_2) = (3, 1, 3, 2)$, the sum-DoF with the optimal allocation of CSIT are shown in Figure 4.1. As a comparison, the sum-DoF for the case $M_1 \leq 2$ is also illustrated in Figure 4.1.

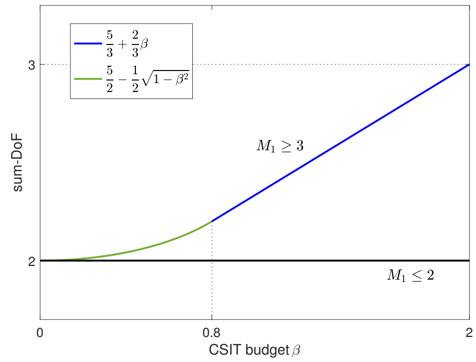

Figure 4.1: Sum-DoF with optimal allocation of CSIT budget β for the case $(N_1, M_2, N_2) = (1, 3, 2)$.

4.4 Proof for $(M_1, N_1, M_2, N_2) = (3, 1, 3, 2)$

In this section, we prove the outer bound for the setting $(\beta_{12}, \beta_{21}) = (\frac{1}{3}, \frac{1}{2})$. Here (4.28) is the critical bound, i.e., $d_1 + d_2 \leq 2 + \frac{1}{7}$. For the outer bound, let us assume that each transmitter has perfect CSIT on its direct link, i.e., $\beta_{11} = \beta_{22} = 1$. This will not reduce the DoF, since a transmitter can always ignore the CSIT.

4.4.1 Equivalent Channel and Deterministic Model

Without loss of generality, for the purpose of deriving a DoF outer bound, we can perform a sequence of invertible operations similar as [16] at transmitters and receivers to convert the channel to its simplest form. First, for Transmitter 1, since it has three antennas, Transmitter 1 can zero-force its first antenna, i.e., X_{11} in Figure 4.2, into the right null space of $\hat{\mathbf{G}}_{21}$. Due to the estimation error $\tilde{\mathbf{G}}_{21}$, the residual interference caused by X_{11} at Receiver 2 has power level $\bar{P}^{-\frac{1}{2}}$. Transmitter 1 can also zero-force its last two antennas, i.e., X_{12}, X_{13}, into the right null space of $\mathbf{G}_{11}$, so that X_{12} and X_{13} will not be heard by Receiver 1.

Similarly for Transmitter 2, it can zero-force its first antenna, i.e., X_{21}, into the right null space of $\mathbf{G}_{22}$, so that X_{21} will not heard by Receiver 2. Transmitter 2 can also zero-force its last two antennas, i.e., X_{22} and X_{23}, into the right null space of $\hat{\mathbf{G}}_{12}$. Then the residual interference caused by X_{22} and X_{23} at Receiver 1 has power level $\bar{P}^{-\frac{1}{3}}$.

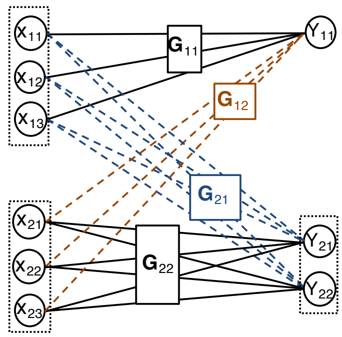

Figure 4.2: Example $(M_1, N_1, M_2, N_2, \beta_{12}, \beta_{21}) = (3, 1, 3, 2, \frac{1}{3}, \frac{1}{2})$

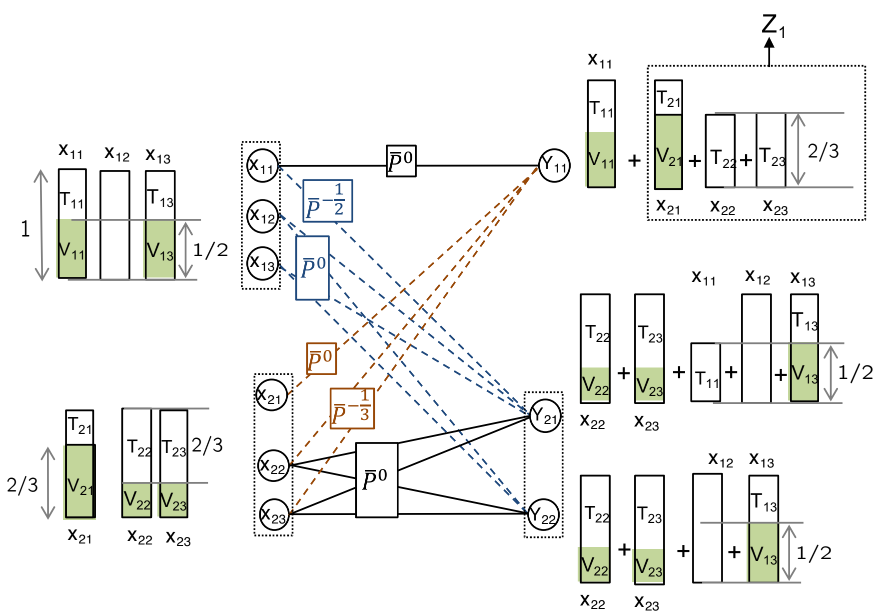

Figure 4.3: Equivalent channel and signal levels at each receiving antenna, where the value on each link is equal to its channel strength level.

Then for Receiver 2, since it has two antennas, Receiver 2 can zero-force the first antenna from Transmitter 1 so that it will not heard by the second antenna at Receiver 2. We now end up with the following simple channel model (see Figure 4.3),

$$Y_{11}(t)=G_{11}^{[11]}(t)X_{11}(t)+G_{12}^{[11]}(t)X_{21}(t)+\bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[12]}(t)X_{22}(t)+\bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[13]}(t)X_{23}(t)+\Gamma_{11}(t) \tag{4.30}$$

$$Y_{21}(t)=\bar{P}^{-\frac{1}{2}}\tilde{G}_{21}^{[11]}(t)X_{11}(t)+G_{21}^{[12]}(t)X_{12}(t)+G_{21}^{[13]}(t)X_{13}(t)+G_{22}^{[12]}(t)X_{22}(t)$$
$$+G_{22}^{[13]}(t)X_{23}(t)+\Gamma_{21}(t) \tag{4.31}$$

$$Y_{22}(t)=G_{21}^{[22]}(t)X_{12}(t)+G_{21}^{[23]}(t)X_{13}(t)+G_{22}^{[22]}(t)X_{22}(t)+G_{22}^{[23]}(t)X_{23}(t)+\Gamma_{22}(t) \tag{4.32}$$

Now following the AIS framework of [6] the deterministic equivalent channel (whose capacity region is within a bounded gap from the original channel) is obtained as

$$\bar{Y}_{11}(t)=\lfloor G_{11}^{[11]}(t)\bar{X}_{11}(t)\rfloor+\lfloor G_{12}^{[11]}(t)\bar{X}_{21}(t)\rfloor+\lfloor \bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[12]}(t)\bar{X}_{22}(t)\rfloor+\lfloor \bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[13]}(t)\bar{X}_{23}(t)\rfloor \tag{4.33}$$

$$\bar{Y}_{21}(t)=\lfloor \bar{P}^{-\frac{1}{2}}\tilde{G}_{21}^{[11]}(t)\bar{X}_{11}(t)\rfloor+\lfloor G_{21}^{[12]}(t)\bar{X}_{12}(t)\rfloor+\lfloor G_{21}^{[13]}(t)\bar{X}_{13}(t)\rfloor$$
$$+\lfloor G_{22}^{[12]}(t)\bar{X}_{22}(t)\rfloor+\lfloor G_{22}^{[13]}(t)\bar{X}_{23}(t)\rfloor \tag{4.34}$$

$$\bar{Y}_{22}(t)=\lfloor G_{21}^{[22]}(t)\bar{X}_{12}(t)\rfloor+\lfloor G_{21}^{[23]}(t)\bar{X}_{13}(t)\rfloor+\lfloor G_{22}^{[22]}(t)\bar{X}_{22}(t)\rfloor+\lfloor G_{22}^{[23]}(t)\bar{X}_{23}(t)\rfloor \tag{4.35}$$

and $\bar{X}_{ji}(t)\in\{0,1,\cdots,\lfloor\bar{P}\rfloor\}$, $\forall i\in\{1,2,3\}, j\in\{1,2\}$.

Let us define $Z_1(t)$, $Z_{21}(t)$ and $Z_{22}(t)$ as the interference term at each receiving antenna, i.e.,

$$Z_1(t)=\lfloor G_{12}^{[11]}(t)\bar{X}_{21}(t)\rfloor+\lfloor \bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[12]}(t)\bar{X}_{22}(t)\rfloor+\lfloor \bar{P}^{-\frac{1}{3}}\tilde{G}_{12}^{[13]}(t)\bar{X}_{23}(t)\rfloor \tag{4.36}$$

$$Z_{21}(t)=\lfloor \bar{P}^{-\frac{1}{2}}\tilde{G}_{21}^{[11]}(t)\bar{X}_{11}(t)\rfloor+\lfloor G_{21}^{[12]}(t)\bar{X}_{12}(t)\rfloor+\lfloor G_{21}^{[13]}(t)\bar{X}_{13}(t)\rfloor \tag{4.37}$$

$$Z_{22}(t)=\lfloor G_{21}^{[22]}(t)\bar{X}_{12}(t)\rfloor+\lfloor G_{21}^{[23]}(t)\bar{X}_{13}(t)\rfloor \tag{4.38}$$

Also as shown in Figure 4.3, we define $T_{1i}(t)=(\bar{X}_{1i}(t))^{\frac{1}{2}}$, $V_{1i}(t)=(\bar{X}_{1i}(t))_{\frac{1}{2}}$, $T_{21}(t)=(\bar{X}_{21}(t))^{\frac{1}{3}}$,

$V_{21}(t)=(\bar{X}_{21}(t))_{\frac{2}{3}}$, $T_{2j}(t)=(\bar{X}_{2j}(t))^{\frac{2}{3}}$ and $V_{2j}(t)=(\bar{X}_{2j}(t))_{\frac{1}{3}}$, $\forall\ i \in \{1,3\}$, $\forall\ j \in \{2,3\}$.

4.4.2 Outer Bound

For Receiver 2, we have

$$
\begin{aligned}
n(R_2-\epsilon) &\leq I(\bar{X}_{22}^n, \bar{X}_{23}^n; \bar{Y}_{21}^n, \bar{Y}_{22}^n) && (4.39)\\
&= H(\bar{Y}_{21}^n, \bar{Y}_{22}^n) - H(\bar{Y}_{21}^n, \bar{Y}_{22}^n \mid \bar{X}_{22}^n, \bar{X}_{23}^n) && (4.40)\\
&\leq H(\bar{Y}_{21}^n, \bar{Y}_{22}^n) - H(Z_{21}^n, Z_{22}^n) && (4.41)\\
&\leq 2n \log \bar{P} - H(T_{11}^n, \bar{X}_{12}^n, T_{13}^n) + n\ o(\log \bar{P}) && (4.42)
\end{aligned}
$$

(4.42) is true as from Theorem 4 in [8] we have $H(Z_{21}^n, Z_{22}^n) \geq H(T_{11}^n, \bar{X}_{12}^n, T_{13}^n)$.[1] So we have

$$
n(R_2-\epsilon)+H(T_{11}^n, \bar{X}_{12}^n, T_{13}^n) \leq 2n \log \bar{P} + no(\log \bar{P}). \quad (4.43)
$$

For Receiver 1, we have

$$
\begin{aligned}
n(R_1-\epsilon) &\leq I(\bar{X}_{11}^n; \bar{Y}_1^n) = H(\bar{Y}_1^n) - H(\bar{Y}_1^n \mid \bar{X}_{11}^n) && (4.44)\\
&\leq n \log \bar{P} - H(Z_1^n) + n\ o(\log \bar{P}) && (4.45)
\end{aligned}
$$

Note that for Receiver 1,

$$
\begin{aligned}
\bar{X}_{11}^n &= (T_{11}^n, V_{11}^n) && (4.46)\\
H(\bar{X}_{11}^n) &= I(\bar{X}_{11}^n; \bar{Y}_1^n) && (4.47)
\end{aligned}
$$

So, replacing (4.46) in (4.47), we have

$$
H(T_{11}^n)+H(V_{11}^n \mid T_{11}^n) = I(T_{11}^n; \bar{Y}_1^n)+I(V_{11}^n; \bar{Y}_1^n \mid T_{11}^n)
$$

[1] This is proved in details in proof of Theorem 4.3.

$$H(V_{11}^n \mid T_{11}^n) = I(V_{11}^n; \bar{Y}_1^n \mid T_{11}^n) \tag{4.48}$$

$$= I(V_{11}^n; (\bar{Y}_1^n)^{\frac{1}{2}}, (\bar{Y}_1^n)_{\frac{1}{2}} \mid T_{11}^n) \tag{4.49}$$

$$= I(V_{11}^n; (\bar{Y}_1^n)^{\frac{1}{2}} \mid T_{11}^n) + I(V_{11}^n; (\bar{Y}_1^n)_{\frac{1}{2}} \mid T_{11}^n, (\bar{Y}_1^n)^{\frac{1}{2}}) \tag{4.50}$$

$$= H((\bar{Y}_1^n)_{\frac{1}{2}} \mid T_{11}^n, (\bar{Y}_1^n)^{\frac{1}{2}}) - H((\bar{Y}_1^n)_{\frac{1}{2}} \mid \bar{X}_{11}^n, (\bar{Y}_1^n)^{\frac{1}{2}}) \tag{4.51}$$

$$\leq \frac{1}{2} n \log \bar{P} - H((Z_1^n)_{\frac{1}{2}} \mid (Z_1^n)^{\frac{1}{2}}) + n\ o(\log \bar{P}) \tag{4.52}$$

(4.48) is because $H(T_{11}^n) = I(T_{11}^n; \bar{Y}_1^n)$. (4.51) is because given T_{11}^n, V_{11}^n and $(\bar{Y}_1^n)^{\frac{1}{2}}$ are independent, i.e., $I(V_{11}^n; (\bar{Y}_1^n)^{\frac{1}{2}} \mid T_{11}^n) = 0$. Scaling equation (4.45) by 3 and equation (4.52) by 4 and then summing up, we have

$$\begin{aligned} &3n(R_1 - \epsilon) + 4H(V_{11}^n \mid T_{11}^n) \\ \leq &3n \log \bar{P} - 3H(Z_1^n) + 2n \log \bar{P} - 4H((Z_1^n)_{\frac{1}{2}} \mid (Z_1^n)^{\frac{1}{2}}) \\ &+ n\ o(\log \bar{P}) \end{aligned} \tag{4.53}$$

$$\leq 5n \log \bar{P} - 3H(Z_1^n) - 4H((Z_1^n)_{\frac{1}{2}} | (Z_1^n)^{\frac{1}{2}}) + no(\log \bar{P}) \tag{4.54}$$

$$\leq 5n \log \bar{P} - 3H(T_{21}^n, T_{22}^n, T_{23}^n) + no(\log \bar{P}) \tag{4.55}$$

(4.55) is because according to Lemma 1 of [8],

$$3H(T_{21}^n, T_{22}^n, T_{23}^n) \leq 3H(Z_1^n) + 4H((Z_1^n)_{\frac{1}{2}} \mid (Z_1^n)^{\frac{1}{2}})$$

Then we have

$$\begin{aligned} &3n(R_1 - \epsilon) + 4H(V_{11}^n \mid T_{11}^n) + 3H(T_{21}^n, T_{22}^n, T_{23}^n) \\ &\leq 5n \log \bar{P} + n\ o(\log \bar{P}) \end{aligned} \tag{4.56}$$

Note that,

$$H(V_{22}^n, V_{23}^n) \leq \frac{2}{3} n \log \bar{P} + n\ o(\log \bar{P}). \tag{4.57}$$

Therefore,

$$\begin{aligned}
&7nR_1 + 7nR_2 \\
\leq &3nR_1 + 4H(\bar{X}_{11}^n) + 4nR_2 + 3H(\bar{X}_{22}^n, \bar{X}_{23}^n) && (4.58) \\
\leq &3nR_1 + 4H(\bar{X}_{11}^n, \bar{X}_{12}^n, T_{13}^n) + 4nR_2 + 3H(T_{21}^n, \bar{X}_{22}^n, \bar{X}_{23}^n) && (4.59) \\
\leq &3nR_1 + 4H(T_{11}^n, \bar{X}_{12}^n, T_{13}^n) + 4H(V_{11}^n \mid T_{11}^n) + 4nR_2 \\
&+ 3H(T_{21}^n, T_{22}^n, T_{23}^n) + 3H(V_{22}^n, V_{23}^n) && (4.60) \\
\leq &15n \log \bar{P} + n\ o(\log \bar{P}) && (4.61)
\end{aligned}$$

where the equation (4.60) is obtained by scaling equation (4.43) by 4 and equation (4.57) by 3 and then summing up with (4.56). Normalizing by $n \log(\bar{P})$ and taking limits, first with $n \to \infty$ and then $\bar{P} \to \infty$, we obtain the desired bound, $d_1 + d_2 \leq 2 + \frac{1}{7}$. ■

4.5 Proof for Theorem 4.1

In this section, we prove the outer bounds for Theorem 4.1. Note that the outer bounds (4.14) and (4.15) are trivial, i.e., (4.14) is the single user bound, (4.15) is the outer bound with perfect CSIT. We will prove the following bound for the two-user MIMO IC with $N_1 \leq N_2$ and $N_1 < M_2$,

$$\frac{d_1}{\min(M_1, N_1)} + \frac{d_2}{\min(M_2, N_2) - (N_1 - M_1)^+} \leq \frac{\min(M_2, N_2) + \min(M_2 - N_1, N_2)\beta_{12}}{\min(M_2, N_2) - (N_1 - M_1)^+} \tag{4.62}$$

This bound is equivalent to (4.16) in Theorem 4.1, (4.19) in Theorem 4.2 and (4.25) in Theorem 4.3. For the outer bound (4.62), we will present the proof only for the real setting here, extension to complex settings follows along the lines of similar extensions in [6, 9]. It is assumed that each transmitter has perfect CSIT on its direct link. This will not reduce the DoF, since transmitter can always ignore the CSIT.

4.5.1 Equivalent Channel

Without loss of generality, for the purpose of deriving a DoF outer bound, we can perform a sequence of invertible operations similar as [16] at transmitters and receivers to convert the channel to its simplest form. Specifically, at Transmitter 2, we aim to obtain an equivalent channel where the inputs are partitioned as $\mathbf{X}_2(t) = [\mathbf{X}_{2a}^T(t)\ \mathbf{X}_{2b}^T(t)\ \mathbf{X}_{2c}^T(t)\ \mathbf{X}_{2d}^T(t)]^T$. We have

$$\mathbf{X}_{2a}(t) = \left[X_2^1(t) \quad X_2^2(t) \quad \cdots \quad X_2^{\min(N_1,(M_2-N_2)^+)}(t)\right]^T \tag{4.63}$$

$$\mathbf{X}_{2b}(t) = \left[X_2^{\min(N_1,(M_2-N_2)^+)+1}(t) \quad \cdots \quad X_2^{N_1}(t)\right]^T \tag{4.64}$$

$$\mathbf{X}_{2c}(t) = \left[X_2^{N_1+1}(t) \quad \cdots \quad X_2^{N_1+(M_2-N_1-N_2)^+}(t)\right]^T \tag{4.65}$$

$$\mathbf{X}_{2d}(t) = \left[X_2^{N_1+(M_2-N_1-N_2)^++1}(t) \quad \cdots \quad X_2^{M_2}(t)\right]^T \tag{4.66}$$

$|\mathbf{X}_{2a}(t)| = \min(N_1,(M_2-N_2)^+)$, $|\mathbf{X}_{2b}(t)| = (N_1-(M_2-N_2)^+)^+$, $|\mathbf{X}_{2c}(t)| = (M_2-N_1-N_2)^+$, $|\mathbf{X}_{2d}(t)| = \min(M_2-N_1, N_2)$. Under the generic channel coefficient assumption, the matrices $\mathbf{G}_{22}(t)$ and $\hat{\mathbf{G}}_{12}(t)$ have $(M_2-N_2)^+$ and M_2-N_1 dimension right null space, respectively, almost surely. Note that there is a $(M_2-N_2-N_1)^+$ dimension intersection space of these two null spaces. Thus Transmitter 2 can zero-force $\mathbf{X}_{2c}(t)$ into this intersection null space, so that $\mathbf{X}_{2c}(t)$ cannot be heard by Receiver 2. Due to the estimation error $\tilde{\mathbf{G}}_{12}$, the residual interference caused by $\mathbf{X}_{2c}(t)$ at Receiver 1 has power level $\bar{P}^{-\beta_{12}}$.

Then Transmitter 2 can also zero-force $\mathbf{X}_{2a}(t)$ and $\mathbf{X}_{2d}(t)$ into the null space of $\mathbf{G}_{22}(t)$ and $\hat{\mathbf{G}}_{12}(t)$, respectively. So that $\mathbf{X}_{2a}(t)$ cannot be heard by Receiver 2, $\mathbf{X}_{2d}(t)$ is heard by Receiver 1 only with power level $\bar{P}^{-\beta_{12}}$. $\mathbf{X}_{2b}(t)$ can be heard by either of the receiver. Note that $\mathbf{X}_{2b}(t)$ and $\mathbf{X}_{2c}(t)$ cannot exist simultaneously, i.e., $\mathbf{X}_{2c}(t)$ do not exist when $N_1{+}N_2{-}M_2 \geq 0$ and $\mathbf{X}_{2b}(t)$ do not exist when $N_1{+}N_2{-}M_2 \leq 0$.

Then for Receiver 1, we aim to obtain an equivalent channel where the outputs are partitioned as $\mathbf{Y}'_1(t) = \begin{bmatrix}\mathbf{Y}'^T_{1a}(t) & \mathbf{Y}'^T_{1b}(t)\end{bmatrix}^T$. We have

$$\mathbf{Y}'_{1a}(t) = \begin{bmatrix} Y'^1_1(t) & Y'^2_1(t) & \cdots & Y'^{\min(M_1,N_1)}_1(t)\end{bmatrix}^T \tag{4.67}$$

$$\mathbf{Y}'_{1b}(t) = \begin{bmatrix} Y'^{\min(M_1,N_1)+1}_1(t) & \cdots & Y'^{N_1}_1(t)\end{bmatrix}^T \tag{4.68}$$

Under the generic channel coefficient assumption, the matrices $\mathbf{G}_{11}(t)$ have $(N_1{-}M_1)^+$ dimension left null space almost surely. Thus Receiver 1 can zero-force signals from Transmitter 1 into this null space, so that its last $(N_1{-}M_1)^+$ antennas i.e., $\mathbf{Y}'_{1b}(t)$, cannot hear $\mathbf{X}_1(t)$. We now end up with the following simple channel model,

$$\mathbf{Y}'_1(t) = \begin{bmatrix}\mathbf{Y}'_{1a}(t)\\ \mathbf{Y}'_{1b}(t)\end{bmatrix} \tag{4.69}$$

$$\mathbf{Y}'_{1a}(t) = \mathbf{G}'_{11}(t)\mathbf{X}_1(t) + \mathbf{G}'_{121}(t)\begin{bmatrix}\mathbf{X}_{2a}(t)\\ \mathbf{X}_{2b}(t)\end{bmatrix} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}'_{121}(t)\begin{bmatrix}\mathbf{X}_{2c}(t)\\ \mathbf{X}_{2d}(t)\end{bmatrix} + \mathbf{\Gamma}_{11}(t) \tag{4.70}$$

$$\mathbf{Y}'_{1b}(t) = \mathbf{G}'_{122}(t)\begin{bmatrix}\mathbf{X}_{2a}(t)\\ \mathbf{X}_{2b}(t)\end{bmatrix} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}'_{122}(t)\begin{bmatrix}\mathbf{X}_{2c}(t)\\ \mathbf{X}_{2d}(t)\end{bmatrix} + \mathbf{\Gamma}_{12}(t) \tag{4.71}$$

$$\mathbf{Y}'_2(t) = \mathbf{G}'_{21}(t)\mathbf{X}_1(t) + \mathbf{G}'_{22}(t)\begin{bmatrix}\mathbf{X}_{2b}(t)\\ \mathbf{X}_{2d}(t)\end{bmatrix} + \mathbf{\Gamma}_2(t) \tag{4.72}$$

where $\mathbf{G}'_{121}(t)$ and $\tilde{\mathbf{G}}'_{121}(t)$ has size $\min(M_1,N_1)\times N_1$ and $\min(M_1,N_1)\times(M_2-N_1)$, respectively. $\mathbf{G}'_{122}(t)$ and $\tilde{\mathbf{G}}'_{122}(t)$ has size $(N_1{-}M_1)^+\times N_1$ and $(N_1{-}M_1)^+\times(M_2{-}N_1)$, respectively. In addition, the partial CSIT quality of $\mathbf{G}'_{121}(t)$ and $\mathbf{G}'_{122}(t)$ is equal to β_{12}. $\tilde{\mathbf{G}}'_{121}(t)$ and $\tilde{\mathbf{G}}'_{122}(t)$ are known by both receivers only to finite precision.

$\mathbf{G}'_{22}(t)$ has size $N_2 \times \min(N_2, M_2)$ and the same partial CSIT quality with $\mathbf{G}_{22}(t)$. $\mathbf{G}'_{11}(t)$ has size $\min(M_1, N_1) \times M_1$ and the same partial CSIT quality with $\mathbf{G}_{11}(t)$. $\mathbf{G}'_{21}(t)$ has the same size and the same partial CSIT quality with $\mathbf{G}_{21}(t)$.

4.5.2 Deterministic Model

Based on the equivalent channel in Subsection 4.5.1, we define the deterministic channel model with input $\bar{\mathbf{X}}_k(t) = [\bar{X}_k^1(t)\ \bar{X}_k^2(t)\ \cdots\ \bar{X}_k^{M_k}(t)]^T, \forall\, k \in \{1, 2\}$, such that

$$\bar{\mathbf{Y}}_1(t) = \begin{bmatrix} \bar{\mathbf{Y}}_{1a}(t) \\ \bar{\mathbf{Y}}_{1b}(t) \end{bmatrix} \tag{4.73}$$

$$\bar{\mathbf{Y}}_{1a}(t) = \left\lfloor \mathbf{G}'_{11}(t)\bar{\mathbf{X}}_1(t) \right\rfloor + \left\lfloor \mathbf{G}'_{121}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2a}(t) \\ \bar{\mathbf{X}}_{2b}(t) \end{bmatrix} \right\rfloor + \left\lfloor \tilde{\mathbf{G}}'_{121}(t) \begin{bmatrix} (\bar{\mathbf{X}}_{2c}(t))^{1-\beta_{12}} \\ (\bar{\mathbf{X}}_{2d}(t))^{1-\beta_{12}} \end{bmatrix} \right\rfloor \tag{4.74}$$

$$\bar{\mathbf{Y}}_{1b}(t) = \left\lfloor \mathbf{G}'_{122}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2a}(t) \\ \bar{\mathbf{X}}_{2b}(t) \end{bmatrix} \right\rfloor + \left\lfloor \tilde{\mathbf{G}}'_{122}(t) \begin{bmatrix} (\bar{\mathbf{X}}_{2c}(t))^{1-\beta_{12}} \\ (\bar{\mathbf{X}}_{2d}(t))^{1-\beta_{12}} \end{bmatrix} \right\rfloor \tag{4.75}$$

$$\bar{\mathbf{Y}}_2(t) = \left\lfloor \mathbf{G}'_{21}(t)\bar{\mathbf{X}}_1(t) \right\rfloor + \left\lfloor \mathbf{G}'_{22}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2b}(t) \\ \bar{\mathbf{X}}_{2d}(t) \end{bmatrix} \right\rfloor \tag{4.76}$$

and $\bar{X}_k^i(t) \in \{0, 1, \cdots, \lfloor \bar{P} \rfloor\}$, $\forall i \in \mathcal{I}_{M_k}$. The size of $\bar{\mathbf{X}}_{2a}(t)$, $\bar{\mathbf{X}}_{2b}(t)$, $\bar{\mathbf{X}}_{2c}(t)$, $\bar{\mathbf{X}}_{2d}(t)$ is the same with $\mathbf{X}_{2a}(t)$, $\mathbf{X}_{2b}(t)$, $\mathbf{X}_{2c}(t)$, $\mathbf{X}_{2d}(t)$. Note that $(\bar{\mathbf{X}}_{2c}(t))_{\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d}(t))_{\beta_{12}}$ can be omitted since they are under noise floor at Receiver 1.

4.5.3 Proof for Outer Bound (4.62)

Note that $o(\log(P))$ and $o(n)$ terms are omitted in the rest of this chapter since they are inconsequential for DoF. For Receiver 1, we have

$$nR_1$$

$$\le H(\bar{\mathbf{Y}}_{1a}^n, \bar{\mathbf{Y}}_{1b}^n) - H(\bar{\mathbf{Y}}_{1a}^n, \bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \tag{4.77}$$

$$
\begin{aligned}
&= H(\bar{\mathbf{Y}}_{1b}^n) + H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n) - H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) - H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) && (4.78)\\
&= H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n) - H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) && (4.79)\\
&\leq n \min(M_1, N_1) \log \bar{P} - H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) && (4.80)\\
&= n \min(M_1, N_1) \log \bar{P} - H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n)\\
&\quad + \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.81)\\
&\leq n \min(M_1, N_1) \log \bar{P} - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})\\
&\quad + \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} (n N_1 \log \bar{P} - n R_1) && (4.82)\\
&= \frac{\min(M_1, N_1) \min(M_2, N_2)}{\min(M_2, N_2) - N_1} n \log \bar{P} - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})\\
&\quad - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} n R_1 - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}})\\
&\quad + \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) && (4.83)\\
&\leq \frac{\min(M_1, N_1) \min(M_2, N_2)}{\min(M_2, N_2) - N_1} n \log \bar{P} - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} n R_2 - \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} n R_1\\
&\quad + \frac{\min(M_1, N_1)}{\min(M_2, N_2) - N_1} \beta_{12} \min(M_2 - N_1, N_2) n \log \bar{P} && (4.84)
\end{aligned}
$$

where (4.79) is true as $\bar{\mathbf{Y}}_{1b}^n$ and $\bar{\mathbf{X}}_1^n$ are independent, i.e., $H(\bar{\mathbf{Y}}_{1b}^n) - H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) = 0$. (4.82) is concluded as for Receiver 1, we have

$$
\begin{aligned}
n R_1 &\leq I(\bar{\mathbf{X}}_1^n; \bar{\mathbf{Y}}_1^n) = H(\bar{\mathbf{Y}}_1^n) - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.85)\\
&\leq n N_1 \log \bar{P} - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.86)
\end{aligned}
$$

Thus $H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \leq n N_1 \log \bar{P} - n R_1$. Now, consider the following lemma.

Lemma 4.1. *For for the two-user deterministic MIMO IC defined in Subsection 4.5.2 with* $N_1 \leq N_2$ *and* $N_1 < M_2$*, we have,*

$$
H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) \leq H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + \frac{\min(M_2, N_2) - N_1}{\min(M_1, N_1)} H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) \qquad (4.87)
$$

See Section 4.8 for proof of Lemma 4.1. (4.84) is true as $nR_2 \leq H(\bar{\mathbf{X}}_{2b}^n, \bar{\mathbf{X}}_{2d}^n)$ and $H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) \leq n\beta_{12} \min(M_2 - N_1, N_2) \log \bar{P}$. Therefore, we have

$$\begin{aligned} & n[\min(M_2, N_2) - (N_1 - M_1)^+]R_1 + n\min(M_1, N_1)R_2 \\ \leq & n\min(M_1, N_1)\min(M_2, N_2)\log\bar{P} + n\beta_{12}\min(M_1, N_1)\min(M_2 - N_1, N_2)\log\bar{P} \end{aligned} \tag{4.88}$$

This implies that

$$\begin{aligned} & [\min(M_2, N_2) - (N_1 - M_1)^+]d_1 + \min(M_1, N_1)d_2 \\ \leq & \min(M_1, N_1)\min(M_2, N_2) + \beta_{12}\min(M_1, N_1)\min(M_2 - N_1, N_2). \end{aligned} \tag{4.89}$$

This concludes the proof for Theorem 4.1. We next prove the remaining bounds in Theorem 4.2.

4.6 Proof for Theorem 4.2: Outer Bound

Note that the outer bounds (4.17) and (4.18) are trivial, i.e., (4.17) is the single user bound, (4.18) is the outer bound with perfect CSIT. (4.19) is proved in Section 4.5. Thus in this subsection, we prove the outer bound (4.20), (4.21) and (4.22) in Theorem 4.2.

4.6.1 Equivalent Channel

In addition to the equivalent channel in Section 4.5.1, for the channel with $\max(M_1, N_1) \leq N_2 \leq M_2$, we can further simplify the channel outputs at Receiver 2 so that they are partitioned as $\mathbf{Y}_2'(t) = \begin{bmatrix} \mathbf{Y}_{2a}'^T(t) & \mathbf{Y}_{2b}'^T(t) \end{bmatrix}^T$. We have

$$\mathbf{Y}_{2a}'(t) = \begin{bmatrix} Y_2'^1(t) & Y_2'^2(t) & \cdots & Y_2'^{M_1}(t) \end{bmatrix}^T \tag{4.90}$$

$$\mathbf{Y}_{2b}'(t) = \begin{bmatrix} Y_2'^{M_1+1}(t) & \cdots & Y_2'^{N_2}(t) \end{bmatrix}^T \tag{4.91}$$

Under the generic channel coefficient assumption, the matrices $\mathbf{G}_{21}(t)$ have N_2-M_1 dimension left null space almost surely. Thus Receiver 2 can zero-force the signal from Transmitter 1 into this null space, so that its last $N_2 - M_1$ antennas, i.e., $\mathbf{Y}'_{2b}(t)$, cannot hear $\mathbf{X}_1(t)$. We now end up with the following simple channel model at Receiver 2,

$$\mathbf{Y}'_2(t) = \begin{bmatrix} \mathbf{Y}'_{2a}(t) \\ \mathbf{Y}'_{2b}(t) \end{bmatrix} \tag{4.92}$$

$$\mathbf{Y}'_{2a}(t) = \mathbf{G}'_{211}(t)\mathbf{X}_1(t) + \mathbf{G}'_{221}(t) \begin{bmatrix} \mathbf{X}_{2b}(t) \\ \mathbf{X}_{2d}(t) \end{bmatrix} + \mathbf{\Gamma}'_{21}(t) \tag{4.93}$$

$$\mathbf{Y}'_{2b}(t) = \mathbf{G}'_{222}(t) \begin{bmatrix} \mathbf{X}_{2b}(t) \\ \mathbf{X}_{2d}(t) \end{bmatrix} + \mathbf{\Gamma}'_{22}(t) \tag{4.94}$$

where $\mathbf{G}'_{211}(t)$ has size $M_1 \times M_1$ and the partial CSIT quality is equal to β_{21}. $\mathbf{G}'_{221}(t)$ has size $M_1 \times \min(M_2, N_2)$ and the partial CSIT quality is equal to β_{22}. $\mathbf{G}'_{222}(t)$ has size $(N_2 - M_1) \times \min(M_2, N_2)$and the partial CSIT quality is equal to β_{22}. The received signal at Receiver 1 is the same as Subsection 4.5.1.

4.6.2 Deterministic Model

For the deterministic channel model, the received signal at Receiver 2 is now as follows

$$\bar{\mathbf{Y}}_2(t) = \begin{bmatrix} \bar{\mathbf{Y}}_{2a}(t) \\ \bar{\mathbf{Y}}_{2b}(t) \end{bmatrix} \tag{4.95}$$

$$\bar{\mathbf{Y}}_{2a}(t) = \left\lfloor \mathbf{G}'_{211}(t)\bar{\mathbf{X}}_1(t) \right\rfloor + \left\lfloor \mathbf{G}'_{221}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2b}(t) \\ \bar{\mathbf{X}}_{2d}(t) \end{bmatrix} \right\rfloor \tag{4.96}$$

$$\bar{\mathbf{Y}}_{2b}(t) = \left\lfloor \mathbf{G}'_{222}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2b}(t) \\ \bar{\mathbf{X}}_{2d}(t) \end{bmatrix} \right\rfloor \tag{4.97}$$

The input signals and $\bar{\mathbf{Y}}_1(t)$ are still the same with Subsection 4.5.2.

4.6.3 Proof for Outer Bound (4.20)

In this subsection, we prove the outer bound (4.20) in Theorem 4.2 for the channel with $\max(M_1, N_1) \leq N_2 \leq M_2$ and $N_1 + M_1 < N_2$. For Receiver 2, we have

$$
\begin{aligned}
& nR_2 \\
\leq & H(\bar{\mathbf{Y}}_{2a}^n, \bar{\mathbf{Y}}_{2b}^n) - H(\bar{\mathbf{Y}}_{2a}^n, \bar{\mathbf{Y}}_{2b}^n \mid \bar{\mathbf{X}}_2^n) && (4.98) \\
= & H(\bar{\mathbf{Y}}_{2b}^n) + H(\bar{\mathbf{Y}}_{2a}^n \mid \bar{\mathbf{Y}}_{2b}^n) - H(\bar{\mathbf{Y}}_{2b}^n \mid \bar{\mathbf{X}}_2^n) - H(\bar{\mathbf{Y}}_{2a}^n \mid \bar{\mathbf{Y}}_{2b}^n, \bar{\mathbf{X}}_2^n) && (4.99) \\
\leq & H(\bar{\mathbf{Y}}_{2a}^n \mid \bar{\mathbf{Y}}_{2b}^n) + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}, (\bar{\mathbf{Y}}_{2b}^n)_{\beta_{12}}) - H(\bar{\mathbf{Y}}_{2a}^n \mid \bar{\mathbf{Y}}_{2b}^n, \bar{\mathbf{X}}_2^n) && (4.100) \\
\leq & nM_1 \log \bar{P} + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) + H((\bar{\mathbf{Y}}_{2b}^n)_{\beta_{12}} \mid (\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - H(\bar{\mathbf{X}}_1^n) && (4.101) \\
\leq & n[M_1 + (N_2 - M_1)\beta_{12}] \log \bar{P} + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - nR_1 && (4.102) \\
= & n[M_1 + (N_2 - M_1)\beta_{12}] \log \bar{P} + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - nR_1 - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \\
& - \frac{N_2 - M_1 - N_1}{\min(M_1, N_1)} H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) + \frac{N_2 - M_1 - N_1}{\min(M_1, N_1)} H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) && (4.103) \\
\leq & n[M_1 + (N_2 - M_1)\beta_{12}] \log \bar{P} - nR_1 + (nN_1 \log \bar{P} - nR_1) \\
& + \frac{N_2 - M_1 - N_1}{\min(M_1, N_1)} (n \min(M_1, N_1) \log \bar{P} - nR_1) && (4.104) \\
= & n[N_2 + (N_2 - M_1)\beta_{12}] \log \bar{P} - 2nR_1 - \frac{N_2 - M_1 - N_1}{\min(M_1, N_1)} nR_1 && (4.105)
\end{aligned}
$$

(4.100) is true as $H(\bar{\mathbf{Y}}_{2b}^n \mid \bar{\mathbf{X}}_2^n) = 0$. (4.101) follows as $\bar{\mathbf{Y}}_{2a}^n$ has M_1 antennas, given $\bar{\mathbf{X}}_2^n$, one can decode $\bar{\mathbf{X}}_1^n$ from $\bar{\mathbf{Y}}_{2a}^n$. Finally, $H((\bar{\mathbf{Y}}_{2b}^n)_{\beta_{12}} \mid (\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) \leq n(N_2 - M_1)\beta_{12} \log \bar{P}$ results in (4.102). (4.104) is concluded from (4.80), (4.86), $H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \leq nN_1 \log \bar{P} - nR_1$ and $H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) \leq n \min(M_1, N_1) \log \bar{P} - nR_1$.

Lemma 4.2. *For for the two-user deterministic MIMO IC defined in Subsection 4.6.2 with* $\max(M_1, N_1) \leq N_2 \leq M_2$ *and* $N_1 + M_1 < N_2$*, we have,*

$$
H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) \leq H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + \frac{N_2 - M_1 - N_1}{\min(M_1, N_1)} H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n) \tag{4.106}
$$

See Section 4.9 for proof of Lemma 4.2. Thus, we have

$$\begin{aligned}&n(N_2 - N_1 - M_1 + 2\min(M_1, N_1))R_1 + n\min(M_1, N_1)R_2\\ \leq& n\min(M_1, N_1)[N_2 + (N_2 - M_1)\beta_{12}]\log \bar{P} \end{aligned} \tag{4.107}$$

This implies that

$$\begin{aligned}&[N_2 - \max(M_1, N_1) + \min(M_1, N_1)]d_1 + \min(M_1, N_1)d_2 \leq \\ &\qquad \min(M_1, N_1)[N_2 + (N_2 - M_1)\beta_{12}]. \end{aligned} \tag{4.108}$$

4.6.4 Proof for Outer Bound (4.21)

In this subsection, we prove the outer bound (4.21) in Theorem 4.2 for the channel with $\max(M_1, N_1) \leq N_2 \leq M_2$ and $N_1 + M_1 \geq N_2$. By (4.86) + (4.102), we have

$$\begin{aligned}&n2R_1 + nR_2\\ \leq& n(N_1 + M_1 + (N_2 - M_1)\beta_{12})\log \bar{P} + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) &(4.109)\\ \leq& n(N_1 + M_1 + (N_2 - M_1)\beta_{12})\log \bar{P} &(4.110)\end{aligned}$$

(4.110) can be concluded from Theorem 4 of [8]. First of all, let us present Theorem 4 of [8].

Theorem 4.4. ([8], Theorem 4) *Consider KM non-negative numbers $\{\lambda_{km} : k \in [K], m \in [M]\}$ and random variables $X_j(t) \in \mathcal{X}_{\max_{k\in[K]}\{\lambda_{k,1}+\lambda_{k,2}+\cdots+\lambda_{k,M}\}}$, $j \in [N]$, $t \in \mathbb{N}$, independent of $\mathcal{G}$, and $\forall k \in [K], K \leq N$, define*

$$\begin{aligned}Z_k(t) &= L_k^b(t)(X_1(t), X_2(t), \cdots, X_N(t)) &(4.111)\\ Z_{k,1}(t) &= L_{k1}^{\vec{\gamma}_{k1}\vec{\delta}_{k1}}(t)((X_j(t))^{\sum_{r=1}^{i}\lambda_{kr}}_{\sum_{r=1}^{i-1}\lambda_{kr}}, i \in I_{k,1}, j \in [N]) &(4.112)\\ Z_{k,2}(t) &= L_{k2}^{\vec{\gamma}_{k2}\vec{\delta}_{k2}}(t)((X_j(t))^{\sum_{r=1}^{i}\lambda_{kr}}_{\sum_{r=1}^{i-1}\lambda_{kr}}, i \in I_{k,2}, j \in [N]) &(4.113)\\ &\vdots\end{aligned}$$

$$Z_{k,l_k}(t) \;=\; L_{kl_k}^{\vec{\gamma}_{kl_k}\vec{\delta}_{kl_k}}(t)((X_j(t))_{\sum_{r=1}^{i-1}\lambda_{kr}}^{\sum_{r=1}^{i}\lambda_{kr}}, i \in I_{k,l_k}, j \in [N]) \tag{4.114}$$

The channel uses are indexed by $t \in \mathbb{N}$. $I_{kk'} \subset [M], k \in [K], k' \in [l_k]$, *such that* $i < j \Rightarrow m(k,i) \geq m(k,j)$, *where*

$$m(a,b) \triangleq \min\{m : m \in I_{a,b}\}.$$

If for all $k \in [K]$ *and for each* $s \in \{1, 2, \cdots, l_k - 1\}$,

$$\mathcal{T}(Z_{k,s+1}) + \mathcal{T}(Z_{k,s+2}) + \cdots + \mathcal{T}(Z_{k,l_k}) \;\leq\; \lambda_{k,1} + \lambda_{k,2} + \cdots + \lambda_{k,(m(k,s)-1)} \tag{4.115}$$

then

$$H(Z_1^{[n]}, \cdots, Z_K^{[n]} \mid W, \mathcal{G}) \;\geq\; H(Z_{1,1}^{[n]}, \cdots, Z_{K,l_K}^{[n]} \mid W) + Kn\, o(\log \bar{P}) \tag{4.116}$$

In order to see how (4.110) is concluded, observe that any component of $\bar{\mathbf{Y}}_1$ is a linear combination of random variables including $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Any component of $(\bar{\mathbf{Y}}_{2b})^{1-\beta_{12}}$ is a linear combination of random variables including $(\bar{\mathbf{X}}_{2b})^{1-\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}_1$ has totally N_1 components which is larger or equal to the number of total components of $(\bar{\mathbf{Y}}_{2b})^{1-\beta_{12}}$. Therefore, replacing $[Z_{1,1}(t), \cdots, Z_{1,N_2-M_1}(t)]$ with the $(N_2 - M_1) \times 1$ vector $(\bar{\mathbf{Y}}_{2b}(t))^{1-\beta_{12}}$ and replacing $[Z_1(t), \cdots, Z_{N_1}(t)]$ with the $N_1 \times 1$ vector $\bar{\mathbf{Y}}_1(t)$, from (4.116) we have

$$H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) \;\leq\; H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \tag{4.117}$$

Note that $K = 1$ and $W = \bar{\mathbf{X}}_1^n$ are assumed. (4.110) implies that

$$2d_1 + d_2 \leq N_1 + M_1 + (N_2 - M_1)\beta_{12}. \tag{4.118}$$

4.6.5 Proof for Outer Bound (4.22)

In this subsection, we prove the outer bound (4.22) in Theorem 4.2 for the channel with $\max(M_1, N_1) \leq N_2 \leq M_2$ and $N_1 + M_1 \geq N_2$. We first split the $\bar{\mathbf{Y}}_1(t)$ in Subsection 4.5.2

into two parts as follows

$$\bar{\mathbf{Y}}'_{1a}(t) = \left[\bar{Y}_1^1(t) \quad \bar{Y}_1^2(t) \quad \cdots \quad \bar{Y}_1^{M_1+N_1-N_2}(t)\right]^T \tag{4.119}$$

$$\bar{\mathbf{Y}}'_{1b}(t) = \left[\bar{Y}_1^{M_1+N_1-N_2+1}(t) \quad \cdots \quad \bar{Y}_1^{N_1}(t)\right]^T \tag{4.120}$$

Then from (4.86), we have

$$\begin{aligned}
&nR_1\\
\leq& nN_1 \log \bar{P} - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.121)\\
\leq& nN_1 \log \bar{P} - H(\bar{\mathbf{Y}}'^n_{1b} \mid \bar{\mathbf{X}}_1^n) - H(\bar{\mathbf{Y}}'^n_{1a} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}'^n_{1b}) && (4.122)\\
=& nN_1 \log \bar{P} - H(\bar{\mathbf{Y}}'^n_{1b} \mid \bar{\mathbf{X}}_1^n) + H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - H(\bar{\mathbf{Y}}'^n_{1a} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}'^n_{1b})\\
&- \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.123)\\
\leq& nN_1 \log \bar{P} - H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - H(\bar{\mathbf{Y}}'^n_{1a} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}'^n_{1b}) - \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n)\\
&+ \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.124)\\
\leq& nN_1 \log \bar{P} - H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}}) - \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})\\
&+ \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.125)\\
\leq& nN_1 \log \bar{P} - \{nR_2 + nR_1 - n[M_1 + (N_2 - M_1)\beta_{12}] \log \bar{P}\}\\
&- \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) + \frac{N_1-N_2+M_1}{N_2-N_1} H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.126)\\
\leq& n[N_1 + M_1 + (N_2 - M_1)\beta_{12}] \log \bar{P} - nR_2 - nR_1\\
&- \frac{N_1-N_2+M_1}{N_2-N_1} [H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) - H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}})]\\
&+ \frac{N_1-N_2+M_1}{N_2-N_1} (nN_1 \log \bar{P} - nR_1) && (4.127)\\
\leq& n[\frac{N_2 M_1}{N_2-N_1} + (N_2 - M_1)\beta_{12}] \log \bar{P} - nR_2 - \frac{M_1}{N_2-N_1} nR_1\\
&- \frac{N_1-N_2+M_1}{N_2-N_1} [nR_2 - H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}})] && (4.128)\\
\leq& n[\frac{N_2 M_1}{N_2-N_1} + (N_2 - M_1)\beta_{12}] \log \bar{P} - \frac{M_1}{N_2-N_1} (nR_1 + nR_2)
\end{aligned}$$

$$+\frac{N_1-N_2+M_1}{N_2-N_1}[n\beta_{12}\min(M_2-N_1,N_2)\log\bar{P}] \tag{4.129}$$

$$=n\frac{N_2M_1+[M_1(N_2-M_1)+(M_1+N_1-N_2)\min(M_1,M_1+M_2-N_1-N_2)]\beta_{12}}{N_2-N_1}\log\bar{P}$$

$$-\frac{M_1}{N_2-N_1}(nR_1+nR_2) \tag{4.130}$$

(4.124) is concluded similar to (4.117). Any component of $\bar{\mathbf{Y}}'_{1b}$ is a linear combination of random variables including $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Any component of $(\bar{\mathbf{Y}}_{2b})^{1-\beta_{12}}$ is a linear combination of random variables including $(\bar{\mathbf{X}}_{2b})^{1-\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}'_1$ has totally N_2-M_1 components which is equal to the number of total components of $(\bar{\mathbf{Y}}_{2b})^{1-\beta_{12}}$. So similar to (4.117) we have, (4.124) is concluded, i.e., $H((\bar{\mathbf{Y}}^n_{2b})^{1-\beta_{12}})\leq H(\bar{\mathbf{Y}}'^n_{1b}\mid\bar{\mathbf{X}}^n_1)$. (4.125) is according to the following Lemma 4.3

Lemma 4.3. *For for the two-user deterministic MIMO IC defined in Subsection 4.6.2 with* $\max(M_1,N_1)\leq N_2\leq M_2$ *and* $N_1+M_1\geq N_2$, *we have,*

$$\frac{N_1-N_2+M_1}{N_2-N_1}H(\bar{\mathbf{X}}^n_{2b},(\bar{\mathbf{X}}^n_{2d})^{1-\beta_{12}})\leq\frac{N_1-N_2+M_1}{N_2-N_1}H(\bar{\mathbf{Y}}^n_1\mid\bar{\mathbf{X}}^n_1)+H(\bar{\mathbf{Y}}'^n_{1a}\mid\bar{\mathbf{X}}^n_1,\bar{\mathbf{Y}}'^n_{1b}) \tag{4.131}$$

See Section 4.10 for proof of Lemma 4.3. (4.126) is because of (4.102), i.e., $H((\bar{\mathbf{Y}}^n_{2b})^{1-\beta_{12}})\geq nR_2+nR_1-n[M_1+(N_2-M_1)\beta_{12}]\log\bar{P}$. (4.127) is because of (4.86). (4.129) is concluded as $H((\bar{\mathbf{X}}^n_{2d})_{\beta_{12}})\leq n\beta_{12}\min(M_2-N_1,N_2)\log\bar{P}$. Therefore, we have

$$n(N_2-N_1+M_1)R_1+nM_1R_2$$

$$\leq n\{N_2M_1+[M_1(N_2-M_1)+(M_1+N_1-N_2)\min(M_1,M_1+M_2-N_1-N_2)]\beta_{12}\}\log\bar{P} \tag{4.132}$$

This implies that

$$(N_2-N_1+M_1)d_1+M_1d_2\leq N_2M_1$$

$$+[M_1(N_2-M_1)+(M_1+N_1-N_2)\min(M_1,M_1+M_2-N_1-N_2)]\beta_{12} \tag{4.133}$$

This concludes the proof for Theorem 4.2.

4.7 Proof for Theorem 4.3

Note that the outer bounds (4.23) and (4.24) are trivial since they are the single user bounds. (4.25) is proved in Section 4.5. Thus in this subsection, we will prove (4.26), (4.27) and (4.28) in Theorem 4.3 with $N_1 \leq N_2 < \min(M_1, M_2)$.

4.7.1 Equivalent Channel

Without loss of generality, for the purpose of deriving a DoF outer bound, we can perform a sequence of invertible operations similar as [16] at transmitters and receivers to convert the channel to its simplest form. Specifically, at Transmitter 1, we aim to obtain an equivalent channel where the inputs are partitioned as $\mathbf{X}_1(t) = [\mathbf{X}_{1a}^T(t)\ \mathbf{X}_{1b}^T(t)\ \mathbf{X}_{1c}^T(t)\ \mathbf{X}_{1d}^T(t)]^T$. We have

$$\mathbf{X}_{1a}(t) = \left[X_1^1(t) \quad X_1^2(t) \quad \cdots \quad X_1^{\min(N_2, M_1-N_1)}(t)\right]^T \tag{4.134}$$

$$\mathbf{X}_{1b}(t) = \left[X_1^{\min(N_2, M_1-N_1)+1}(t) \quad \cdots \quad X_1^{N_2}(t)\right]^T \tag{4.135}$$

$$\mathbf{X}_{1c}(t) = \left[X_1^{N_2+1}(t) \quad \cdots \quad X_1^{N_2+(M_1-N_2-N_1)^+}(t)\right]^T \tag{4.136}$$

$$\mathbf{X}_{1d}(t) = \left[X_1^{N_2+(M_1-N_2-N_1)^++1}(t) \quad \cdots \quad X_1^{M_1}(t)\right]^T \tag{4.137}$$

where $|\mathbf{X}_{1a}(t)| = \min(N_2, M_1-N_1)$, $|\mathbf{X}_{1b}(t)| = (N_2-M_1+N_1)^+$, $|\mathbf{X}_{1c}(t)| = (M_1-N_2-N_1)^+$, $|\mathbf{X}_{1d}(t)| = \min(M_1-N_2, N_1)$.

Under the generic channel coefficient assumption, the matrices $\mathbf{G}_{11}(t)$ and $\hat{\mathbf{G}}_{21}(t)$ have M_1-N_1 and M_1-N_2 dimension right null space, respectively, almost surely. Note that there is a $(M_1-N_1-N_2)^+$ dimension intersection space of these two null spaces. Thus Transmitter 1 can zero-force $\mathbf{X}_{1c}(t)$ into this intersection null space, so that $\mathbf{X}_{1c}(t)$ cannot be heard by Receiver 1. Due to the estimation error $\tilde{\mathbf{G}}_{21}$, the residual interference caused by $\mathbf{X}_{1c}(t)$ at Receiver 2 has power level $\bar{P}^{-\beta_{21}}$.

Then Transmitter 1 can also zero-force $\mathbf{X}_{1a}(t)$ and $\mathbf{X}_{1d}(t)$ into the null space of $\mathbf{G}_{11}(t)$ and $\hat{\mathbf{G}}_{21}(t)$, respectively. So that $\mathbf{X}_{1a}(t)$ cannot be heard by Receiver 1, $\mathbf{X}_{1d}(t)$ is heard by Receiver 2 only with power level $\bar{P}^{-\beta_{21}}$. $\mathbf{X}_{1b}(t)$ can be heard by either of the receiver. Note that $\mathbf{X}_{1b}(t)$ and $\mathbf{X}_{1c}(t)$ cannot exist simultaneously, i.e., $\mathbf{X}_{1c}(t)$ do not exist when $N_1+N_2-M_1 \geq 0$ and $\mathbf{X}_{1b}(t)$ do not exist when $N_1+N_2-M_1 \leq 0$. The invertible operations for Transmitter 2 follows similar to 4.5.1. Then for Receiver 1, we aim to obtain an equivalent channel where the outputs are partitioned as $\mathbf{Y}_1'(t) = \begin{bmatrix}\mathbf{Y}_{1a}'^T(t) & \mathbf{Y}_{1b}'^T(t)\end{bmatrix}^T$. We have

$$\mathbf{Y}_{1a}'(t) = \begin{bmatrix} Y_1'^1(t) & Y_1'^2(t) & \cdots & Y_1'^{(N_2+N_1-M_1)^+}(t) \end{bmatrix}^T \tag{4.138}$$

$$\mathbf{Y}_{1b}'(t) = \begin{bmatrix} Y_1'^{(N_2+N_1-M_1)^++1}(t) & \cdots & Y_1'^{N_1}(t) \end{bmatrix}^T \tag{4.139}$$

Under the generic channel coefficient assumption, Receiver 1 can zero-force $\mathbf{X}_{1b}(t)$ and $\mathbf{X}_{1d}(t)$, so that $\mathbf{X}_{1b}(t)$ cannot be heard by its last $\min(M_1 - N_2, N_1)$ antennas i.e., $\mathbf{Y}_{1b}'(t)$, $\mathbf{X}_{1d}(t)$ cannot be heard by its first $(N_2 + N_1 - M_1)^+$ antennas i.e., $\mathbf{Y}_{1a}'(t)$. We now end up with the following simple channel model,

$$\mathbf{Y}_1'(t) = \begin{bmatrix} \mathbf{Y}_{1a}'(t) \\ \mathbf{Y}_{1b}'(t) \end{bmatrix} \tag{4.140}$$

$$\mathbf{Y}_{1a}'(t) = \mathbf{G}_{111}'(t)\mathbf{X}_{1b}(t) + \mathbf{G}_{121}'(t)\begin{bmatrix} \mathbf{X}_{2a}(t) \\ \mathbf{X}_{2b}(t) \end{bmatrix} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{121}'(t)\begin{bmatrix} \mathbf{X}_{2c}(t) \\ \mathbf{X}_{2d}(t) \end{bmatrix} + \mathbf{\Gamma}_{11}'(t) \tag{4.141}$$

$$\mathbf{Y}_{1b}'(t) = \mathbf{G}_{112}'(t)\mathbf{X}_{1d}(t) + \mathbf{G}_{122}'(t)\begin{bmatrix} \mathbf{X}_{2a}(t) \\ \mathbf{X}_{2b}(t) \end{bmatrix} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{122}'(t)\begin{bmatrix} \mathbf{X}_{2c}(t) \\ \mathbf{X}_{2d}(t) \end{bmatrix} + \mathbf{\Gamma}_{12}'(t) \tag{4.142}$$

$$\mathbf{Y}_2'(t) = \mathbf{G}_{21}'(t)\begin{bmatrix} \mathbf{X}_{1a}(t) \\ \mathbf{X}_{1b}(t) \end{bmatrix} + \sqrt{P^{-\beta_{21}}}\tilde{\mathbf{G}}_{21}'(t)\begin{bmatrix} \mathbf{X}_{1c}(t) \\ \mathbf{X}_{1d}(t) \end{bmatrix} + \mathbf{G}_{22}'(t)\begin{bmatrix} \mathbf{X}_{2b}(t) \\ \mathbf{X}_{2d}(t) \end{bmatrix} + \mathbf{\Gamma}_2'(t) \tag{4.143}$$

where $\mathbf{G}'_{111}(t)$ has size $(N_2+N_1-M_1)^+ \times (N_2+N_1-M_1)^+$. $\mathbf{G}'_{121}(t)$ and $\tilde{\mathbf{G}}'_{121}(t)$ has size $(N_2+N_1-M_1)^+ \times N_1$ and $(N_2+N_1-M_1)^+ \times (M_2-N_1)$, respectively. $\tilde{\mathbf{G}}'_{112}(t)$ has size $\min(M_1-N_2, N_1) \times \min(M_1 - N_2, N_1)$. $\mathbf{G}'_{122}(t)$ and $\tilde{\mathbf{G}}'_{122}(t)$ has size $\min(M_1 - N_2, N_1) \times N_1$ and $\min(M_1 - N_2, N_1) \times (M_2 - N_1)$, respectively. In addition, the partial CSIT quality of $\mathbf{G}'_{111}(t)$ and $\mathbf{G}'_{112}(t)$ is equal to β_{11}. The partial CSIT quality of $\mathbf{G}'_{121}(t)$ and $\mathbf{G}'_{122}(t)$ is equal to β_{12}. $\tilde{\mathbf{G}}'_{121}(t)$ and $\tilde{\mathbf{G}}'_{122}(t)$ are known by both receivers only to finite precision.

$\mathbf{G}'_{21}(t)$ has size $N_2 \times N_2$ and the partial CSIT quality is equal to β_{21}. $\tilde{\mathbf{G}}'_{21}(t)$ has size $N_2 \times (M_1 - N_2)$ and is known by both receivers only to finite precision. $\mathbf{G}'_{22}(t)$ has size $N_2 \times \min(N_2, M_2)$and the partial CSIT quality is equal to β_{22}.

4.7.2 Deterministic Model

Based on the equivalent channel in Subsection 4.7.1, we define the deterministic channel model with input $\bar{\mathbf{X}}_k(t) = [\bar{X}_k^1(t)\ \bar{X}_k^2(t)\ \cdots\ \bar{X}_k^{M_k}(t)]^T, \forall\ k \in \{1, 2\}$, such that

$$\bar{\mathbf{Y}}_1(t) = \begin{bmatrix} \bar{\mathbf{Y}}_{1a}(t) \\ \bar{\mathbf{Y}}_{1b}(t) \end{bmatrix} \tag{4.144}$$

$$\bar{\mathbf{Y}}_{1a}(t) = \left\lfloor \mathbf{G}'_{111}(t)\bar{\mathbf{X}}_{1b}(t) \right\rfloor + \left\lfloor \mathbf{G}'_{121}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2a}(t) \\ \bar{\mathbf{X}}_{2b}(t) \end{bmatrix} \right\rfloor + \left\lfloor \tilde{\mathbf{G}}'_{121}(t) \begin{bmatrix} (\bar{\mathbf{X}}_{2c}(t))^{1-\beta_{12}} \\ (\bar{\mathbf{X}}_{2d}(t))^{1-\beta_{12}} \end{bmatrix} \right\rfloor \tag{4.145}$$

$$\bar{\mathbf{Y}}_{1b}(t) = \left\lfloor \mathbf{G}'_{112}(t)\bar{\mathbf{X}}_{1d}(t) \right\rfloor + \left\lfloor \mathbf{G}'_{122}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2a}(t) \\ \bar{\mathbf{X}}_{2b}(t) \end{bmatrix} \right\rfloor + \left\lfloor \tilde{\mathbf{G}}'_{122}(t) \begin{bmatrix} (\bar{\mathbf{X}}_{2c}(t))^{1-\beta_{12}} \\ (\bar{\mathbf{X}}_{2d}(t))^{1-\beta_{12}} \end{bmatrix} \right\rfloor \tag{4.146}$$

$$\bar{\mathbf{Y}}_2(t) = \left\lfloor \mathbf{G}'_{21}(t) \begin{bmatrix} \bar{\mathbf{X}}_{1a}(t) \\ \bar{\mathbf{X}}_{1b}(t) \end{bmatrix} \right\rfloor + \left\lfloor \tilde{\mathbf{G}}'_{21}(t) \begin{bmatrix} (\bar{\mathbf{X}}_{1c}(t))^{1-\beta_{21}} \\ (\bar{\mathbf{X}}_{1d}(t))^{1-\beta_{21}} \end{bmatrix} \right\rfloor + \left\lfloor \mathbf{G}'_{22}(t) \begin{bmatrix} \bar{\mathbf{X}}_{2b}(t) \\ \bar{\mathbf{X}}_{2d}(t) \end{bmatrix} \right\rfloor \tag{4.147}$$

and $\bar{X}_k^i(t) \in \{0, 1, \cdots, \lfloor \bar{P} \rfloor\}$, $\forall i \in \mathcal{I}_{M_k}$. The size of $\bar{\mathbf{X}}_{ij}(t)$ is the same with $\mathbf{X}_{ij}(t), \forall i \in \{1, 2\}, j \in \{a, b, c, d\}$ in Subsection 4.7.1. Note that $(\bar{\mathbf{X}}_{2c}(t))_{\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d}(t))_{\beta_{12}}$ are omitted in (4.145) and (4.146) since they are under noise floor at Receiver 1. $(\bar{\mathbf{X}}_{1c}(t))_{\beta_{21}}$ and $(\bar{\mathbf{X}}_{1d}(t))_{\beta_{21}}$ are omitted in (4.147) since they are under noise floor at Receiver 2.

4.7.3 Proof for Outer Bound (4.26)

In this subsection, we prove the outer bound (4.26) in Theorem 4.3. For Receiver 2, we have

$$nR_2 \le H(\bar{\mathbf{Y}}_2^n) - H(\bar{\mathbf{Y}}_2^n \mid \bar{\mathbf{X}}_2^n) \tag{4.148}$$

$$\le nN_2 \log \bar{P} - H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \tag{4.149}$$

$$= nN_2 \log \bar{P} - H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) - H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}) + H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}) \tag{4.150}$$

$$\le nN_2 \log \bar{P} - nR_1 + n \min(M_1 - N_2, N_1)\beta_{21} \log \bar{P} \tag{4.151}$$

(4.149) is true similar to (4.117). Any component of $\bar{\mathbf{Y}}_2$ is a linear combination of random variables including $\bar{\mathbf{X}}_{1b}$ and $(\bar{\mathbf{X}}_{1d})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}_2$ has totally N_2 components which is equal to the sum of total components of $\bar{\mathbf{X}}_{1b}$ and $(\bar{\mathbf{X}}_{1d})^{1-\beta_{12}}$. So similar to (4.117) we have, $H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \le H(\bar{\mathbf{Y}}_2^n \mid \bar{\mathbf{X}}_2^n)$. (4.151) is because $H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}) \le n \min(M_1 - N_2, N_1)\beta_{21} \log \bar{P}$. Thus, we have

$$nR_1 + nR_2 \le n[N_2 + \min(M_1 - N_2, N_1)\beta_{21}] \log \bar{P} \tag{4.152}$$

This implies that

$$d_1 + d_2 \le N_2 + \min(M_1 - N_2, N_1)\beta_{21} \tag{4.153}$$

4.7.4 Proof for Outer Bound (4.27)

In this subsection, we prove the outer bound (4.27) in Theorem 4.3. First, we further split $\bar{\mathbf{X}}_{2d}(t)$ into two parts, i.e., $\bar{\mathbf{X}}_{2d}(t) = [\bar{\mathbf{X}}_{2d1}^T(t)\ \bar{\mathbf{X}}_{2d2}^T(t)]^T$. We have

$$\bar{\mathbf{X}}_{2d1}(t) = \left[\bar{X}_2^{N_1+(M_2-N_1-N_2)^+ +1}(t) \quad \cdots \quad \bar{X}_2^{M_2+N_1-N_2}(t)\right]^T \tag{4.154}$$

$$\bar{\mathbf{X}}_{2d2}(t) = \left[\bar{X}_2^{M_2+N_1-N_2+1}(t) \quad \cdots \quad \bar{X}_2^{M_2}(t)\right]^T \tag{4.155}$$

where $|\mathbf{X}_{2d1}(t)| = \min(M_2 - N_2, N_1)$, $|\mathbf{X}_{2d2}(t)| = N_2 - N_1$. For Receiver 1, we have

$$
\begin{aligned}
& nR_1 \\
\leq & H(\bar{\mathbf{Y}}_1^n) - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.156) \\
\leq & nN_1 \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d1}^n)^{1-\beta_{12}}) && (4.157) \\
= & nN_1 \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d1}^n)^{1-\beta_{12}}) - H((\bar{\mathbf{X}}_{2d2}^n)^{1-\beta_{12}}) + H((\bar{\mathbf{X}}_{2d2}^n)^{1-\beta_{12}}) \\
& - H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) && (4.158) \\
\leq & nN_1 \log \bar{P} - nR_2 + H((\bar{\mathbf{X}}_{2d2}^n)^{1-\beta_{12}}) + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) && (4.159) \\
\leq & nN_1 \log \bar{P} - nR_2 + n(N_2 - N_1)(1 - \beta_{12}) \log \bar{P} + n \min(M_2 - N_1, N_2)\beta_{12} \log \bar{P} && (4.160)
\end{aligned}
$$

(4.157) is true similar to (4.117). Any component of $\bar{\mathbf{Y}}_1$ is a linear combination of random variables including $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d1})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}_1$ has totally N_1 components which is equal to the sum of total components of $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d1})^{1-\beta_{12}}$. So similar to (4.117), we have, $H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d1}^n)^{1-\beta_{12}}) \leq H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n)$. (4.160) is concluded as follows.

$$
\begin{aligned}
H((\bar{\mathbf{X}}_{2d2}^n)^{1-\beta_{12}}) & \leq n(N_2 - N_1)(1 - \beta_{12}) \log \bar{P} && (4.161) \\
H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) & \leq n \min(M_2 - N_1, N_2)\beta_{12} \log \bar{P} && (4.162)
\end{aligned}
$$

Therefore, we have

$$
nR_1 + nR_2 \leq n[N_2 + \min(M_2 - N_2, N_1)\beta_{12}] \log \bar{P} \quad (4.163)
$$

This implies that

$$
d_1 + d_2 \leq N_2 + \min(M_2 - N_2, N_1)\beta_{12} \quad (4.164)
$$

4.7.5 Proof for Outer Bound (4.28)

In this subsection, we first prove the following lemma.

Lemma 4.4. *For for the two-user deterministic MIMO IC defined in Subsection 4.7.2, we* *have,*

$$
\begin{aligned}
&H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \\
\leq & n \min(M_1 - N_2, N_1)\beta_{21} \log \bar{P} - H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \qquad (4.165)
\end{aligned}
$$

Proof. For Receiver 1,

$$
H(\bar{\mathbf{X}}_{1b}^n, \bar{\mathbf{X}}_{1d}^n) = I(\bar{\mathbf{X}}_{1b}^n, \bar{\mathbf{X}}_{1d}^n; \bar{\mathbf{Y}}_1^n) \qquad (4.166)
$$

$$
\begin{aligned}
&H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) + H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \\
=& I(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}; \bar{\mathbf{Y}}_1^n) + I((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}; \bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \qquad (4.167)
\end{aligned}
$$

Note that $H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) = I(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}; \bar{\mathbf{Y}}_1^n)$, then we have

$$
\begin{aligned}
&H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \\
=& I((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}; \bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \qquad (4.168) \\
=& I((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}; \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \\
&+ I((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}; (\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \qquad (4.169) \\
\leq& H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \\
&- H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, \bar{\mathbf{X}}_{1d}^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \qquad (4.170) \\
\leq& n \min(M_1 - N_2, N_1)\beta_{21} \log \bar{P} - H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \qquad (4.171)
\end{aligned}
$$

(4.170) is because given $\{\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}\}$, $(\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}$ and $\{\bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}\}$ are independent, i.e., $I((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}}; \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) = 0$. ■

Case $\beta_{12}+\beta_{21}<1$**:**

We now prove the outer bound (4.28) in Theorem 4.3 when $\beta_{12}+\beta_{21}<1$. For Receiver 1,

$$
\begin{aligned}
& nR_1 \\
\leq & nN_1 \log \bar{P} - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) && (4.172)\\
= & nN_1 \log \bar{P} - H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) - \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \\
& + \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) && (4.173)\\
\leq & nN_1 \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) \\
& + \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) && (4.174)\\
\leq & nN_1 \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) \\
& + \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} [n \min(M_1-N_2, N_1)\beta_{21} \log \bar{P} - H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}})] \\
& && (4.175)\\
= & n[N_1 + (N_2-N_1)(1-\beta_{12})] \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) - H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) \\
& - \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} [H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \\
& + H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) - H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}})] && (4.176)\\
\leq & n[N_1 + (N_2-N_1)(1-\beta_{12})] \log \bar{P} - nR_2 + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) \\
& - \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} [nR_1 - H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}})] && (4.177)\\
\leq & n[N_1 + (N_2-N_1)(1-\beta_{12})] \log \bar{P} - nR_2 + n \min(M_2-N_1, N_2)\beta_{12} \log \bar{P} \\
& - \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} [nR_1 - (nN_2 \log \bar{P} - nR_2)] && (4.178)\\
= & n[N_2 + \min(M_2-N_2, N_1)\beta_{12}] \log \bar{P} + n \frac{N_2(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} \log \bar{P} \\
& - \frac{(N_2-N_1)(1-\beta_{12})}{\min(M_1-N_2, N_1)\beta_{21}} [nR_1 + nR_2] - nR_2 && (4.179)
\end{aligned}
$$

(4.174) is according to the proof for MIMO BC in [13] [2], we have,

$$\min(M_1{-}N_2,N_1)\beta_{21}H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + (N_2{-}N_1)(1{-}\beta_{12})H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}})$$
$$\geq \min(M_1{-}N_2,N_1)\beta_{21}H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) \tag{4.180}$$

(4.175) follows from Lemma 4.4 and (4.178) is true from (4.149) and (4.162). Thus, we have

$$\begin{aligned} &n[\min(M_1{-}N_2,N_1)\beta_{21}{+}(N_2{-}N_1)(1{-}\beta_{12})](R_1+R_2) \\ \leq& n[\min(M_1{-}N_2,N_1)\beta_{21}+(N_2{-}N_1)(1{-}\beta_{12})]N_2\log\bar{P} \\ &+n\min(M_1{-}N_2,N_1)\min(M_2{-}N_2,N_1)\beta_{12}\beta_{21}\log\bar{P} \end{aligned} \tag{4.181}$$

This implies that

$$d_1+d_2 \leq N_2 + \frac{\min(M_1{-}N_2,N_1)\min(M_2{-}N_2,N_1)\beta_{12}\beta_{21}}{\min(M_1{-}N_2,N_1)\beta_{21}{+}(N_2{-}N_1)(1{-}\beta_{12})} \tag{4.182}$$

Case $\beta_{12}+\beta_{21}\geq 1$:

We now prove the outer bound (4.28) in Theorem 4.3 when $\beta_{12}+\beta_{21}\geq 1$. For Receiver 1,

$$nR_1$$
$$\leq nN_1\log\bar{P}-H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \tag{4.183}$$
$$\begin{aligned} =&nN_1\log\bar{P}-H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n)-\frac{N_2{-}N_1}{\min(M_1{-}N_2,N_1)}H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \\ &+\frac{N_2{-}N_1}{\min(M_1{-}N_2,N_1)}H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \end{aligned} \tag{4.184}$$
$$\leq nN_1\log\bar{P}-H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})+\frac{N_2{-}N_1}{\min(M_1{-}N_2,N_1)}H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \tag{4.185}$$

[2]The random variables W_1 is arbitrary and can be replaced with $\bar{\mathbf{X}}_1^n$.

$$
\begin{aligned}
&\leq nN_1 \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) \\
&\quad + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} [n \min(M_1 - N_2, N_1)\beta_{21} \log \bar{P} - H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}})]
\end{aligned} \tag{4.186}
$$

$$
\begin{aligned}
&= n[N_1 + (N_2 - N_1)\beta_{21}] \log \bar{P} - H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}) - H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) \\
&\quad + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} [H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) - H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) \\
&\quad - H((\bar{\mathbf{X}}_{1d}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}})]
\end{aligned} \tag{4.187}
$$

$$
\begin{aligned}
&\leq n[N_1 + (N_2 - N_1)\beta_{21}] \log \bar{P} - nR_2 + H((\bar{\mathbf{X}}_{2d}^n)_{\beta_{12}}) \\
&\quad + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} [H(\bar{\mathbf{X}}_{1b}^n, (\bar{\mathbf{X}}_{1d}^n)^{1-\beta_{21}}) - nR_1]
\end{aligned} \tag{4.188}
$$

$$
\begin{aligned}
&\leq n[N_1 + (N_2 - N_1)\beta_{21}] \log \bar{P} - nR_2 + n \min(M_2 - N_1, N_2)\beta_{12} \log \bar{P} \\
&\quad + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} [(nN_2 \log \bar{P} - nR_2) - nR_1]
\end{aligned} \tag{4.189}
$$

$$
\begin{aligned}
&= n[N_1 + (N_2 - N_1)\beta_{21}] \log \bar{P} - nR_2 + n \min(M_2 - N_1, N_2)\beta_{12} \log \bar{P} \\
&\quad + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} [(nN_2 \log \bar{P} - nR_2) - nR_1]
\end{aligned} \tag{4.190}
$$

(4.185) is true as from the MIMO BC proof in [13] [3], we have,

$$
\begin{aligned}
&H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) + \frac{N_2 - N_1}{\min(M_1 - N_2, N_1)} H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_{21}} \mid \bar{\mathbf{X}}_1^n, \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_{21}}) \\
&\geq H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})
\end{aligned} \tag{4.191}
$$

(4.186) is concluded from Lemma 4.4. (4.189) is true from (4.149) and (4.162). Then we have,

$$
\begin{aligned}
&n \min(M_1 - N_1, N_2)(R_1 + R_2) \\
&\leq n[\min(M_1 - N_2, N_1)N_1 + (N_2 - N_1)N_2]N_2 \log \bar{P} \\
&\quad + n \min(M_1 - N_2, N_1)[(N_2 - N_1)\beta_{21} + \min(M_2 - N_1, N_2)\beta_{12}] \log \bar{P}
\end{aligned} \tag{4.192}
$$

[3] The random variables W_1 is arbitrary and can be replaced with $\bar{\mathbf{X}}_1^n$.

This implies that

$$d_1 + d_2 \leq N_2 + \min(M_1 - N_2, N_1) \frac{N_1 - N_2 + (N_2 - N_1)\beta_{21} + \min(M_2 - N_1, N_2)\beta_{12}}{\min(M_1 - N_1, N_2)} \tag{4.193}$$

This concludes the proof for Theorem 4.3.

4.8 Proof of Lemma 4.1

As $\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}$ are independent from $\bar{\mathbf{X}}_1^n$, (4.87) can be rewritten as,

$$\begin{aligned} &\min(M_1, N_1) H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n) + (\min(M_2, N_2) - N_1) H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \\ \leq &(\min(M_2, N_2) - N_1 + \min(M_1, N_1)) H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \end{aligned} \tag{4.194}$$

Get the two nonnegative numbers r_1 and r_2, where

$$r_1 \leq N_1 - \min(N_1, (M_2 - N_2)^+) \tag{4.195}$$

$$r_2 \leq \min(M_2 - N_1, N_2) \tag{4.196}$$

$$r_1 + r_2 = (N_1 - M_1)+ \tag{4.197}$$

Note that these numbers r_1 and r_2 exists as $(N_1 - M_1)+ \leq \min(M_2, N_2)$. Define the vectors $\bar{\mathbf{X}}_{2ba}(t)$, $\bar{\mathbf{X}}_{2bb}(t)$, $\bar{\mathbf{X}}_{2da}(t)$, $\bar{\mathbf{X}}_{2db}(t)$ and $\bar{\mathbf{X}}_{i,m}(t)$ as,

$$\bar{\mathbf{X}}_{2ba}(t) = [\bar{\mathbf{X}}_{2b}(t)]_{0,(N_1-(M_2-N_2)^+)^+-r_1} \tag{4.198}$$

$$\bar{\mathbf{X}}_{2bb}(t) = [\bar{\mathbf{X}}_{2b}(t)]_{(N_1-(M_2-N_2)^+)^+-r_1,r_1} \tag{4.199}$$

$$\bar{\mathbf{X}}_{2da}(t) = [\bar{\mathbf{X}}_{2d}(t)]_{0,\min(M_2-N_1,N_2)-r_2} \tag{4.200}$$

$$\bar{\mathbf{X}}_{2db}(t) = [\bar{\mathbf{X}}_{2d}(t)]_{\min(M_2-N_1,N_2)-r_2,r_2} \tag{4.201}$$

$$\bar{\mathbf{X}}_{i,m}(t) = [\bar{\mathbf{X}}_{2ba}(t); \bar{\mathbf{X}}_{2da}(t)]_{i,m} \tag{4.202}$$

Thus, we have,

$$\min(M_1,N_1)H(\bar{\mathbf{X}}_{2ba}^n,\bar{\mathbf{X}}_{2bb}^n,(\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}},(\bar{\mathbf{X}}_{2db}^n)^{1-\beta_{12}})+(\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_{1b}^n\mid\bar{\mathbf{X}}_1^n)$$

$$\leq\min(M_1,N_1)H(\bar{\mathbf{Y}}_{1b}^n,\bar{\mathbf{X}}_{2ba}^n,(\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}}\mid\bar{\mathbf{X}}_1^n)+(\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_{1b}^n\mid\bar{\mathbf{X}}_1^n) \tag{4.203}$$

$$=\min(M_1,N_1)H(\bar{\mathbf{X}}_{2ba}^n,(\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}}\mid\bar{\mathbf{Y}}_{1b}^n,\bar{\mathbf{X}}_1^n)$$

$$+(\min(M_1,N_1)+\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_{1b}^n\mid\bar{\mathbf{X}}_1^n) \tag{4.204}$$

$$\leq(\min(M_1,N_1)+\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_{1a}^n\mid\bar{\mathbf{Y}}_{1b}^n,\bar{\mathbf{X}}_1^n)$$

$$+(\min(M_1,N_1)+\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_{1b}^n\mid\bar{\mathbf{X}}_1^n) \tag{4.205}$$

$$=(\min(M_1,N_1)+\min(M_2,N_2)-N_1)H(\bar{\mathbf{Y}}_1^n\mid\bar{\mathbf{X}}_1^n) \tag{4.206}$$

Any component of $\bar{\mathbf{Y}}_{1b}$ is a linear combination of $\bar{\mathbf{X}}_{2a}$, $\bar{\mathbf{X}}_{2b}$, $(\bar{\mathbf{X}}_{2c})^{1-\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Since Theorem 4.4 requires that all the random variables are over the same alphabet, we need to add complementary variables to $(\bar{\mathbf{X}}_{2c})^{1-\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Specifically, we define $\bar{\mathbf{X}}_{2c}^*=\left\lfloor\bar{P}^{1-\beta_{12}}\right\rfloor\bar{\mathbf{X}}_{2c}^c+(\bar{\mathbf{X}}_{2c})^{1-\beta_{12}}$ and $\bar{\mathbf{X}}_{2d}^*=\left\lfloor\bar{P}^{1-\beta_{12}}\right\rfloor\bar{\mathbf{X}}_{2d}^c+(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$, where each component of $\bar{\mathbf{X}}_{2c}^c$ and $\bar{\mathbf{X}}_{2d}^c$ is an integer valued random variables over the alphabet $\{0,1,\cdots,\left\lfloor\bar{P}^{\beta_{12}}\right\rfloor\}$. Then $\bar{\mathbf{Y}}_{1b}^*$ is the linear combination of $\bar{\mathbf{X}}_{2a}$, $\bar{\mathbf{X}}_{2b}$ $\bar{\mathbf{X}}_{2c}^*$ and $\bar{\mathbf{X}}_{2d}^*$, i.e., (4.203) is ture.
Now each random variable of $\bar{\mathbf{Y}}_{1b}^*$ is over the same alphabet, i.e., $\{0,1,\cdots,\left\lfloor\bar{P}\right\rfloor\}$. The vector $\bar{\mathbf{Y}}_{1b}^*$ has totally $(N_1-M_1)^+$ components which is equal to the sum of total components of $\bar{\mathbf{X}}_{2bb}$ and $(\bar{\mathbf{X}}_{2db})^{1-\beta_{12}}$. By applying Theorem 4.4, we have

$$H(\bar{\mathbf{X}}_{2bb},(\bar{\mathbf{X}}_{2db})^{1-\beta_{12}}|W)\leq H(\bar{\mathbf{Y}}_{1b}^*|W). \tag{4.207}$$

Note that (4.207) holds even with $\bar{\mathbf{X}}_{2c}^c=\mathbf{0}$ and $\bar{\mathbf{X}}_{2d}^c=\mathbf{0}$. Therefore, when $\bar{\mathbf{X}}_{2c}^c=\mathbf{0}$ and $\bar{\mathbf{X}}_{2d}^c=\mathbf{0}$, $H(\bar{\mathbf{X}}_{2bb},(\bar{\mathbf{X}}_{2db})^{1-\beta_{12}}|W)\leq H(\bar{\mathbf{Y}}_{1b}^*|W)=H(\bar{\mathbf{Y}}_{1b}|W)$, i.e., (4.203) is true. (4.204) and (4.206) comes from the chain rule, and finally, (4.205) is true as

$$\min(M_1,N_1)H(\bar{\mathbf{X}}_{2ba}^n,(\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}}\mid W) \tag{4.208}$$

$$\leq \sum_{i=1}^{\min(M_1,N_1)+\min(M_2,N_2)-N_1} H(\bar{\mathbf{X}}^n_{i,\min(M_1,N_1)} \mid W) \tag{4.209}$$

$$\leq (\min(M_1,N_1)+\min(M_2,N_2)-N_1) H(\bar{\mathbf{Y}}^n_{1a} \mid W) \tag{4.210}$$

where $\bar{\mathbf{X}}_{i,\min(M_1,N_1)}$ is defined in (4.202). (4.209) is true from sub-modularity properties of entropy function, i.e.

$$mH(X_1, X_2, \cdots, X_n) \leq \sum_{i=0}^{n-1} H(X_{i+1}, X_{i+2}, \cdots, X_{i+m}) \tag{4.211}$$

where $0 < m \leq n$, and X_{n+k} is defined as X_k. (4.210) follows similar to (4.203) as any component of $\bar{\mathbf{Y}}_{1a}$ is a linear combination of random variables including of components $\bar{\mathbf{X}}_{2ba}$ and $(\bar{\mathbf{X}}_{2da})^{1-\beta_{12}}$, and the vector $\bar{\mathbf{Y}}_{1a}$ has totally $\min(N_1, M_1)$ components.

4.9 Proof of Lemma 4.2

As $(\bar{\mathbf{Y}}^n_{2b})^{1-\beta_{12}}$ and $\bar{\mathbf{X}}^n_1$ are independent from each other, (4.106) can be rewritten as,

$$\min(M_1,N_1) H((\bar{\mathbf{Y}}^n_{2b})^{1-\beta_{12}} \mid \bar{\mathbf{X}}^n_1) + (N_2 - M_1 - N_1) H(\bar{\mathbf{Y}}^n_{1b} \mid \bar{\mathbf{X}}^n_1)$$
$$\leq (N_2 - \max(M_1,N_1)) H(\bar{\mathbf{Y}}^n_1 \mid \bar{\mathbf{X}}^n_1) \tag{4.212}$$

If $N_1 \leq M_1$, (4.212) is immediately concluded from Theorem 4.4 as $H(\bar{\mathbf{Y}}^n_{1b} \mid \bar{\mathbf{X}}^n_1) = 0$. So we can assume $M_1 < N_1$. Similar to proof of Lemma 4.1, define $\bar{\mathbf{Y}}_{2ba}(t)$ and $\bar{\mathbf{Y}}_{2bb}(t)$ as,

$$\bar{\mathbf{Y}}_{2ba}(t) = [\bar{\mathbf{Y}}_{2b}(t)]_{0,N_1-M_1} \tag{4.213}$$

$$\bar{\mathbf{Y}}_{2bb}(t) = [\bar{\mathbf{Y}}_{2b}(t)]_{N_1-M_1,N_2-N_1} \tag{4.214}$$

$$\bar{\mathbf{Y}}_{i,m}(t) = [\bar{\mathbf{Y}}_{2bb}(t)]_{i,m} \tag{4.215}$$

See Definition 4.5 for (4.213) and (4.214). Thus, we have,

$$\min(M_1, N_1)H((\bar{\mathbf{Y}}_{2b}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-M_1-N_1)H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \quad (4.216)$$

$$= M_1H((\bar{\mathbf{Y}}_{2ba}^n)^{1-\beta_{12}}, (\bar{\mathbf{Y}}_{2bb}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-M_1-N_1)H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \quad (4.217)$$

$$\leq M_1H(\bar{\mathbf{Y}}_{1b}^n, (\bar{\mathbf{Y}}_{2bb}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-M_1-N_1)H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \quad (4.218)$$

$$= M_1H((\bar{\mathbf{Y}}_{2bb}^n)^{1-\beta_{12}} \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n)+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \quad (4.219)$$

$$\leq (N_2-N_1)H(\bar{\mathbf{Y}}_{1a}^n \mid \bar{\mathbf{Y}}_{1b}^n, \bar{\mathbf{X}}_1^n)+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}^n \mid \bar{\mathbf{X}}_1^n) \quad (4.220)$$

$$= (N_2-N_1)H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \quad (4.221)$$

(4.218) is concluded from Theorem 4.4. Any component of $\bar{\mathbf{Y}}_{1b}$ is a linear combination of random variables including $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Any component of $(\bar{\mathbf{Y}}_{2ba})^{1-\beta_{12}}$ is a linear combination of random variables including $(\bar{\mathbf{X}}_{2b})^{1-\beta_{12}}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}_{1b}$ has totally $N_1 - M_1$ components which is equal to the number of total components of $(\bar{\mathbf{Y}}_{2ba})^{1-\beta_{12}}$. So from Theorem 4.4 we have, $H((\bar{\mathbf{Y}}_{2ba}^n)^{1-\beta_{12}}|W) \leq H(\bar{\mathbf{Y}}_{1b}^n|W)$. (4.219) and (4.221) are true from chain rule. (4.220) follows as,

$$M_1H((\bar{\mathbf{Y}}_{2bb}^n)^{1-\beta_{12}} \mid W) \quad (4.222)$$

$$\leq \sum_{i=1}^{N_2-N_1} H(\bar{\mathbf{Y}}_{i,M_1}^n \mid W) \quad (4.223)$$

$$\leq (N_2 - N_1)H(\bar{\mathbf{Y}}_{1a}^n \mid W) \quad (4.224)$$

(4.223) is true similar to (4.209) from sub-modularity properties of entropy function as $M_1 + N_1 < N_2$. (4.224) follows similar to (4.218) from Theorem 4.4.

4.10 Proof of Lemma 4.3

(4.131) can be rewritten as,

$$(N_1-N_2+M_1)H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}})+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}'^n \mid \bar{\mathbf{X}}_1^n)$$

$$\leq M_1 H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \tag{4.225}$$

Moreover, as $\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}}$ are independent from $\bar{\mathbf{X}}_1^n$, (4.131) easily can be written as,

$$\begin{aligned}&(N_1-N_2+M_1)H(\bar{\mathbf{X}}_{2b}^n, (\bar{\mathbf{X}}_{2d}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}'^n \mid \bar{\mathbf{X}}_1^n)\\ \leq& M_1 H(\bar{\mathbf{Y}}_1^n \mid \bar{\mathbf{X}}_1^n) \end{aligned} \tag{4.226}$$

Similar to Proof of Lemma 4.1, get the two nonnegative numbers r_1 and r_2, where

$$r_1 \leq N_1 - \min(N_1, (M_2-N_2)^+) \tag{4.227}$$

$$r_2 \leq \min(M_2 - N_1, N_2) \tag{4.228}$$

$$r_1 + r_2 = N_2 - M_1 \tag{4.229}$$

Note that these numbers r_1 and r_2 exists as $N_2 - M_1 \leq \min(M_2, N_2)$. Define the vectors $\bar{\mathbf{X}}_{2ba}(t)$, $\bar{\mathbf{X}}_{2bb}(t)$, $\bar{\mathbf{X}}_{2da}(t)$, $\bar{\mathbf{X}}_{2db}(t)$ and $\bar{\mathbf{X}}_{i,m}(t)$ as,

$$\bar{\mathbf{X}}_{2ba}(t) = [\bar{\mathbf{X}}_{2b}(t)]_{0,(N_1-(M_2-N_2)^+)^+-r_1} \tag{4.230}$$

$$\bar{\mathbf{X}}_{2bb}(t) = [\bar{\mathbf{X}}_{2b}(t)]_{(N_1-(M_2-N_2)^+)^+-r_1,r_1} \tag{4.231}$$

$$\bar{\mathbf{X}}_{2da}(t) = [\bar{\mathbf{X}}_{2d}(t)]_{0,\min(M_2-N_1,N_2)-r_2} \tag{4.232}$$

$$\bar{\mathbf{X}}_{2db}(t) = [\bar{\mathbf{X}}_{2d}(t)]_{\min(M_2-N_1,N_2)-r_2,r_2} \tag{4.233}$$

$$\bar{\mathbf{X}}_{i,m}(t) = [\bar{\mathbf{X}}_{2ba}(t); \bar{\mathbf{X}}_{2da}(t)]_{i,m} \tag{4.234}$$

See Definition 4.5 for the notations in equations (4.230) - (4.234).Thus, we have,

$$(N_1-N_2+M_1)H(\bar{\mathbf{X}}_{2ba}^n, \bar{\mathbf{X}}_{2bb}^n, (\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}}, (\bar{\mathbf{X}}_{2db}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}'^n \mid \bar{\mathbf{X}}_1^n)$$

$$\leq (N_1-N_2+M_1)H(\bar{\mathbf{Y}}_{1b}'^n, \bar{\mathbf{X}}_{2ba}^n, (\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}} \mid \bar{\mathbf{X}}_1^n)+(N_2-N_1)H(\bar{\mathbf{Y}}_{1b}'^n \mid \bar{\mathbf{X}}_1^n) \tag{4.235}$$

$$= (N_1-N_2+M_1)H(\bar{\mathbf{X}}_{2ba}^n, (\bar{\mathbf{X}}_{2da}^n)^{1-\beta_{12}} \mid \bar{\mathbf{Y}}_{1b}'^n, \bar{\mathbf{X}}_1^n)+M_1 H(\bar{\mathbf{Y}}_{1b}'^n \mid \bar{\mathbf{X}}_1^n) \tag{4.236}$$

$$\leq \sum_{i=1}^{M_1} H(\bar{\mathbf{X}}^n_{i,N_1-N_2+M_1} \mid \bar{\mathbf{Y}}'^n_{1b}, \bar{\mathbf{X}}^n_1) + M_1 H(\bar{\mathbf{Y}}'^n_{1b} \mid \bar{\mathbf{X}}^n_1) \tag{4.237}$$

$$\leq M_1 H(\bar{\mathbf{Y}}'^n_{1a} \mid \bar{\mathbf{Y}}'^n_{1b}, \bar{\mathbf{X}}^n_1) + M_1 H(\bar{\mathbf{Y}}'^n_{1b} \mid \bar{\mathbf{X}}^n_1) \tag{4.238}$$

$$= M_1 H(\bar{\mathbf{Y}}^n_1 \mid \bar{\mathbf{X}}^n_1) \tag{4.239}$$

Any component of $\bar{\mathbf{Y}}'_{1b}$ is a linear combination of random variables including $\bar{\mathbf{X}}_{2b}$ and $(\bar{\mathbf{X}}_{2d})^{1-\beta_{12}}$. Moreover, the vector $\bar{\mathbf{Y}}'_{1b}$ has totally $N_2 - M_1$ components which is equal to the sum of total components of $\bar{\mathbf{X}}_{2bb}$ and $(\bar{\mathbf{X}}_{2db})^{1-\beta_{12}}$. So from Theorem 4.4 we have, $H(\bar{\mathbf{X}}^n_{2bb}, (\bar{\mathbf{X}}^n_{2db})^{1-\beta_{12}}|W) \leq H(\bar{\mathbf{Y}}'^n_{1b}|W)$, i.e., (4.235) is true. (4.236) and (4.239) comes from the chain rule. The way $\bar{\mathbf{X}}_{i,N_1-N_2+M_1}$ is defined in (4.234), from sub-modularity properties of entropy function in (4.211), (4.237) is concluded. (4.238) follows similar to (4.235) as any component of $\bar{\mathbf{Y}}'_{1a}$ is a linear combination of random variables including of components $\bar{\mathbf{X}}_{2ba}$ and $(\bar{\mathbf{X}}_{2da})^{1-\beta_{12}}$, and the vector $\bar{\mathbf{Y}}'_{1a}$ has totally $N_1 - N_2 + M_1$ components.

4.11 Proof for Theorem 4.2: Achievability

Note that for Theorem 4.2, the achievability in [20] is only tight when $M_1 \geq N_1$. In this section, we will present the achievability for Theorem 4.2 with $M_1 < N_1$. First, an useful lemma is present as follows.

Consider a multiple access channel with K signal antenna transmitters. The receiver has M antennas. The $M \times 1$ received signal vector Y is represented as follows

$$\mathbf{Y} = \sqrt{P} \sum_{k=1}^{K} \mathbf{H}_k \mathbf{X}_k + \sum_{m=1}^{M} \sqrt{P^{\alpha_m}} \mathbf{G}_m Z_m \tag{4.240}$$

Here, $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K$ the transmitted symbols, normalized to unit transmit power constraint. Z_m are i.i.d. Gaussian zero mean unit variance terms. The $\mathbf{H}_k, \mathbf{G}_n$ are $M \times 1$ generic vectors, i.e., generated from continuous distributions with bounded density, so that any M of them are linearly independent almost surely. All $\alpha_m \in [0, 1]$.

4.11.1 $N_1 + M_1 < N_2$

We first show the proof for the case $N_1 + M_1 < N_2$. In this case, the DoF region can be simplified as follows

$$d_1 \leq M_1 \tag{4.241}$$

$$d_1 + d_2 \leq N_2 \tag{4.242}$$

$$\frac{d_1}{M_1} + \frac{d_2}{N_2 - N_1 + M_1} \leq \frac{N_2 + \min(M_2 - N_1, N_2 - M_1)\beta_{12}}{N_2 - N_1 + M_1} \tag{4.243}$$

Note that (4.243) is a combination of (4.19) and (4.20). There are four corner points in the region. $(M_1, 0)$ and $(0, N_2)$ is trivial. The corner point $(d_1, d_2) = (M_1, \min[N_2 - M_1, N_1 - M_1 + \beta_{12}\min(M_2 - N_1, N_2 - M_1)])$ is achieved in [44]. We now show the proof for the following corner point,

$$(d_1, d_2) = (M_1\bar{d}, N_2 - M_1\bar{d}), \tag{4.244}$$

where $\bar{d} = \min(\frac{\min(M_2 - N_1, N_2 - M_1)\beta_{12}}{N_2 - N_1}, 1)$.

Mathematically, the transmitted signals are,

$$\mathbf{X}_1 = c_o\sqrt{P}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c \tag{4.245}$$

$$\begin{aligned}\mathbf{X}_2 =& c_1\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}_{2j}^p X_{2j}^p + c_2\sqrt{P^{1-\bar{d}}}\sum_{k=\min(N_2-M_1,M_2-N_1)+1}^{N_2}\mathbf{V}_{2k}^p X_{2k}^p \\ &+ c_3\sqrt{P}\sum_{i=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}_{2i}^c X_{2i}^c + c_4\sqrt{P}\sum_{m=\min(N_2-M_1,M_2-N_1)+1}^{N_2-M_1}\mathbf{V}_{2m}^c X_{2m}^c\end{aligned} \tag{4.246}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\bar{d}$ DoF, and the superscript 'c' is used to indicate that these codewords can be decoded by both receivers (common). X_{21}^c,

$X^c_{22}, \cdots, X^c_{2(N_2-M_1)}$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X^c_{2i} and X^c_{2m} carries $\bar{d}-\beta_{12}$ and $\bar{d}$ DoF, respectively. $X^p_{21}, \cdots, X^p_{N_2}$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X^p_{2j} and X^p_{2k} carries $1-\bar{d}+\beta_{12}$ and $1-\bar{d}$ DoF, respectively. The superscript 'p' is used to indicate that these are 'private', i.e., only decoded by the intended receiver. c_o, c_1, c_2, c_3 and c_4 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here $\mathbf{V}^c_{1l}$ are $M_1 \times 1$ generic unit vectors. $\mathbf{V}^c_{2i}$, $\mathbf{V}^c_{2m}$ and $\mathbf{V}^p_{2k}$ are $M_2 \times 1$ generic unit vectors. $\mathbf{V}^p_{2j}$ are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{G}}_{12}\begin{bmatrix}\mathbf{V}^p_{21} & \mathbf{V}^p_{22} & \cdots & \mathbf{V}^p_{2\min(N_2-M_1,M_2-N_1)}\end{bmatrix} = \mathbf{0} \tag{4.247}$$

The received signals are

$$\mathbf{Y}_1 = c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}^c_{1l}X^c_{1l} + (\hat{\mathbf{G}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{12})\mathbf{X}_2 + \mathbf{\Gamma}_1 \tag{4.248}$$

$$\begin{aligned}
&= c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}^c_{1l}X^c_{1l} + c_1\tilde{\mathbf{G}}_{12}\sqrt{P^{1-\bar{d}}}\sum_{j=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}^p_{2j}X^p_{2j} \\
&\quad + c_2\mathbf{G}_{12}\sqrt{P^{1-\bar{d}}}\sum_{k=\min(N_2-M_1,M_2-N_1)+1}^{N_2}\mathbf{V}^p_{2k}X^p_{2k} + c_3\mathbf{G}_{12}\sqrt{P}\sum_{i=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}^c_{2i}X^c_{2i} \\
&\quad + c_4\mathbf{G}_{12}\sqrt{P}\sum_{m=\min(N_2-M_1,M_2-N_1)+1}^{N_2-M_1}\mathbf{V}^c_{2m}X^c_{2m} + \mathbf{\Gamma}_1
\end{aligned} \tag{4.249}$$

$$\begin{aligned}
\mathbf{Y}_2 &= c_o\sqrt{P}\mathbf{G}_{21}\sum_{l=1}^{M_1}\mathbf{V}^c_{1l}X^c_{1l} + c_1\mathbf{G}_{22}\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}^p_{2j}X^p_{2j} \\
&\quad + c_2\mathbf{G}_{22}\sqrt{P^{1-\bar{d}}}\sum_{k=\min(N_2-M_1,M_2-N_1)+1}^{N_2}\mathbf{V}^p_{2k}X^p_{2k} + c_3\mathbf{G}_{22}\sqrt{P}\sum_{i=1}^{\min(N_2-M_1,M_2-N_1)}\mathbf{V}^c_{2i}X^c_{2i} \\
&\quad + c_4\mathbf{G}_{22}\sqrt{P}\sum_{m=\min(N_2-M_1,M_2-N_1)+1}^{N_2-M_1}\mathbf{V}^c_{2m}X^c_{2m} + \mathbf{\Gamma}_2
\end{aligned} \tag{4.250}$$

At Receiver 1, X_{1l}^c, X_{2i}^c and X_{2m}^c can be decoded (Lemma 3.1) by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise.

At the same time, Receiver 2 can decode X_{1l}^c, X_{2i}^c and X_{2m}^c by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise (Lemma 3.1). After X_{1l}^c, X_{2i}^c and X_{2m}^c are decoded and removed, all the remaining signals can then be decoded separately along N_2 interference-free dimensions. This concludes the proof for (4.244).

4.11.2 $N_1 + M_1 \geq N_2$ and $M_2 - N_1 \leq N_2 - M_1$

We now show the proof for the case $N_1 + M_1 \geq N_2$ and $M_2 - N_1 \leq N_2 - M_1$. In this case, the DoF region can be simplified as follows

$$d_1 \leq M_1 \tag{4.251}$$

$$d_1 + d_2 \leq N_2 \tag{4.252}$$

$$\frac{d_1}{M_1} + \frac{d_2}{N_2 - N_1 + M_1} \leq \frac{N_2 + (M_2 - N_1)\beta_{12}}{N_2 - N_1 + M_1} \tag{4.253}$$

Note that now (4.21) and (4.22) are redundant. There are four corner points in the region. $(M_1, 0)$ and $(0, N_2)$ is trivial. The corner point $(d_1, d_2) = (M_1, \min[N_2 - M_1, N_1 - M_1 + \beta_{12}(M_2 - N_1)])$ is achieved in [44]. We now show the proof for the following corner point,

$$(d_1, d_2) = (M_1\bar{d}, N_2 - M_1\bar{d}), \tag{4.254}$$

where $\bar{d} = \min(\frac{(M_2-N_1)\beta_{12}}{N_2-N_1}, 1)$.

Mathematically, the transmitted signals are,

$$\mathbf{X}_1 = c_o\sqrt{P}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c \tag{4.255}$$

$$\mathbf{X}_2 = c_1\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{M_2-N_1}\mathbf{V}_{2j}^p X_{2j}^p + c_2\sqrt{P^{1-\bar{d}}}\sum_{k=M_2-N_1+1}^{N_2}\mathbf{V}_{2k}^p X_{2k}^p$$

$$+ c_3\sqrt{P}\sum_{i=1}^{M_2-N_1}\mathbf{V}_{2i}^c X_{2i}^c + c_4\sqrt{P}\sum_{m=M_2-N_1+1}^{N_2-M_1}\mathbf{V}_{2m}^c X_{2m}^c \tag{4.256}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\bar{d}$ DoF. X_{21}^c, X_{22}^c, $\cdots$, $X_{2(N_2-M_1)}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{2i}^c and X_{2m}^c carries $\bar{d}-\beta_{12}$ and $\bar{d}$ DoF, respectively. X_{21}^p, $\cdots$, $X_{N_2}^p$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X_{2j}^p and X_{2k}^p carries $1-\bar{d}+\beta_{12}$ and $1-\bar{d}$ DoF, respectively. c_o, c_1, c_2, c_3 and c_4 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here $\mathbf{V}_{1l}^c$ are $M_1 \times 1$ generic unit vectors. $\mathbf{V}_{2i}^c$, $\mathbf{V}_{2m}^c$ and $\mathbf{V}_{2k}^p$ are $M_2 \times 1$ generic unit vectors. $\mathbf{V}_{2j}^p$ are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{G}}_{12}\left[\mathbf{V}_{21}^p \quad \mathbf{V}_{22}^p \quad \cdots \quad \mathbf{V}_{2(M_2-N_1)}^p\right] = \mathbf{0} \tag{4.257}$$

The received signals are

$$\mathbf{Y}_1 = c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + (\hat{\mathbf{G}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{12})\mathbf{X}_2 + \mathbf{\Gamma}_1 \tag{4.258}$$

$$\begin{aligned}
&= c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_1\tilde{\mathbf{G}}_{12}\sqrt{P^{1-\bar{d}}}\sum_{j=1}^{M_2-N_1}\mathbf{V}_{2j}^p X_{2j}^p + c_2\mathbf{G}_{12}\sqrt{P^{1-\bar{d}}}\sum_{k=M_2-N_1+1}^{N_2}\mathbf{V}_{2k}^p X_{2k}^p \\
&\quad + c_3\mathbf{G}_{12}\sqrt{P}\sum_{i=1}^{M_2-N_1}\mathbf{V}_{2i}^c X_{2i}^c + c_4\mathbf{G}_{12}\sqrt{P}\sum_{m=M_2-N_1+1}^{N_2-M_1}\mathbf{V}_{2m}^c X_{2m}^c + \mathbf{\Gamma}_1
\end{aligned} \tag{4.259}$$

$$\begin{aligned}
\mathbf{Y}_2 &= c_o\sqrt{P}\mathbf{G}_{21}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_1\mathbf{G}_{22}\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{M_2-N_1}\mathbf{V}_{2j}^p X_{2j}^p \\
&\quad + c_2\mathbf{G}_{22}\sqrt{P^{1-\bar{d}}}\sum_{k=M_2-N_1+1}^{N_2}\mathbf{V}_{2k}^p X_{2k}^p + c_3\mathbf{G}_{22}\sqrt{P}\sum_{i=1}^{M_2-N_1}\mathbf{V}_{2i}^c X_{2i}^c \\
&\quad + c_4\mathbf{G}_{22}\sqrt{P}\sum_{m=M_2-N_1+1}^{N_2-M_1}\mathbf{V}_{2m}^c X_{2m}^c + \mathbf{\Gamma}_2
\end{aligned} \tag{4.260}$$

At Receiver 1, X_{1l}^c, X_{2i}^c and X_{2m}^c can be decoded (Lemma 3.1) by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise.

At the same time, Receiver 2 can decode X_{1l}^c, X_{2i}^c and X_{2m}^c by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise (Lemma 3.1). After X_{1l}^c, X_{2i}^c and X_{2m}^c are decoded and removed, all the remaining signals can then be decoded separately along N_2 interference-free dimensions. This concludes the proof for (4.254).

4.11.3 Case $N_1 + M_1 \geq N_2$ and $M_2 - N_1 > N_2 - M_1$:

We now show the proof for the case $N_1 + M_1 \geq N_2$ and $M_2 - N_1 > N_2 - M_1$. We further divide the case into three sub-cases according to the value of β_{12}.

$0 \leq \beta_{12} \leq \frac{N_2-N_1}{\min(M_2-N_1,N_2)}$

When $0 \leq \beta_{12} \leq \frac{N_2-N_1}{\min(M_2-N_1,N_2)}$, the DoF region can be simplified as follows

$$d_1 \leq M_1 \tag{4.261}$$

$$d_1 + d_2 \leq N_2 \tag{4.262}$$

$$d_1 + \frac{d_2}{2} \leq \frac{1}{2}[M_1 + N_1 + (N_2 - M_1)\beta_{12}] \tag{4.263}$$

$$\frac{d_1}{M_1} + \frac{d_2}{N_2+M_1-N_1} \leq \frac{N_2}{N_2+M_1-N_1} + \left[\frac{N_2 - M_1}{N_2+M_1-N_1} + \frac{(M_1+N_1-N_2)\min(M_1, M_1+M_2-N_1-N_2)}{M_1(N_2 + M_1 - N_1)}\right]\beta_{12} \tag{4.264}$$

Note that now (4.19) is redundant. There are five corner points for this DoF region. $(M_1, 0)$ and $(0, N_2)$ is trivial. The corner point $(d_1, d_2) = (M_1, \min[N_2-M_1, N_1-M_1+\beta_{12}(N_2-M_1)])$ is achieved in [44]. We first show the proof for the following corner point (intersection point of (4.262) and (4.264)),

$(d_1, d_2) =$

$$\left(M_1(\bar{d}-\frac{\min(M_1+M_2-N_1-N_2, M_1)\beta_{12}}{M_1}), N_2-M_1(\bar{d}-\frac{\min(M_1+M_2-N_1-N_2, M_1)\beta_{12}}{M_1})\right) \tag{4.265}$$

where $\bar{d} = \frac{\min(M_2-N_1,N_2)\beta_{12}}{N_2-N_1}$.

Mathematically, each time slot, the transmitted signals are,

$$\mathbf{X}_1 = c_o\sqrt{P}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c \tag{4.266}$$

$$\begin{aligned}\mathbf{X}_2 =& c_1\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^p X_{2j}^p + c_2\sqrt{P^{1-\bar{d}}}\sum_{k=\min(M_2-N_1,N_2)+1}^{N_2}\mathbf{V}_{2k}^p X_{2k}^p \\ &+ c_3\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c\end{aligned} \tag{4.267}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\bar{d}-\frac{(M_1+M_2-N_1-N_2)^+\beta_{12}}{M_1}$ DoF. X_{21}^c, X_{22}^c, $\cdots$, $X_{2(N_2-M_1)}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{2i}^c carries $\bar{d}-\beta_{12}$ DoF. X_{21}^p, $\cdots$, $X_{N_2}^p$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X_{2j}^p and X_{2k}^p carries $1-\bar{d}+\beta_{12}$ and $1-\bar{d}$ DoF, respectively. c_o, c_1, c_2 and c_3 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here $\mathbf{V}_{1l}^c$ are $M_1 \times 1$ generic unit vectors. $\mathbf{V}_{2i}^c$ and $\mathbf{V}_{2k}^p$ are $M_2 \times 1$ generic unit vectors. $\mathbf{V}_{2j}^p$ are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{G}}_{12}\begin{bmatrix}\mathbf{V}_{21}^p & \mathbf{V}_{22}^p & \cdots & \mathbf{V}_{2\min(M_2-N_1,N_2)}^p\end{bmatrix} = \mathbf{0} \tag{4.268}$$

The received signals are

$$\begin{aligned}\mathbf{Y}_1 =& c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + (\hat{\mathbf{G}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{12})\mathbf{X}_2 + \mathbf{\Gamma}_1 \\ =& c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_1\tilde{\mathbf{G}}_{12}\sqrt{P^{1-\bar{d}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^p X_{2j}^p\end{aligned} \tag{4.269}$$

$$+ c_2\mathbf{G}_{12}\sqrt{P^{1-\bar{d}}} \sum_{k=\min(M_2-N_1,N_2)+1}^{N_2} \mathbf{V}_{2k}^{p}X_{2k}^{p} + c_3\mathbf{G}_{12}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^{c}X_{2i}^{c} + \mathbf{\Gamma}_1 \quad (4.270)$$

$$\begin{aligned}\mathbf{Y}_2 =& c_o\sqrt{P}\mathbf{G}_{21}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^{c}X_{1l}^{c} + c_1\mathbf{G}_{22}\sqrt{P^{1-\bar{d}+\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^{p}X_{2j}^{p} \\ &+ c_2\mathbf{G}_{22}\sqrt{P^{1-\bar{d}}}\sum_{k=\min(M_2-N_1,N_2)+1}^{N_2}\mathbf{V}_{2k}^{p}X_{2k}^{p} + c_3\mathbf{G}_{22}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^{c}X_{2i}^{c} + \mathbf{\Gamma}_2 \quad (4.271)\end{aligned}$$

At Receiver 1, X_{1l}^c and X_{2i}^c can be decoded (Lemma 3.1) by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise.

At the same time, Receiver 2 can decode X_{1l}^c and X_{2i}^c by treating the rest N_2 streams carrying X_{2j}^p and X_{2k}^p as white noise (Lemma 3.1). After X_{1l}^c and X_{2i}^c are decoded and removed, all the remaining signals can then be decoded separately along N_2 interference-free dimensions. This concludes the proof for (4.265).

Then we show the proof for the following corner point (intersection point of (4.263) and (4.264)),

$$\begin{aligned}(d_1, d_2) =& (M_1 - \min(M_1, M_1{+}M_2{-}N_1{-}N_2)\beta_{12}, \\ & N_1 - M_1 + \min(M_1, M_1{+}M_2{-}N_1{-}N_2)\beta_{12} + \min(N_2, M_2{-}N_1)\beta_{12}) . \quad (4.272)\end{aligned}$$

Mathematically, each time slot, the transmitted signals are,

$$\mathbf{X}_1 = c_o\sqrt{P}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^{c}X_{1l}^{c} \quad (4.273)$$

$$\mathbf{X}_2 = c_1\sqrt{P^{\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^{p}X_{2j}^{p} + c_2\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^{c}X_{2i}^{c} \quad (4.274)$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\frac{M_1-\min(M_1,M_1+M_2-N_1-N_2)\beta_{12}}{M_1}$ DoF. X_{21}^c, X_{22}^c, $\cdots$, $X_{2(N_2-M_1)}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{2i}^c carries $\frac{N_1-M_1+\min(M_1,M_1+M_2-N_1-N_2)\beta_{12}}{N_2-M_1}$ DoF. $X_{21}^p, \cdots, X_{2(N_2-M_1)}^p$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X_{2j}^p carries β_{12} DoF. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here $\mathbf{V}_{1l}^c$ are $M_1 \times 1$ generic unit vectors. $\mathbf{V}_{2i}^c$ are $M_2 \times 1$ generic unit vectors. $\mathbf{V}_{2j}^p$ are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{G}}_{12}\begin{bmatrix}\mathbf{V}_{21}^p & \mathbf{V}_{22}^p & \cdots & \mathbf{V}_{2\min(M_2-N_1,N_2)}^p\end{bmatrix} = \mathbf{0} \tag{4.275}$$

The received signals are

$$\begin{aligned}\mathbf{Y}_1 =& c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + (\hat{\mathbf{G}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{12})\mathbf{X}_2 + \mathbf{\Gamma}_1 & (4.276)\\ =& c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_2\mathbf{G}_{12}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c + O(1) + \mathbf{\Gamma}_1 & (4.277)\end{aligned}$$

$$\begin{aligned}\mathbf{Y}_2 =& c_o\sqrt{P}\mathbf{G}_{21}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_1\mathbf{G}_{22}\sqrt{P^{\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^p X_{2j}^p\\ &+ c_2\mathbf{G}_{22}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c + \mathbf{\Gamma}_2 & (4.278)\end{aligned}$$

At Receiver 1, since X_{2j}^p remains at the noise floor level, X_{1l}^c and X_{2i}^c can be jointly decoded (Lemma 3.1) as a MAC channel.

At the same time, Receiver 2 can decode X_{1l}^c and X_{2i}^c by treating the rest $\min(M_2-N_1, N_2)$ streams carrying X_{2j}^p as white noise (Lemma 3.1). After X_{1l}^c and X_{2i}^c are decoded and removed, all the remaining signals can then be decoded separately along $\min(M_2-N_1, N_2)$ interference-free dimensions. This concludes the proof for (4.272).

Case $\frac{N_2-N_1}{\min(M_2-N_1,N_2)} < \beta_{12} < \frac{N_2-N_1}{N_2-M_1}$**:**

When $\frac{N_2-N_1}{\min(M_2-N_1,N_2)} < \beta_{12} < \frac{N_2-N_1}{N_2-M_1}$, the DoF region can be simplified as follows

$$d_1 \leq M_1 \tag{4.279}$$

$$d_1 + d_2 \leq N_2 \tag{4.280}$$

$$d_1 + \frac{d_2}{2} \leq \frac{1}{2}[M_1 + N_1 + (N_2 - M_1)\beta_{12}] \tag{4.281}$$

Note that now (4.19) and (4.22) is redundant. There are four corner points for this DoF region. $(M_1, 0)$ and $(0, N_2)$ is trivial. The corner point $(d_1, d_2) = (M_1, \min[N_2 - M_1, N_1 - M_1 + \beta_{12}(N_2 - M_1)])$ is achieved in [44]. We now show the proof for the following corner point (intersection point of (4.280) and (4.281)),

$$(d_1, d_2) = (M_1 + N_1 - N_2 + (N_2 - M_1)\beta_{12}, 2N_2 - M_1 - N_1 - (N_2 - M_1)\beta_{12}) . \tag{4.282}$$

Mathematically, each time slot, the transmitted signals are,

$$\mathbf{X}_1 = c_o\sqrt{P}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c \tag{4.283}$$

$$\mathbf{X}_2 = c_1\sqrt{P^{\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^p X_{2j}^p + c_2\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c \tag{4.284}$$

Here X_{11}^c, X_{12}^c, $\cdots$, $X_{1M_1}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{1l}^c carries $\frac{M_1+N_1-N_2+(N_2-M_1)\beta_{12}}{M_1}$ DoF. X_{21}^c, X_{22}^c, $\cdots$, $X_{2(N_2-M_1)}^c$ are independent Gaussian codewords from unit power codebooks that can be decoded by both receivers. Each X_{2i}^c carries $1-\beta_{12}$ DoF. X_{21}^p, $\cdots$, $X_{\min(M_2-N_1,N_2)}^p$ are independent Gaussian codewords from unit power codebooks that are to be decoded only by User 2. Each X_{2j}^p carries $\frac{N_2-N_1}{\min(M_2-N_1,N_2)}$ DoF. c_o, c_1 and c_2 are scaling factors, $O(1)$ in P, chosen to ensure that the transmit power constraint is satisfied.

Here $\mathbf{V}_{1l}^c$ are $M_1 \times 1$ generic unit vectors. $\mathbf{V}_{2i}^c$ are $M_2 \times 1$ generic unit vectors. $\mathbf{V}_{2j}^p$ are $M_2 \times 1$ unit vectors chosen so that

$$\hat{\mathbf{G}}_{12} \left[\mathbf{V}_{21}^p \quad \mathbf{V}_{22}^p \quad \cdots \quad \mathbf{V}_{2\min(M_2-N_1,N_2)}^p \right] = \mathbf{0} \tag{4.285}$$

The received signals are

$$\mathbf{Y}_1 = c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + (\hat{\mathbf{G}}_{12} + \sqrt{P^{-\beta_{12}}}\tilde{\mathbf{G}}_{12})\mathbf{X}_2 + \mathbf{\Gamma}_1 \tag{4.286}$$

$$= c_o\sqrt{P}\mathbf{G}_{11}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_2\mathbf{G}_{12}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c + O(1) + \mathbf{\Gamma}_1 \tag{4.287}$$

$$\mathbf{Y}_2 = c_o\sqrt{P}\mathbf{G}_{21}\sum_{l=1}^{M_1}\mathbf{V}_{1l}^c X_{1l}^c + c_1\mathbf{G}_{22}\sqrt{P^{\beta_{12}}}\sum_{j=1}^{\min(M_2-N_1,N_2)}\mathbf{V}_{2j}^p X_{2j}^p$$
$$+ c_2\mathbf{G}_{22}\sqrt{P}\sum_{i=1}^{N_2-M_1}\mathbf{V}_{2i}^c X_{2i}^c + \mathbf{\Gamma}_2 \tag{4.288}$$

At Receiver 1, since X_{2j}^p remains at the noise floor level, X_{1l}^c and X_{2i}^c can be jointly decoded (Lemma 3.1) as a MAC channel.

At the same time, Receiver 2 has enough antennas to jointly decode all the streams as a MAC channel. Specifically, a rate tuple

$$(R_{11}^c, \cdots, R_{1M_1}^c, R_{21}^c, \cdots, R_{2(N_2-M_1)}^c, R_{21}^p, \cdots, R_{2\min(M_2-N_1,N_2)}^p)$$

is achievable if the following inequalities are satisfied.

$$\sum_{l\in\mathcal{U}_1} R_{1l}^c + \sum_{i\in\mathcal{U}_2} R_{2i}^c + \sum_{j\in\mathcal{U}_3} R_{2j}^p \leq$$
$$I(\{X_{1l}^c, X_{2i}^c, X_{2j}^p,\ \forall l\in\mathcal{U}_1, i\in\mathcal{U}_2, j\in\mathcal{U}_3\}; Y_2|\{X_{1l}^c, X_{2i}^c, X_{2j}^p,\ \forall l\in\mathcal{U}_1^c, i\in\mathcal{U}_2^c, j\in\mathcal{U}_3^c\}), \tag{4.289}$$

$$\forall\, \mathcal{U}_1 \subseteq \mathcal{I}_{M_1},\ \mathcal{U}_2 \subseteq \mathcal{I}_{N_2-M_1},\ \mathcal{U}_3 \subseteq \{1, 2, \cdots, \min(M_2 - N_1, N_2)\},$$

where $\mathcal{U}_k^c$ is the complementary set of $\mathcal{U}_k$, $\forall k \in \{1, 2, 3\}$.

$$\begin{aligned}
&I(\{X_{1l}^c, X_{2i}^c, X_{2j}^p,\ \forall l \in \mathcal{U}_1, i \in \mathcal{U}_2, j \in \mathcal{U}_3\}; Y_2|\{X_{1l}^c, X_{2i}^c, X_{2j}^p\ \forall l \in \mathcal{U}_1^c, i \in \mathcal{U}_2^c, j \in \mathcal{U}_3^c\}) \\
&= h(Y_2|\{X_{1l}^c, X_{2i}^c, X_{2j}^p,\ \forall l \in \mathcal{U}_1^c, i \in \mathcal{U}_2^c, j \in \mathcal{U}_3^c\}) \\
&- h(Y_2|\{X_{1l}^c, X_{2i}^c, X_{2j}^p, \forall l \in \mathcal{I}_{M_1},\ i \in \mathcal{I}_{N_2-M_1},\ j \in \{1, 2, \cdots, \min(M_2 - N_1, N_2)\}) \quad (4.290) \\
&= [|\mathcal{U}_1| + |\mathcal{U}_2| + \min(N_2 - |\mathcal{U}_1| - |\mathcal{U}_2|, |\mathcal{U}_3|)\beta_{12}] \log P + o(\log P) \quad (4.291)
\end{aligned}$$

$|\mathcal{U}_k|$ is the cardinality of $\mathcal{U}_k$. (4.291) follows from Lemma 3 in [11].

From (4.291), we obtain the achievable DoF region,

$$\begin{aligned}
&\sum_{l \in \mathcal{U}_1} d_{1l}^c + \sum_{i \in \mathcal{U}_2} d_{2i}^c + \sum_{j \in \mathcal{U}_3} d_{2j}^p \leq |\mathcal{U}_1| + |\mathcal{U}_2| + \min(N_2 - |\mathcal{U}_1| - |\mathcal{U}_2|, |\mathcal{U}_3|)\beta_{12}, \quad (4.292) \\
&\forall\, \mathcal{U}_1 \subseteq \mathcal{I}_{M_1},\ \mathcal{U}_2 \subseteq \mathcal{I}_{N_2-M_1},\ \mathcal{U}_3 \subseteq \{1, 2, \cdots, \min(M_2 - N_1, N_2)\}
\end{aligned}$$

Thus it can be verified that the DoF tuple ($d_{1l}^c = \frac{M_1+N_1-N_2+(N_2-M_1)\beta_{12}}{M_1}, d_{2i}^c = 1 - \beta_{12}, d_{2j}^p = \frac{N_2-N_1}{\min(M_2-N_1,N_2)}, \forall\ l \in \mathcal{U}_1, i \in \mathcal{U}_2, j \in \mathcal{U}_3$) satisfies (4.292), i.e.,

$$\begin{aligned}
&|\mathcal{U}_1| \frac{M_1+N_1-N_2+(N_2-M_1)\beta_{12}}{M_1} + |\mathcal{U}_2|(1 - \beta_{12}) + |\mathcal{U}_3| \frac{N_2 - N_1}{\min(M_2 - N_1, N_2)} \\
&\leq |\mathcal{U}_1| + |\mathcal{U}_2| + \min(N_2 - |\mathcal{U}_1| - |\mathcal{U}_2|, |\mathcal{U}_3|)\beta_{12}, \quad (4.293) \\
&\forall\, \mathcal{U}_1 \subseteq \mathcal{I}_{M_1},\ \mathcal{U}_2 \subseteq \mathcal{I}_{N_2-M_1},\ \mathcal{U}_3 \subseteq \{1, 2, \cdots, \min(M_2 - N_1, N_2)\}
\end{aligned}$$

This concludes the proof for (4.282).

Case $\frac{N_2-N_1}{N_2-M_1} \leq \beta_{12} \leq 1$:

When $\frac{N_2-N_1}{N_2-M_1} \leq \beta_{12} \leq 1$, the DoF region can be simplified as follows

$$d_1 \leq M_1 \tag{4.294}$$

$$d_1 + d_2 \leq N_2 \tag{4.295}$$

There are three corner points for this DoF region. $(M_1, 0)$ and $(0, N_2)$ is trivial. The corner point $(d_1, d_2) = (M_1, N_2 - M_1)$ is achieved in [44]. Note that now this DoF region is the same as with perfect CSIT.

4.12 Summary

In this chapter we explored the DoF region of a two-user MIMO IC with arbitrary antenna configurations and arbitrary partial CSIT levels. The aligned image set approach and sum-set inequalities was applied to obtain the tight outer bound. This chapter not only bridges the extremes of known DoF results between perfect and no CSIT, but also show that several new ideas of achievability such as elevated multiplexing and jointly using the signal levels and time slots, is optimal.

Chapter 5

On the Synergistic Benefits of Reconfigurable Antennas and Partial CSIT for the MIMO IC

None-perfect CSIT settings allows transmitter to acquire partial CSIT or delayed CSIT. When only delayed CSIT is available [33, 34], transmitter has the ability to reconstruct all the interference seen in previous symbols so that interference will be aligned into a smaller space across multiple channel uses. If partial CSIT is available, transmitter can employ partial zero-forcing [44, 20] and partial interference alignment schemes to achieve partial DoF gains. In addition, a mixed CSIT setting (with both partial and delayed CSIT) has also been considered in [16].

Reconfigurable antennas are antenna arrays that are capable of switching between a number of preset modes, each corresponding to an independent set of channels to all receivers. It has been shown in [18, 39] that with reconfigurable antennas it is possible to achieve interference alignment for broadcast channel even when no other channel knowledge is available. The scheme is termed *blind interference alignment*.

Blind interference alignment (BIA) schemes create and exploit channel coherence patterns without the knowledge of channel realizations at transmitters, while beamforming schemes rely primarily on channel knowledge available to the transmitters without regard to channel coherence patterns. In order to explore the compatibility of these disparate ideas and the possibility of synergistic gains, this chapter explores the degrees of freedom (DoF) of the two-user $(M_1 \times N_1)(M_2 \times N_2)$ multiple-input multiple-output (MIMO) interference channel (IC) where Transmitter 1 is equipped with reconfigurable antennas and has no channel knowledge while Transmitter 2 has partial channel knowledge but no reconfigurable antennas. Coding schemes are proposed that jointly exploit partial channel knowledge and reconfigurable antennas, demonstrating synergistic DoF gains over what is achievable with either BIA or beamforming by itself. In addition, with the aid of a new outer bound, a complete characterization of the DoF region is obtained for $N_1 < M_2 \leq N_2$.

5.1 System Model

Consider a 2-user MIMO interference channel, where there are N_k antennas at the k^{th} receiver, Transmitter 1 has M_1 reconfigurable antennas and each can switch among N_1 preset modes, and Transmitter 2 has M_2 antennas. It is assumed $M_1 < N_1 < \min\{M_2, N_2\}$. Each transmitter sends an independent message to its corresponding receiver. It is assumed only Transmitter 2 has imperfect CSIT to Receiver 1 and all other CSIT is not known. All receivers have perfect channel knowledge. At time slot $t \in \mathbb{Z}^+$, the channel input-output equations are given by

$$\boldsymbol{y}_1(t) = \mathbf{H}_{11}^{m}(t)\boldsymbol{x}_1(t) + \mathbf{H}_{12}^{\dagger}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_1(t), \tag{5.1}$$

$$\boldsymbol{y}_2(t) = \mathbf{H}_{21}^{m}(t)\boldsymbol{x}_1(t) + \mathbf{H}_{22}(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_2(t), \tag{5.2}$$

Here, $\boldsymbol{x}_k(t) = [x_{k1}(t), x_{k2}(t), ..., x_{kM_k}(t)]^T \in \mathbb{R}^{M_k \times 1}$ is the real signal vector sent from Transmitter k, which satisfies an average power constraint $\mathbb{E}(\|\boldsymbol{x}_k(t)\|^2) \leq P$. $\boldsymbol{y}_k(t) \in \mathbb{R}^{N_k \times 1}$ is the received signal vector at receiver k. $\boldsymbol{z}_k(t) \in \mathbb{R}^{N_k \times 1}$ is the i.i.d. real additive white Gaussian noise (AWGN) at Receiver k, each entry of which is an i.i.d. Gaussian random variable with zero-mean and unit-variance.

$\mathbf{H}_{k1}^{m}(t) \in \mathbb{R}^{N_k \times M_1}$ denote the $N_k \times M_1$ channel matrix associated with M_1 antennas of Transmitter 1 at the mode m and at time slot t to Receiver k, where $m \in \mathcal{I}_{N_1}$. $\mathbf{H}_{22}(t) \in \mathbb{R}^{N_2 \times M_2}$ is the channel matrix from Transmitter 2 to Receiver 2. In addition, for $\mathbf{H}_{12}^{\dagger}(t)$ we assume that

$$\mathbf{H}_{12}^{\dagger}(t) = \hat{\mathbf{H}}_{12}(t) + \sqrt{P^{-\beta}}\tilde{\mathbf{H}}_{12}(t) \tag{5.3}$$

where $\hat{\mathbf{H}}_{12}(t)$ is the estimated channel known to Transmitter 2, while $\tilde{\mathbf{H}}_{12}(t)$ is the estimation error that is unknown to Transmitter 2. We assume each coefficient of $\hat{\mathbf{H}}_{12}(t)$, $\tilde{\mathbf{H}}_{12}(t)$ and all other channel matrices are drawn from continuous distributions, independent of each other. Each coefficient of estimated channel matrix and estimation error matrix is also assumed to be zero mean with unit variance. In addition, the estimation error is assumed to be bounded away from infinity, i.e., there exists constant $0 < \Delta < \infty$ such that $\left|\tilde{H}_{12}(t)\right| \in [0, \Delta]$, $\forall t \in \mathbb{N}$. The parameter β measures the quality of the current channel estimation. If $\beta = 0$, then it corresponds to the case when there is no current CSI. If $\beta \geq 1$, then it corresponds to the case that the current CSI is as good as perfect (for DoF). It is also assumed that the coherence times of all the channels are long enough so that the channels stay constant across N_1 time slots.

5.2 Main Results

The following theorem states the main result of this paper, i.e., the existence of the synergistic benefit of reconfigurable antennas and partial CSIT.

Theorem 5.1. *(Synergistic DoF gain) Consider the 2-user MIMO interference channel with reconfigurable antennas at Tx1, partial CSIT at Tx2 and arbitrary number of antennas at each* *node. As long as reconfigurable antennas and partial CSIT are both useful, i.e., comparing with no CSIT one can achieve more DoF with either reconfigurable antennas or partial CSIT, then by jointly exploiting these two factors, the synergistic DoF gain always exists over what is achievable with either reconfigurable antennas or partial CSIT by itself. In other words, when* $M_1 < N_1 < \min(M_2, N_2)$*, the following DoF tuple can be achievable*

$$(d_1, d_2) = (M_1, \tfrac{(N_1-M_1)\min(M_2,N_2)}{N_1} + \beta \tfrac{M_1(\min(M_2,N_2)-N_1)}{N_1}) \tag{5.4}$$

To prove Theorem 5.1, a toy example is first presented in Section 5.3.1 to highlight the key idea of the proposed achievable scheme, that is, compare to the BIA scheme with no CSIT, Tx2 now can always send extra signals in its interference null space, as long as Rx2 has enough antennas to separate the all signals. Thus the synergistic DoF gain will always exists. A complete proof of Theorem 5.1 is presented in Section 5.4.

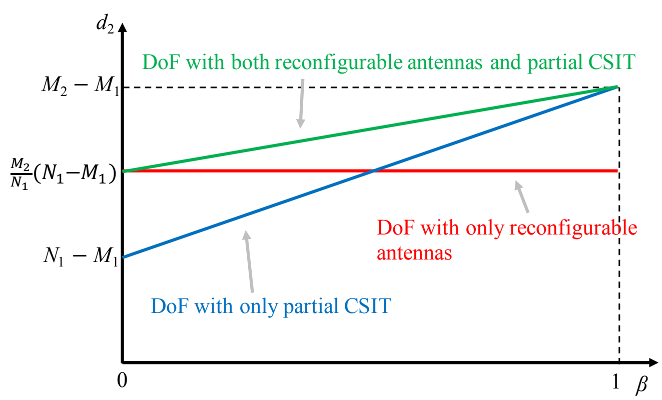

Figure 5.1: The optimal DoF achieved by User 2 when $d_1 = M_1$ with $M_1{<}N_1{<}M_2{\leq}N_2$.

Remark 5.1. *The main contribution of Theorem 5.1 can be summarized in Figure. 5.1 for the case* $M_1{<}N_1{<}M_2{\leq}N_2$*. Consider the optimal DoF achieved by User 2 without hurting User 1 under various assumptions. First,* $N_1 - M_1$ *DoF can be achieved when no channel knowledge is available [46]. If with only partial CSIT, without reconfigurable antennas, as a function of* β*,* $d_2 = N_1 - M_1 + (M_2 - N_1)\beta$ *is achievable [44] which is better than with no CSIT, i.e., partial CSIT is useful. Then, by using reconfigurable antennas alone, i.e., without* *any other channel knowledge, we know* $\frac{M_2(N_1-M_1)}{N_1}$ *DoF are optimally achieved [27] by User 2, i.e., reconfigurable antennas are also useful. Therefore, now by combining reconfigurable antennas and partial CSIT, User 2 can achieve* $\frac{(N_1-M_1)M_2}{N_1} + \beta\frac{M_1(M_2-N_1)}{N_1}$*. The synergistic DoF gain is shown in Figure. 5.1. One can also verify the synergistic DoF gain for the case* $M_1<N_1<N_2<M_2$ *in a same manner.*

Theorem 5.2. *(a DoF outer bound) For the 2-user MIMO interference channel with reconfigurable antennas and partial CSIT, if* $N_1 < \min(M_2, N_2)$*, the DoF region satisfies*

$$\mathcal{D} \in \mathcal{D}_{out} \triangleq \left\{ (d_1, d_2) \in \mathbb{R}_2^+, \begin{array}{lccc} L_1: & d_1 & \leq & \min(M_1, N_1), \\ L_2: & d_1 + d_2 & \leq & \min[N_1 + N_2, \max(M_1, N_2), M_2], \\ L_3: & \frac{d_1}{N_1} + \frac{d_2}{M_2} & \leq & 1 + \frac{\beta \min(M_1,N_1)(M_2-N_1)}{N_1 M_2} \end{array} \right\} \quad (5.5)$$

If $N_1 < M_2 \leq N_2$*, the DoF outer bound given in* (5.5) *is tight, i.e.,* $\mathcal{D} = \mathcal{D}_{out}$*.*

The outer bounds L_1 and L_2 are trivial since they are also the outer bounds for MIMO IC with perfect CSIT. Then we only need to prove L_3. Note that in our setting, the coherence time for each link may larger than 1, so the AIS approach [8] previously used to obtain the outer bounds cannot be used here. To prove L_3, the approach we use is based on the compound setting argument, i.e., imposing a compound setting for channel uncertainty does not decrease the capacity of the original channel. The proof is presented in Section 5.5.

Remark 5.2. *Theorem 5.2 indicates that the scheme proposed in Theorem 5.1 is optimal for the case* $M_1<N_1<M_2\leq N_2$*. To completely characterize the DoF region of the 2-user MIMO interference channel, one need to solve the remaining case in Theorem 5.1 that* $M_1<N_1<N_2<M_2$*. Unlike the solved case where Tx2 can fully use its inference null space since Rx2 has abundant receive antennas, for this case the extra signals transmitted through the inference null space is limited by the number of antennas at Rx2, i.e., Tx2 cannot efficiently use its inference null space with the scheme proposed in Theorem 5.1 . This is the main reason that the case* $M_1<N_1<N_2<M_2$ *remains open. Now in the following theorem,*

we introduce a new scheme that allowing Tx2 to better utilize its interference null space by splitting its transmitted signals into a layered structure. Thus it can achieve more DoF than the scheme in Theorem 5.1 in certain conditions and has the potential to solve this remaining case.

Theorem 5.3. *For the 2-user MIMO interference channel with reconfigurable antennas and partial CSIT and $M_1 < N_1 < N_2 < M_2$, when User 1 achieves its interference-free DoF, i.e., $d_1 = M_1$, the DoF value achieved by User 2 can be further improved to*

$$d_2 = \min[\tfrac{(N_1-M_1)N_2}{N_1} + \beta\tfrac{M_1(N_2-M_1)}{N_1}, N_2 - M_1] \tag{5.6}$$

if the dimension of the interference-null space at Transmitter 2, i.e., $\dim[\text{null}(\mathbf{H}_{12}^{\dagger})] = M_2 - N_1$ *satisfied the following condition:*

$$M_2 - N_1 \geq \begin{cases} \max(N_2 - M_1, N_1), & \text{if } \frac{N_2-N_1}{N_2-M_1} \leq \frac{1}{2} \\ \max(N_2 - M_1, N_2), & \text{if } \frac{N_2-N_1}{N_2-M_1} > \frac{1}{2} \end{cases} \tag{5.7}$$

The condition (5.7) indicates that in order for transmitted signals at Tx2 to have a layered structure, Tx2 need to have enough dimensions for its interference null space. A toy example is first presented in Section 5.3.2 to highlight the key idea of this new scheme. A complete proof for Theorem 5.3 is presented in Section 5.6.

5.3 Examples and Discussions on Synergistic Benefits

As mentioned previously, the most interesting aspects of the reconfigurable antennas and partial CSIT problem are the synergistic DoF gains. The key idea to obtain this gain is to split the signal space into two part, one subspace corresponds to the case with only reconfigurable antennas and no CSIT, while another subspace corresponds to the case with perfect CSIT. When jointly exploiting reconfigurable antennas and partial channel knowledge, we can achieve more DoF than we could by exploiting each of them individually. Representative toy examples of this phenomenon are presented next.

5.3.1 Example 1: Separable Signal Space by Antenna Dimension

As a special case in Theorem 5.1, let us start with the setting $(M_1, M_2, N_1, N_2) = (1, 3, 2, 3)$ in Figure 5.2, where User 1 has reconfigurable antenna and achieves $d_1 = 1$, i.e., his maximum DoF. Suppose Tx2 has partial CSIT level $\beta = \frac{1}{4}$ for his interference carrying link to Rx1. Remarkably, as shown in Figure 5.2, User 2 can achieve 1.625 DoF without hurting User 1.

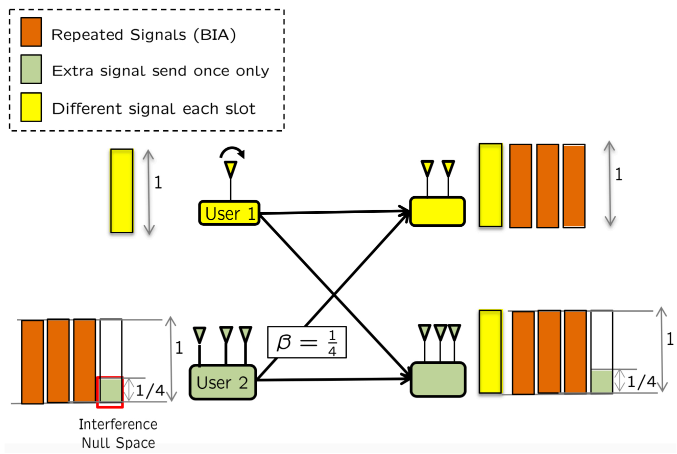

Figure 5.2: Illustration of achievable scheme for Example 1.

To accomplish this, we first introduce the scheme based on BIA that can achieve $d_2 = 1.5$ with only reconfigurable antennas [27]. It operates over two channel uses. During the first channel use, User 1 sends 1 symbol, User 2 sends 3 symbols (signal subspace denoted by orange color in Figure 5.2). Rx1 sees as many linear combinations of these symbols as the number of receive antennas. During the second channel use, User 1 switches his antenna mode and sends a new symbol, while User 2 repeats the same 3 symbols. The repetition of symbols aligns interference at Rx1, because only 2 linear combinations of the 3 symbols are seen over 2 channel uses, and Rx1 is able to subtract the output of one channel use from the output of the other channel use to eliminate interference, leaving it with two interference free observations of the 2 desired symbols which can therefore be resolved. The reconfigurable antenna at Tx1 is important to make sure that the two desired symbols do not align with each other. User 2 has enough receive antennas to decode all symbols. Thus, User 1 achieves 1 DoF and User 2 achieves 1.5 DoF.

Now, based on the BIA scheme, with partial CSIT, Tx2 can send an additional stream with $\frac{1}{4}$ DoF (signal subspace denoted by green color in Figure 5.2) at first channel use. This stream will not be heard by Rx1, i.e., through the null-space of the estimated channel $\hat{\mathbf{H}}_{12}$, within which Transmitter 2 must not exceed the power level $\bar{P}^{\frac{1}{4}}$. So that User 1 will still achieve 1 DoF, meanwhile this additional stream can be decoded by Rx2 since it still has enough antennas. It turns out that User 2 can achieve 3.25 DoF over two channel uses, and $d_2 = 1.625$ is achieved per channel use without hurting User 1. And according to Theorem 5.2, this is also the optimal DoF for this channel.

It is also known that with only partial CSIT and no reconfigurable antenna, User 2 can only achieve 1.25 DoF [44] without hurting User 1. Therefore, for this example, we obtained a $\frac{1}{8}$ DoF gain by using both reconfigurable antennas and partial CSIT.

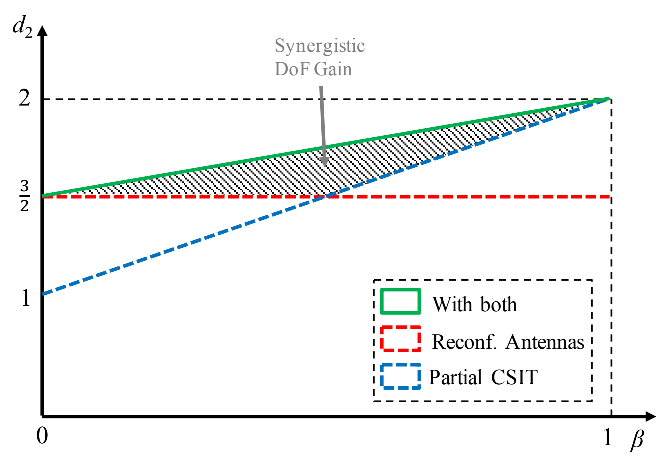

Figure 5.3: Synergistic DoF gain for Example 1.

By extending this scheme to the arbitrary partial CSIT level, i.e., $0 \leq \beta \leq 1$, the DoF achieved by User 2 as a function of β is $d_2 = 1.5 + \frac{\beta}{2}$ (as shown in Figure 5.3). To emphasize the synergistic benefit, the DoF achieved with only partial CSIT and with only reconfigurable antennas are also shown in Figure 5.3.

5.3.2 Example 2: Separable Signal Space by Power Level

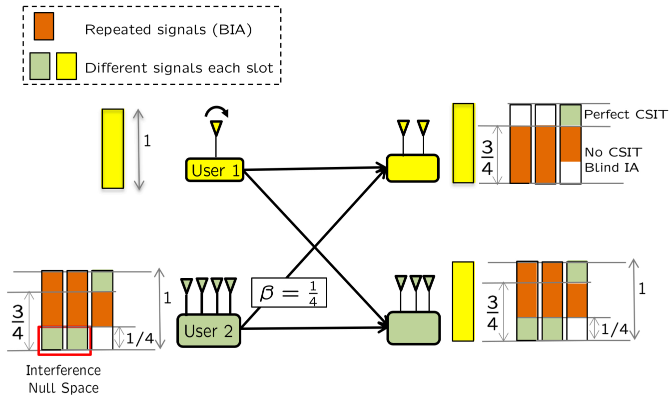

Figure 5.4: Illustration of achievable scheme for Example 2.

We now increase the transmitting antennas at Tx2 by 1 from Example 1. Consider the setting $(M_1, M_2, N_1, N_2) = (1, 4, 2, 3)$ as Figure 5.4, and all other settings remain the same. This is a special case in Theorem 5.3. Now if we use the same scheme as Example 1, Rx2 will not have enough antennas to separate the all signals if Tx2 fully uses its two-dimension interference null space. Thus Tx2 can still only send 1 stream through this null space and achieve $1.5 + \frac{\beta}{2}$ DoF for User 2. Although there is still a synergistic DoF gain for such a scheme as shown in Figure 5.5, it doesn't use Tx2's null space efficiently. Next we introduce a new scheme for this example.

In this case, Tx2 needs to carefully design its transmitted signal, so that it arrives at Rx1 with a layered structure from the power level prospective. Then, Rx1 can treat its top-level received signal as with perfect CSIT and its bottom-level received signal as with only reconfigurable antennas (as shown in Figure 5.4). And thus Rx2 can decode its desired signal in a successive way and achieve 1.75 DoF without hurting User 1. The proposed scheme is as follows.

Designing transmitted signals:

The scheme operates over two channel uses. Mathematically, the transmitted signals at first channel use are,

$$\boldsymbol{x}_1 = \bar{P}X_{11} + \bar{P}^{\frac{3}{4}}X_{12} + \bar{P}^{\frac{1}{4}}X_{13} \tag{5.8}$$

$$\boldsymbol{x}_2 = \mathbf{v}_{21}X_{21} + \mathbf{v}_{22}X_{22} + \mathbf{v}_{23}X_{23} \tag{5.9}$$

where X_{1i} are independent Gaussian codewords from unit power codebooks. X_{11} and X_{13} each carries $\frac{1}{4}$ DoF, X_{12} carries $\frac{1}{2}$ DoF. The precoding vector $\mathbf{v}_{2j}$ at Tx2 each has size 4×1. $\mathbf{v}_{21}$ and $\mathbf{v}_{22}$ are chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}[\mathbf{v}_{21}\ \mathbf{v}_{22}] = \mathbf{0}$. $\mathbf{v}_{23}$ can be chosen as a random generic vector. X_{2i} has a layered structure as follows,

$$X_{21} = \bar{P}X_{24}^{b} + \bar{P}^{\frac{3}{4}}X_{21}^{b} + \bar{P}^{\frac{1}{4}}X_{21}^{p} \tag{5.10}$$

$$X_{22} = \bar{P}X_{25}^{b} + \bar{P}^{\frac{3}{4}}X_{22}^{b} + \bar{P}^{\frac{1}{4}}X_{22}^{p} \tag{5.11}$$

$$X_{23} = \bar{P}X_{23}^{p} + \bar{P}^{\frac{3}{4}}X_{23}^{b} \tag{5.12}$$

here X_{2j}^{b} and X_{2i}^{p} are all independent Gaussian codewords from unit power codebooks, with the difference that X_{2j}^{b} will be repeated by Tx2 at the second channel use. They correspond to the signals decoded at Rx1 by BIA. X_{21}^{b}, X_{22}^{b} and X_{23}^{b} each carries $\frac{1}{2}$ DoF and all other codewords each carries $\frac{1}{4}$ DoF.

Now the received signals are,

$$\boldsymbol{y}_1 = \mathbf{H}_{11}^{1}\boldsymbol{x}_1 + \bar{P}^{-\frac{1}{4}}\tilde{\mathbf{H}}_{12}(\mathbf{v}_{21}X_{21} + \mathbf{v}_{22}X_{22}) + \mathbf{H}_{12}\mathbf{v}_{23}X_{23} + \boldsymbol{z}_1 \tag{5.13}$$

$$= \bar{P}[\mathbf{H}_{11}^{1}X_{11} + \mathbf{H}_{12}\mathbf{v}_{23}X_{23}^{p}] \tag{5.14}$$

$$+ \bar{P}^{\frac{3}{4}}[\mathbf{H}_{11}^{1}X_{12} + \tilde{\mathbf{H}}_{12}(\mathbf{v}_{21}X_{24}^{b} + \mathbf{v}_{22}X_{25}^{b}) + \mathbf{H}_{12}\mathbf{v}_{23}X_{23}^{b}] \tag{5.15}$$

$$+ \bar{P}^{\frac{1}{2}}[\tilde{\mathbf{H}}_{12}(\mathbf{v}_{21}X_{21}^{b} + \mathbf{v}_{22}X_{22}^{b})] \tag{5.16}$$

$$+ \bar{P}^{\frac{1}{4}}\mathbf{H}_{11}^{1}X_{13} + \tilde{\mathbf{H}}_{12}(\mathbf{v}_{21}X_{21}^{p} + \mathbf{v}_{22}X_{22}^{p}) + \boldsymbol{z}_1 \tag{5.17}$$

$$
\begin{aligned}
\boldsymbol{y}_2 =& \mathbf{H}_{21}^1 \boldsymbol{x}_1 + \mathbf{H}_{22}(\mathbf{v}_{21} X_{21} + \mathbf{v}_{22} X_{22} + \mathbf{v}_{23} X_{23}) + \boldsymbol{z}_2 && (5.18)\\
=& \bar{P}[\mathbf{H}_{21}^1 X_{11} + \mathbf{H}_{22}(\mathbf{v}_{21} X_{24}^b + \mathbf{v}_{22} X_{25}^b + \mathbf{v}_{23} X_{23}^p)] && (5.19)\\
&+ \bar{P}^{\frac{3}{4}}[\mathbf{H}_{21}^1 X_{12} + \mathbf{H}_{22}(\mathbf{v}_{21} X_{21}^b + \mathbf{v}_{22} X_{22}^b + \mathbf{v}_{23} X_{23}^b)] && (5.20)\\
&+ \bar{P}^{\frac{1}{4}}[\mathbf{H}_{21}^1 X_{13} + \mathbf{H}_{22}(\mathbf{v}_{21} X_{21}^p + \mathbf{v}_{22} X_{22}^p)] + \boldsymbol{z}_2 && (5.21)
\end{aligned}
$$

At the second channel use, Tx1 switch its antenna mode. Tx1 and Tx2 send new codewords X'_{1i} and X'^p_{2i} carry the same number of DoF as X_{1i} and X^p_{2i} in the first channel use, in addition, their power lever and precoding vectors are also the same as X_{1i} and X^p_{2i}. Tx2 also send the repeated codewords X^b_{2j} with the same power lever and precoding vectors as the first channel use. Then the received signals are,

$$
\begin{aligned}
\boldsymbol{y}'_1 =& \bar{P}[\mathbf{H}_{11}^2 X'_{11} + \mathbf{H}_{12}\mathbf{v}_{23} X'^p_{23}] && (5.22)\\
&+ \bar{P}^{\frac{3}{4}}[\mathbf{H}_{11}^2 X'_{12} + \tilde{\mathbf{H}}_{12}(\mathbf{v}_{21} X_{24}^b + \mathbf{v}_{22} X_{25}^b) + \mathbf{H}_{12}\mathbf{v}_{23} X_{23}^b] && (5.23)\\
&+ \bar{P}^{\frac{1}{2}}[\tilde{\mathbf{H}}_{12}(\mathbf{v}_{21} X_{21}^b + \mathbf{v}_{22} X_{22}^b)] && (5.24)\\
&+ \bar{P}^{\frac{1}{4}}\mathbf{H}_{11}^2 X'_{13} + \tilde{\mathbf{H}}_{12}(\mathbf{v}_{21} X'^p_{21} + \mathbf{v}_{22} X'^p_{22}) + \boldsymbol{z}_1 && (5.25)\\
\boldsymbol{y}'_2 =& \bar{P}[\mathbf{H}_{21}^2 X'_{11} + \mathbf{H}_{22}(\mathbf{v}_{21} X_{24}^b + \mathbf{v}_{22} X_{25}^b + \mathbf{v}_{23} X'^p_{23})] && (5.26)\\
&+ \bar{P}^{\frac{3}{4}}[\mathbf{H}_{21}^2 X'_{12} + \mathbf{H}_{22}(\mathbf{v}_{21} X_{21}^b + \mathbf{v}_{22} X_{22}^b + \mathbf{v}_{23} X_{23}^b)] && (5.27)\\
&+ \bar{P}^{\frac{1}{4}}[\mathbf{H}_{21}^2 X'_{13} + \mathbf{H}_{22}(\mathbf{v}_{21} X'^p_{21} + \mathbf{v}_{22} X'^p_{22})] + \boldsymbol{z}_2 && (5.28)
\end{aligned}
$$

Decoding at Rx1:

Rx1 can first decode X_{11} and X^p_{23} from $\boldsymbol{y}_1$ by treating all other signals as white noise. This is possible because X_{11} and X^p_{23} each carries $\frac{1}{4}$ DoF. They are received at Rx1 with power level $\sim \bar{P}$ and the equivalent noise floor is $\sim \bar{P}^{\frac{3}{4}}$ (refer to the top $\frac{1}{4}$ level at Rx1 in Figure 5.4). Once decode X_{11} and X^p_{23}, Rx1 can subtract them from $\boldsymbol{y}_1$. Note that Rx1 can also decode X'_{11} and X'^p_{23} in a same manner and subtract them from $\boldsymbol{y}'_1$.

Now the remaining interference at Rx1 are aligned into a two-dimension subspace (refer to the bottom $\frac{3}{4}$ level at Rx1 in Figure 5.4). This is because Rx1 observes the same linear combinations of X_{2j}^b from two channel use, and X_{21}^p, $X_{22}'^p$, X_{21}^p and $X_{22}'^p$ are all received below the noise floor. Therefore, Rx 1 can use the BIA approach to decode X_{1i}, X_{12}, X_{1i}' and X_{12}'. Then Rx1 achieves a total of 2 DoF in two channel uses, and $d_1 = 1$ per channel use is achieved.

Decoding at Rx2:

Rx2 is able to decode all the signal from both transmitter. Specifically, the received signals at Rx2 have a three-layered structure. X_{11}, X_{11}', X_{24}^b, X_{25}^b, X_{23}^p and $X_{23}'^p$ are received at Rx2 with power level $\sim \bar{P}$, and each carry $\frac{1}{4}$ DoF. They are the top layer of $\boldsymbol{y}_1$ and $\boldsymbol{y}_1'$ (refer to the top $\frac{1}{4}$ level at Rx2 in Figure 5.4). Rx2 can decode all the codewords on its first layer by treating all other signals as white noise. This is possible because the equivalent noise floor is $\sim \bar{P}^{\frac{3}{4}}$.

After Rx2 decodes codewords on its first layer, it subtracts them from the received signals and decode all the codewords on the middle layer (refer to the middle $\frac{1}{2}$ level at Rx2 in Figure 5.4) by treating the signals on the bottom layer as white noise. Then by subtracting the codewords decoded on the middle layer, Rx2 can decode the remaining codewords on the bottom layer (refer to the bottom $\frac{1}{4}$ level at Rx2 in Figure 5.4). Thus, Rx2 achieves 3.5 DoF over two channel user, and $d_2 = 1.75$ per channel use can be achieved. Note that through constructing specific channel coefficients for the channel matrices at each channel use, one can easily verify the linear independence of the received vector of each codeword at Rx2 to guarantee they are decodable.

By extending this scheme to the arbitrary partial CSIT level, i.e., $0 \leq \beta \leq 1$, the DoF achieved by User 2 as a function of β is $d_2 = 1.5 + \max(\beta, 0.5)$ (as shown in Figure 5.5). To emphasize the synergistic benefit, the DoF achieved with only partial CSIT and with only reconfigurable antennas are also shown in Figure 5.5.

One may notice that this signal design scheme is limited by the dimension of the interference

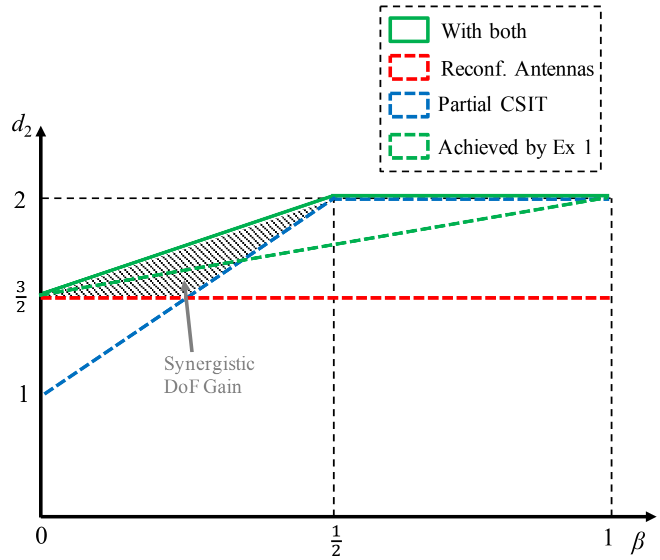

Figure 5.5: Synergistic DoF gain for Example 2.

null space at Tx2. So for a general case with $N_2 < M_2$, a synergistic DoF gain like Figure 5.5 can only be obtain when Tx2 have enough dimensions of its interference null space (Condition (5.7) is satisfied).

5.3.3 Discussions

Example 2 shows that when a simple scheme combining BIA and partial zero-forcing like Example 1 cannot utilize the interference null space efficiently, we can further improve the synergistic DoF gain for User 2. The main idea is letting Tx2 to use its interference null space in a novel way, so that, other than only transmitting signals in the null space that will not be heard by Rx1, now Tx2 can also align part of its transmitted signals at a lower power level at Rx1 so that they will occupy a smaller subspace. Then Rx1 can use BIA to eliminate interference. While at the same time, this same part of signals will still arrive at Rx2 with its normal higher power level for Rx2 to decode.

Despite the differences between the two examples, the two proposed schemes has one thing in common, i.e., the transmitted signals at Tx2 are stilted into two parts either in signal space or in power level. We now provide some intuitions behind our schemes and the existence condition for synergistic DoF gain stated in Theorem 5.1.

Theorem 5.1 claims a sufficient condition for a synergistic DoF gain to exist for a 2-user MIMO interference channel, i.e., if both reconfigurable antennas and partial CSIT are useful. This condition can be understand as the existence of two separate signal subspace at Tx2, i.e., one subspace for BIA and the other for partial zero-forcing. Specifically, for the cases where reconfigurable antennas is useful, $N_1 > M_1$ is always satisfied, i.e., Rx1 can tolerate some interference from Tx2. Then with reconfigurable antennas at Tx1, Tx2 can achieve more DoF compare to with no CSIT, by using BIA approach to reduce its interference dimension causes at Rx1. On the other hand, when partial CSIT at Tx2 is useful, there is always a interference null space from Tx2 to Rx1. Then Tx2 can achieve more DoF compare to with no CSIT by sending part of its signals though the null space so that they will not be heard by Rx1.

Now, we can see that the nature of BIA and partial zero-forcing approach does not conflict in 2-user MIMO interference channel, i.e., BIA utilize the signal subspace will be seen by Rx1 and partial zero-forcing utilize the signal subspace will not be seen by Rx1. If any one of the two signal subspaces does not exist, these two approaches cannot be combined intuitively. Therefore, we conjecture that the condition in Theorem 5.1 is no only sufficient but also necessary for synergistic DoF gain to exist in 2-user MIMO interference channel.

5.4 Proof for Theorem 5.1

5.4.1 Part I

To prove Theorem 5.1, the first step is to prove that $M_1 < N_1 < \min(M_2, N_2)$ is the only case in 2-user interference channel that both reconfigurable antennas and partial CSIT are useful. We will first identify the cases where partial CSIT is useful. Then it is sufficient to shown whether reconfigurable antennas is useful among these cases.

First, for the 2-user interference channel with partial CSIT at Tx2 and without reconfigurable antennas, one can classify such a channel into two cases: 1) $N_1 < N_2$ and 2) $N_1 \geq N_2$.

Case 1 can be further divided into two sub cases: a) $M_2 \leq N_1 < N_2$, b) $N_1 < \min(M_2, N_2)$. Any CSIT is not useful for sub case 1.a, since the DoF region with perfect CSIT is the same with no CSIT. For sub case 1.b, partial CSIT at Tx2 is useful [44, 20].

Case 2 can be further divided into three sub cases: a) $M_1 \leq N_2 \leq N_1$, b) $N_2 < M_1$ and $N_2 \leq \min(M_2, N_1)$, c) $N_2 \leq N_1$ and $M_2 < N_2 < M_1$. Any CSIT is not useful for sub case 2.a, since the DoF region with perfect CSIT is the same with no CSIT. For sub case 2.b, partial CSIT at Tx2 is not useful, since the bound for no CSIT in [21] still holds here as long as Tx1 has no CSIT. For sub case 2.c, partial CSIT at Tx2 is also not useful, since any 2-user channel with partial CSIT only at Tx2 and $M_2 < \min(N_1, N_2)$ is equivalent to a channel without any CSIT.

Now we know that the only sub case for partial CSIT at Tx2 to be useful is sub case 1.b. Then for sub case 1.b, if there is no CSIT everywhere, Tx1 has reconfigurable antennas, reconfigurable antennas is useful only when $M_1 \geq N_1$[27].

Therefore, both reconfigurable antennas and partial CSIT are useful only when $M_1 < N_1 < \min(M_2, N_2)$.

5.4.2 Part II

We now prove the achievable DoF in (5.4). Consider N_1 time slots. During these N_1 time slots, Tx1 switches its reconfigurable antenna mode each time to go through all N_1 modes. Tx2 send messages through the same channel during N_1 time slots. It is sufficient to show that $(N_1 M_1, (N_1 - M_1)\min(M_2, N_2) + \beta M_1(\min(M_2, N_2) - N_1))$ DoF can be achieved in N_1 time slots.

The key idea is that, in addition to the original scheme with only reconfigurable antennas, partial CSIT allows Tx2 to transmit additional messages without hurting the original scheme.

Mathematically, the overall transmitted signals in N_1 time slots are

$$\mathbf{X}_1 = \bar{P}\widehat{\boldsymbol{x}}_1 \tag{5.29}$$

$$\mathbf{X}_2 = \bar{P}\widehat{\mathbf{V}}^1\widehat{\boldsymbol{x}}_{21} + \bar{P}^{\beta}\widehat{\mathbf{V}}^2\widehat{\boldsymbol{x}}_{22} \tag{5.30}$$

Here, $\widehat{\boldsymbol{x}}_1 = [x_1^{[1]}, x_1^{[2]}, ..., x_1^{[N_1 M_1]}]^T$, $\widehat{\boldsymbol{x}}_2 = [\widehat{\boldsymbol{x}}_{21}^T, \widehat{\boldsymbol{x}}_{22}^T]^T$, $\widehat{\boldsymbol{x}}_{21} = [x_{21}^{[1]}, x_{21}^{[2]}, ..., x_{21}^{[(N_1-M_1)\min(M_2,N_2)]}]^T$, $\widehat{\boldsymbol{x}}_{21} = [x_{22}^{[1]}, x_{22}^{[2]}, ..., x_{22}^{M_1(M_2-N_1)}]^T$. $x_1^{[i]}$ and $x_{2k}^{[j]}$ are independent Gaussian codewords from unit power codebooks. Each $x_1^{[i]}$ and $x_{21}^{[j]}$ carries 1 DoF, each $x_{22}^{[j]}$ carries β DoF.

The achievable scheme is a combination of the scheme in [27] (Section IV.C) and partial zero-forcing precoding. $\widehat{\boldsymbol{x}}_1$ and $\widehat{\boldsymbol{x}}_{21}$ are transmitted in the same manner as [27] according to the reconfigurable antenna switching pattern, while $\widehat{\boldsymbol{x}}_{22}$ is transmitted through the null space of the estimated channel matrix. Since the major part of this scheme is the same with [27], we only highlight the differences here.

$\widehat{\mathbf{V}}^2$ is the partial zero-forcing matrix for $\widehat{\boldsymbol{x}}_{22}$ at Tx2 with following structures

$$\widehat{\mathbf{V}}^2_{M_2 N_1 \times M_1(\min(M_2,N_2)-N_1)} = \begin{bmatrix} \mathbf{I}_{M_1} \otimes \mathbf{V}_{M_2\times(\min(M_2,N_2)-N_1)} \\ \mathbf{0}_{M_2(N_1-M_1)\times M_1(\min(M_2,N_2)-N_1)} \end{bmatrix} \tag{5.31}$$

where $\mathbf{V}_{M_2\times(\min(M_2,N_2)-N_1)}$ is chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}\mathbf{V} = \mathbf{0}$. $\widehat{\mathbf{V}}^1$ is the same with [27].

Note that $\widehat{\boldsymbol{x}}_{22}$ is transmitted with power level $\sim \bar{P}^{\beta}$, it will be received at Rx1 below the noise floor. Thus Rx1 is equivalent to a receiver in [27] with only reconfigurable antennas at transmitter. Therefore Rx1 is able to decode $\widehat{\boldsymbol{x}}_1$ and $\widehat{\boldsymbol{x}}_{21}$.

Now we consider Rx2, we first show that it can decode $\widehat{\boldsymbol{x}}_1$ and $\widehat{\boldsymbol{x}}_{21}$. Then by subtracting the decoded message, Rx2 has enough antennas to decode $\widehat{\boldsymbol{x}}_{22}$.

To decode $\widehat{\boldsymbol{x}}_1$ and $\widehat{\boldsymbol{x}}_{21}$, Rx2 first zero-forcing $\widehat{\boldsymbol{x}}_{22}$ each time slot through a receiving zero-forcing matrix $\mathbf{R}_{(\min(M_2,N_2)-N_1)\times N_2}$. $\mathbf{R}$ is chosen so that $\mathbf{R}\mathbf{H}_{22}\mathbf{V} = \mathbf{0}$. Now Rx 2 is equivalent to a receiver in [27] with $(N_2-M_2)^{+}+N_1$ antennas. Since this equivalent receiver still has more antennas than Rx1 and their channel are statistically equivalent. If Rx1 can decode $\widehat{\boldsymbol{x}}_1$ and $\widehat{\boldsymbol{x}}_{21}$, then Rx2, having at least as many antennas, must also be able to decode the same messages. This concludes the proof.

5.5 Proof for Theorem 5.2

In order to prove the tightness of Theorem 5.2 for the case $N_1 < M_2 \leq N_2$, it is sufficient to prove that all the corner points in the DoF region can be achieved. When $M_1 < N_1 < M_2 \leq N_2$, the corner point $(d_1, d_2) = (M_1, M_2 - \frac{M_1 M_2}{N_1} + \beta \frac{M_1(M_2 - N_1)}{N_1})$ can be achieved by Theorem 5.1. All the other corner points in can be achieved with only partial CSIT as [44, 20].

We now prove the outer bound. A key step for the information-theoretic DoF outer bound proof is to first perform a change of basis operation, corresponding to invertible linear transformations at Tx22.

5.5.1 Change of basis

By the invertible transformations, we aim to obtain an equivalent channel at time slot t where the input at Tx2 are partitioned as $\boldsymbol{x}_2(t) = [\boldsymbol{x}_{21}^T(t), \boldsymbol{x}_{22}^T(t)]^T$. Let $|\boldsymbol{x}_{2i}|$ indicate the size/number of antennas in $\boldsymbol{x}_{2i}(t)$. We want $|\boldsymbol{x}_{21}| = M_2 - N_1$ and $|\boldsymbol{x}_{22}| = N_1$. This implies that $\boldsymbol{x}_{21}(t) = [x_{21}(t), x_{22}(t), ..., x_{2(M_2-N_1)}(t)]^T \in \mathbb{R}^{(M_2-N_1)\times 1}$ and $\boldsymbol{x}_{22}(t) = [x_{2(M_2-N_1+1)}(t), ..., x_{2M_2}(t)]^T \in \mathbb{R}^{N_1 \times 1}$.

Under the generic channel coefficient assumption, the matrices $\hat{\mathbf{H}}_{12}(t)$ have a $M_2 - N_1$ dimension right null space almost surely. Thus Tx2 can zero-force $\boldsymbol{x}_{21}$ into this null space. Due to the estimation error $\tilde{\mathbf{H}}_{12}(t)$, the residual interference caused by $\boldsymbol{x}_{21}$ at Rx1 has power level $\bar{P}^{-\beta}$.

Now the original channel can be transformed into a channel without any CSIT and having the following input-output equations

$$\boldsymbol{y}_1'(t) = \mathbf{H}_{11}^m(t)\boldsymbol{x}_1(t) + \mathbf{H}_{121}'(t)\boldsymbol{x}_{22}(t) + \bar{P}^{-\beta}\mathbf{H}_{122}'(t)\boldsymbol{x}_{21}(t) + \boldsymbol{z}_1(t), \tag{5.32}$$

$$\boldsymbol{y}_2'(t) = \mathbf{H}_{21}^m(t)\boldsymbol{x}_1(t) + \mathbf{H}_{22}'(t)\boldsymbol{x}_2(t) + \boldsymbol{z}_2(t), \tag{5.33}$$

where $\mathbf{H}_{122}'(t)$ has size $N_1 \times (M_2 - N_1)$. $\mathbf{H}_{121}'(t)$ has size $N_1 \times N_1$. $\mathbf{H}_{22}'(t)$ has the same size with $\mathbf{H}_{22}(t)$.

5.5.2 Deterministic Model and Compound Setting

After the change of basis operation, we discretize the channel (5.32) and (5.33) to a deterministic channel model.

$$\bar{\boldsymbol{y}}_1(t) = \lfloor \mathbf{H}_{11}^m(t)\bar{\boldsymbol{x}}_1(t) \rfloor + \lfloor \mathbf{H}_{121}'(t)\bar{\boldsymbol{x}}_{22}(t) \rfloor + \left\lfloor \mathbf{H}_{122}'(t)(\bar{\boldsymbol{x}}_{21}(t))^{1-\beta} \right\rfloor, \tag{5.34}$$

$$\bar{\boldsymbol{y}}_2(t) = \lfloor \mathbf{H}_{21}^m(t)\bar{\boldsymbol{x}}_1(t) \rfloor + \lfloor \mathbf{H}_{22}'(t)\bar{\boldsymbol{x}}_2(t) \rfloor . \tag{5.35}$$

where the input $\bar{\boldsymbol{x}}_k(t) = [\bar{x}_{k1}(t), \bar{x}_{k2}(t), ..., \bar{x}_{kM_k}(t)]^T$ and $\bar{x}_{ki}(t) \in \left\{0, 1, ..., \lceil \bar{P} \rceil\right\}$, $\forall i \in \mathcal{I}_{M_k}$. The size of $\bar{\boldsymbol{x}}_{21}(t)$ and $\bar{\boldsymbol{x}}_{22}(t)$ is the same with $\boldsymbol{x}_{21}(t)$ and $\boldsymbol{x}_{22}(t)$.

Now, let us impose a compound setting on $\mathbf{H}_{121}'(t)$ and $\mathbf{H}_{122}'(t)$, which is consistent with the outer bound argument. To see this, suppose we first introduce another $M_2 - 1$ auxiliary receivers, that are statistically equivalent to the original Rx1 and require the same message W_1. Since the additional receivers have the same decoding capabilities as the original receivers, the capacity region is not decreased. If we denote $\bar{\boldsymbol{y}}_1(t)$, $\mathbf{H}_{121}'(t)$ and $\mathbf{H}_{122}'(t)$ in original Rx1 as $\bar{\boldsymbol{y}}_1^{[1]}(t)$, $\mathbf{H}_{121}^{[1]}(t)$ and $\mathbf{H}_{122}^{[1]}(t)$, then the total M_2 outputs require W_1 are as follows.

$$\bar{\boldsymbol{y}}_1^{[1]}(t) = \lfloor \mathbf{H}_{11}^m(t)\bar{\boldsymbol{x}}_1(t) \rfloor + \left\lfloor \mathbf{H}_{121}^{[1]}(t)\bar{\boldsymbol{x}}_{22}(t) \right\rfloor + \left\lfloor \mathbf{H}_{122}^{[1]}(t)(\bar{\boldsymbol{x}}_{21}(t))^{1-\beta} \right\rfloor, \tag{5.36}$$

$$\bar{\boldsymbol{y}}_1^{[2]}(t) = \lfloor \mathbf{H}_{11}^m(t)\bar{\boldsymbol{x}}_1(t) \rfloor + \left\lfloor \mathbf{H}_{121}^{[2]}(t)\bar{\boldsymbol{x}}_{22}(t) \right\rfloor + \left\lfloor \mathbf{H}_{122}^{[2]}(t)(\bar{\boldsymbol{x}}_{21}(t))^{1-\beta} \right\rfloor, \tag{5.37}$$

$$\vdots$$

$$\bar{\boldsymbol{y}}_1^{[M_2]}(t) = \lfloor \mathbf{H}_{11}^m(t)\bar{\boldsymbol{x}}_1(t) \rfloor + \left\lfloor \mathbf{H}_{121}^{[M_2]}(t)\bar{\boldsymbol{x}}_{22}(t) \right\rfloor + \left\lfloor \mathbf{H}_{122}^{[M_2]}(t)(\bar{\boldsymbol{x}}_{21}(t))^{1-\beta} \right\rfloor, \tag{5.38}$$

Note that in the compound setting the transmitter knows that the cross-channel matrix can take any one of values $\left\{\mathbf{H}_{12i}^{[1]}(t), ..., \mathbf{H}_{12i}^{[M_2]}(t)\right\}$. Essentially we have M_2 Rx1s controlled by the same inputs.

5.5.3 Outer Bound

We now prove the bound $\frac{d_1}{N_1} + \frac{d_2}{M_2} \leq 1 + \frac{\beta \min(M_1, N_1)(M_2 - N_1)}{N_1 M_2}$.

We start with the first N_1 receivers from the compound setting that requires W_1. For $i \in \mathcal{I}_{N_1}$, from Fano's inequality, we have

$$\begin{aligned}
& nR_1^{[i]} \\
\leq & I(W_1; \bar{\boldsymbol{y}}_1^{[i]n}) + o(n) && (5.39)\\
= & H(\bar{\boldsymbol{y}}_1^{[i]n}) - H(\bar{\boldsymbol{y}}_1^{[i]n} \mid W_1) + o(n) && (5.40)\\
\leq & H(\bar{\boldsymbol{y}}_1^{[i]n}) - H(\bar{\boldsymbol{y}}_1^{[i]n}, (\bar{\boldsymbol{x}}_{22}^n)^\beta \mid W_1) + no(\log(\bar{P})) + o(n) && (5.41)\\
= & H(\bar{\boldsymbol{y}}_1^{[i]n}) - H((\bar{\boldsymbol{x}}_{22}^n)^\beta \mid W_1) - H(\bar{\boldsymbol{y}}_1^{[i]n} \mid W_1, (\bar{\boldsymbol{x}}_{22}^n)^\beta) \\
& + no(\log(\bar{P})) + o(n) && (5.42)\\
\leq & nN_1 \log(\bar{P}) - H((\bar{\boldsymbol{x}}_{22}^n)^\beta) - H(\bar{\boldsymbol{y}}_1^{[i]n} \mid W_1, (\bar{\boldsymbol{x}}_{22}^n)^\beta) \\
& + no(\log(\bar{P})) + o(n) && (5.43)
\end{aligned}$$

where (5.41) follows from the fact that with W_1, Rx1 has enough antennas to decode $(\bar{\boldsymbol{x}}_{22}^n)^\beta$ from the top β power level of $\bar{\boldsymbol{y}}_1^{[i]n}$. (5.42) follows from the chain rule. (5.43) is obtained because Rx1 has only N_1 antennas, W_1 is independent with W_2.

Then for the rest $M_2 - N_1$ receivers that requires W_1, for $j \in \{N_1 + 1, N_1 + 2, ..., M_2\}$, we have

$$
\begin{aligned}
& nR_1^{[j]} \\
\leq & I(W_1; \bar{\boldsymbol{y}}_1^{[j]n}) + o(n) && (5.44) \\
= & I(W_1; (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta}, (\bar{\boldsymbol{y}}_1^{[j]n})_{1-\beta}) + o(n) && (5.45) \\
= & I(W_1; (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta}) + I(W_1; (\bar{\boldsymbol{y}}_1^{[j]n})_{1-\beta} \mid (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta}) + o(n) && (5.46) \\
\leq & n\beta \min(M_1, N_1) \log(\bar{P}) + H((\bar{\boldsymbol{y}}_1^{[j]n})_{1-\beta} \mid (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta}) \\
& - H((\bar{\boldsymbol{y}}_1^{[j]n})_{1-\beta}, (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta} \mid (\bar{\boldsymbol{y}}_1^{[j]n})^{\beta}, W_1) + no(\log(\bar{P})) + o(n) && (5.47) \\
\leq & n\beta \min(M_1, N_1) \log(\bar{P}) + n(1-\beta)N_1 \log(\bar{P}) \\
& - H(\bar{\boldsymbol{y}}_1^{[j]n} \mid (\bar{\boldsymbol{x}}_{22}^{n})^{\beta}, W_1) + no(\log(\bar{P})) + o(n) && (5.48) \\
= & n[N_1 - \beta(N_1 - M_1)^+] \log(\bar{P}) - H(\bar{\boldsymbol{y}}_1^{[j]n} \mid (\bar{\boldsymbol{x}}_{22}^{n})^{\beta}, W_1) \\
& + no(\log(\bar{P})) + o(n) && (5.49)
\end{aligned}
$$

where (5.47) follows from the fact that for a one-to-one $M_1 \times N_1$ MIMO channel, if the noise floor is $\boldsymbol{z} \sim \mathcal{N}(0, P^{1-\beta})$, then the capacity of this channel is $\beta \min(M_1, N_1) \log(\bar{P}) + o(\log(\bar{P}))$. (5.48) is obtained because Rx1 has only N_1 antennas and $(\bar{\boldsymbol{y}}_1^{[j]n})_{1-\beta}$ is the lower $1-\beta$ power level of $\bar{\boldsymbol{y}}_1^{[j]n}$.

Then from (5.43) and (5.49) , we have

$$
\begin{aligned}
& nM_2R_1 \leq \sum_{i=1}^{N_1} R_1^{[i]} + \sum_{j=N_1+1}^{M_2} R_1^{[j]} \\
\leq & nN_1^2 \log(\bar{P}) + n(M_2 - N_1)[N_1 - \beta(N_1 - M_1)^+] \log(\bar{P}) \\
& - N_1 H((\bar{\boldsymbol{x}}_{22}^{n})^{\beta}) - \sum_{k=1}^{M_2} H(\bar{\boldsymbol{y}}_1^{[k]n} \mid (\bar{\boldsymbol{x}}_{22}^{n})^{\beta}, W_1) \\
& + no(\log(\bar{P})) + o(n) && (5.50) \\
\leq & n[N_1 M_2 - \beta(M_2 - N_1)(N_1 - M_1)^+] \log(\bar{P}) - N_1 H((\bar{\boldsymbol{x}}_{22}^{n})^{\beta}) \\
& - N_1 H((\bar{\boldsymbol{x}}_{21}^{n})^{1-\beta}, \bar{\boldsymbol{x}}_{22}^{n} \mid (\bar{\boldsymbol{x}}_{22}^{n})^{\beta}) + no(\log(\bar{P})) + o(n) && (5.51)
\end{aligned}
$$

$$
\begin{aligned}
=&n[N_1M_2-\beta(M_2-N_1)(N_1-M_1)^+]\log(\bar{P}) - N_1H((\bar{\boldsymbol{x}}_{22}^n)^\beta)\\
&- N_1H((\bar{\boldsymbol{x}}_{21}^n)^{1-\beta}, \bar{\boldsymbol{x}}_{22}^n \mid (\bar{\boldsymbol{x}}_{22}^n)^\beta) - N_1H((\bar{\boldsymbol{x}}_{21}^n)_\beta)\\
&+ N_1H((\bar{\boldsymbol{x}}_{21}^n)_\beta) + no(\log(\bar{P})) + o(n) \qquad (5.52)\\
\le&n[N_1M_2-\beta(M_2-N_1)(N_1-M_1)^+]\log(\bar{P})\\
&- N_1H((\bar{\boldsymbol{x}}_{21}^n)_\beta, (\bar{\boldsymbol{x}}_{21}^n)^{1-\beta}, \bar{\boldsymbol{x}}_{22}^n) + n\beta N_1(M_2-N_1)\log(\bar{P})\\
&+ no(\log(\bar{P})) + o(n) \qquad (5.53)\\
\le&n[N_1M_2+\beta\min(M_1,N_1)(M_2-N_1)]\log(\bar{P}) - nN_1R_2\\
&+ no(\log(\bar{P})) + o(n) \qquad (5.54)
\end{aligned}
$$

(5.53) is because $(\bar{\boldsymbol{x}}_{21}^n)_\beta$ is the lower β power level of $\bar{\boldsymbol{x}}_{21}^n$.

(5.51) is because $\sum_{k=1}^{M_2} H(\bar{\boldsymbol{y}}_1^{[k]n} \mid (\bar{\boldsymbol{x}}_{22}^n)^\beta, W_1) \ge N_1H((\bar{\boldsymbol{x}}_{21}^n)^{1-\beta}, \bar{\boldsymbol{x}}_{22}^n \mid (\bar{\boldsymbol{x}}_{22}^n)^\beta)$. This can be proved by the Corollary 1 in [41] [1] , it implies that when a collection of N_1 generic linear combinations (i.e., $\{\bar{\boldsymbol{y}}_1^{[k]n}|W_1\}$) of the M_2 variables can reconstruct these variables (i.e., $\{(\bar{\boldsymbol{x}}_{21}^n)^{1-\beta}, \bar{\boldsymbol{x}}_{22}^n\}$) N_1 times, then the summation of the entropy must carry at least their proportional share of the total entropy of these M_2 variables.

By arranging terms of (5.54), we have

$$
\begin{aligned}
&nM_2R_1 + nN_1R_2\\
&\le n[N_1M_2+\beta\min(M_1,N_1)(M_2-N_1)]\log(\bar{P}) + no(\log(\bar{P})) + o(n) \qquad (5.55)
\end{aligned}
$$

which implies that the $M_2d_1+N_1d_2 \le N_1M_2+\beta\min(M_1,N_1)(M_2-N_1)$. This concludes the proof.

[1] Corollary 1 in [41] was originally for continues channel model, but it can be easily verified that it also works on discrete channel model.

5.6 Proof for Theorem 5.3

5.6.1 Case 1: $\frac{N_2-N_1}{N_2-M_1} \leq \frac{1}{2}$

For this case, we prove that $d_1 = M_1$, $d_2 = \frac{(N_1-M_1)N_2}{N_1} + \beta\frac{M_1(N_2-M_1)}{N_1}$ can be achieved if $\beta < \frac{N_2-N_1}{N_2-M_1}$. When $\beta \geq \frac{N_2-N_1}{N_2-M_1}$, User 2 achieves the same DoF as with perfect CSIT.

The scheme operates over N_1 channel uses. Similar to the scheme in Example 2, the key step of the proposed scheme is to split the transmitted signals into four layers, i.e., denote as layer 1, 2, 3 and 4 from top to bottom. Mathematically, the transmitted signals at each channel use are,

$$\boldsymbol{x}_1 = \bar{P}\mathbf{X}_1^1 + \bar{P}^{1-\frac{2N_1-N_2-M_1}{N_2-N_1}\beta}\mathbf{X}_1^2 + \bar{P}^{1-\frac{N_1-M_1}{N_2-N_1}\beta}\mathbf{X}_1^3 + \bar{P}^{\beta}\mathbf{X}_1^4 \tag{5.56}$$

$$\boldsymbol{x}_2 = \bar{P}\mathbf{v}_{21}\mathbf{X}_2^{1p} + \bar{P}^{1-\frac{2N_1-N_2-M_1}{N_2-N_1}\beta}\mathbf{v}_{22}\mathbf{X}_2^{2p} + \bar{P}^{1-\frac{N_1-M_1}{N_2-N_1}\beta}\mathbf{X}_2^b + \bar{P}^{\beta}\mathbf{v}_{23}\mathbf{X}_2^{4p} \tag{5.57}$$

where $\mathbf{X}_1^i$ and $\mathbf{X}_2^{ip}$ is the signals on i-th layer at Tx1 and Tx2, respectively.

$\mathbf{X}_1^1 = [X_{11}^1, X_{12}^1, ..., X_{1M_1}^1]^T$, $\mathbf{X}_2^{1p} = [X_{21}^{1p}, X_{22}^{1p}, ..., X_{2(N_2-M_1)}^{1p}]^T$, X_{1j}^1 and X_{2j}^{1p} are independent Gaussian codewords from unit power codebooks, each carries $1-2\beta$ DoF.

$\mathbf{X}_1^2 = [X_{11}^2, X_{12}^2, ..., X_{1M_1}^2]^T$, $\mathbf{X}_2^{2p} = [X_{21}^{2p}, X_{22}^{2p}, ..., X_{2(N_2-N_1)}^{2p}]^T$, $\mathbf{X}_1^4 = [X_{11}^4, X_{12}^4, ..., X_{1M_1}^4]^T$, $\mathbf{X}_2^{4p} = [X_{21}^{4p}, X_{22}^{4p}, ..., X_{2(N_2-M_1)}^{4p}]^T$, X_{1j}^2, X_{1j}^4, X_{2j}^{2p} and X_{2j}^{4p} are independent Gaussian codewords from unit power codebooks, each carries β DoF.

$\mathbf{X}_1^3 = [X_{11}^3, X_{12}^3, ..., X_{1M_1}^3]^T$, X_{1j}^3 are independent Gaussian codewords from unit power codebooks, each carries $1 - \frac{N_2-M_1}{N_2-N_1}\beta$ DoF.

$\mathbf{v}_{21}$ and $\mathbf{v}_{22}$ are precoding matrices with size $M_2 \times (N_2-M_1)$ and $M_2 \times (N_2-N_1)$, respectively, can be chosen as random generic matrices. $\mathbf{v}_{23}$ with size $M_2 \times (N_2 - M_1)$ is chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}\mathbf{v}_{23} = \mathbf{0}$.

$\mathbf{X}_2^b$ is N_2 linear combinations of all the BIA signals sent by Tx2 among N_1 channel uses. It is designed jointly across all the N_1 channel use, let us use $\widehat{\mathbf{X}}_2^b$ to denote the overall BIA signals.

$$\widehat{\mathbf{X}}_2^b = (\mathbf{P}_{N_1\times(N_1-M_1)} \otimes \mathbf{v}_{M_2\times N_2})(\mathbf{X}_2^{3b} + \bar{P}^{\beta} \begin{bmatrix} \mathbf{X}_2^{2b} \\ \mathbf{0}_{(N_2-N_1)(N_1-M_1)\times 1} \end{bmatrix}) \tag{5.58}$$

$\mathbf{X}_2^{3b} = [X_{21}^{3b}, X_{22}^{3b}, ..., X_{2N_2(N_1-M_1)}^{3b}]^T$, $\mathbf{X}_2^{2b} = [X_{21}^{2b}, X_{22}^{2b}, ..., X_{2N_1(N_1-M_1)}^{2b}]^T$, X_{2j}^{2b} and X_{2j}^{3b} are independent Gaussian codewords from unit power codebooks, carry $1 - \frac{N_2-M_1}{N_2-N_1}\beta$ and β DoF, respectively. Note that X_{2j}^{3b} is transmitted at Layer 3 of Tx2 and X_{2j}^{2b} is transmitted at Layer 2 of Tx2. $\mathbf{v}_{M_2\times N_2} = [\mathbf{v}_{M_2\times N_1}^1 \;\; \mathbf{v}_{M_2\times(N_2-N_1)}^2]$ is the precoding matrix for $\mathbf{X}_2^b$ at each channel use. $\mathbf{v}^1$ is chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}\mathbf{v}^1 = \mathbf{0}$. $\mathbf{v}^2$ can be chosen as random generic matrix. $\mathbf{P}_{N_1\times(N_1-M_1)}$ is chosen in a same manner as [27] (Section IV.C) Note that the condition (5.7) in Theorem 5.3 ensures that the dimension of interference null space at Tx2 are large enough for $\mathbf{v}^1$ and $\mathbf{v}_{23}$ to be exist.

Now with the above designed transmitted signals, the received signal at Rx1 can be divided into two part, i.e., top Layer 1, 2 can be treated as with perfect CSIT and bottom Layer 3, 4 can be treated as without CSIT. This is possible because with partial zero-forcing, $\mathbf{X}_2^{4p}$ arrives at RX1 with power strength below the noise floor and $\mathbf{X}_2^b$ arrives at Rx1 below Layer 2. Therefore each channel use, from its top two layers Rx1 can decode $\mathbf{X}_1^1$, $\mathbf{X}_1^2$, $\mathbf{X}_2^{1p}$ and $\mathbf{X}_2^{2p}$ as a MAC channel by treating everything else as white noise and then subtract them out of its received signal. Now Rx1 can eliminate the remaining interference term, i.e., $\mathbf{X}_2^b$, with BIA approach and decode $\mathbf{X}_1^1$ and $\mathbf{X}_1^2$. For Rx2, it has enough antennas to decode all the signals at each layer.

Therefore, $(d_1, d_2) = (N_1M_1, (N_1 - M_1)N_2 + \beta M_1(N_2 - M_1))$ is achieved over N_1 channel uses. This concludes the proof.

5.6.2 Case 2: $\frac{N_2-N_1}{N_2-M_1} > \frac{1}{2}$

For this case, the transmitted signals at Tx2 are split into three layers, i.e., denote as layer 1, 2 and 3 from top to bottom. The scheme still operates over N_1 channel uses. The transmitted signals at Tx1 is omitted here since they alway occupy the M_1 dimension subspace at both receivers and cannot affect the designing of Tx2's signal. Mathematically, the transmitted signals at Tx2 for each channel use are,

$$\boldsymbol{x}_2 = \bar{P}\mathbf{v}_{21}\mathbf{X}_2^{1p} + \bar{P}\mathbf{X}_2^b + \bar{P}^{\beta}\mathbf{v}_{23}\mathbf{X}_2^{3p} \tag{5.59}$$

$\mathbf{X}_2^{1p} = [X_{21}^{1p}, X_{22}^{1p}, ..., X_{2(N_2-N_1)}^{1p}]^T$, X_{2j}^{1p} are independent Gaussian codewords from unit power codebooks, each carries $\frac{N_1-M_1}{N_2-N_1}\beta$ DoF.
$\mathbf{X}_2^{3p} = [X_{21}^{4p}, X_{22}^{4p}, ..., X_{2(N_2-M_1)}^{4p}]^T$, X_{2j}^{3p} are independent Gaussian codewords from unit power codebooks, each carries β DoF.
$\mathbf{v}_{21}$ is the precoding matrix with size $M_2 \times (N_2 - N_1)$ which can be chosen as random generic matrix.
$\mathbf{v}_{23}$ with size $M_2 \times (N_2 - M_1)$ is chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}\mathbf{v}_{23} = \mathbf{0}$.
$\mathbf{X}_2^b$ is N_2 linear combinations of all the BIA signals sent by Tx2 among N_1 channel uses. It is designed jointly across all the N_1 channel use, let us use $\widehat{\mathbf{X}}_2^b$ to denote the overall BIA signals.

$$\widehat{\mathbf{X}}_2^b = (\mathbf{P}_{N_1\times(N_1-M_1)} \otimes \mathbf{v}_{M_2\times N_2})\mathbf{X}_2^{1b} \tag{5.60}$$

$\mathbf{X}_2^{1b} = [X_{21}^{31b}, X_{22}^{1b}, ..., X_{2N_2(N_1-M_1)}^{1b}]^T$, X_{2j}^{2b} is independent Gaussian codewords from unit power codebooks, carry $1-\beta$ DoF. $\mathbf{v}_{M_2\times N_2}$ is the precoding matrix for $\mathbf{X}_2^b$ at each channel use and is chosen from the right null space of $\hat{\mathbf{H}}_{12}$, i.e., $\hat{\mathbf{H}}_{12}\mathbf{v} = \mathbf{0}$. $\mathbf{P}_{N_1\times(N_1-M_1)}$ is chosen in a same manner as [27] (Section IV.C)
Note that the condition (5.7) in Theorem 5.3 ensures that the dimension of interference null space at Tx2 are large enough for $\mathbf{v}$ to be exist.
Now with the above designed transmitted signals, the transmitted signals from Tx2 received at Rx1 can be divided into two part, i.e., top layer $\mathbf{X}_2^{1p}$ can be treated as with perfect CSIT and bottom layer $\mathbf{X}_2^b$ can be treated as without CSIT. This is possible because with partial

zero-forcing, $\mathbf{X}_2^{3p}$ arrives at RX1 with power strength below the noise floor and $\mathbf{X}_2^{b}$ arrives at Rx1 with power level $\bar{P}^{1-\beta}$. Therefore similar to Subection 5.6.1, Rx1 can decode its desired signals and $\mathbf{X}_2^{1p}$ and eliminate the interference term $\mathbf{X}_2^{b}$ with BIA approach. For Rx2, it has enough antenna to decode all the signals at each layer.

Therefore, $(d_1, d_2) = (N_1M_1, (N_1 - M_1)N_2 + \beta M_1(N_2 - M_1))$ is achieved over N_1 channel uses. This concludes the proof.

5.7 Summary

In this chapter, the two-user MIMO interference channel with reconfigurable antennas and partial CSIT was considered. The novel achievability scheme was proposed to jointly exploiting both reconfigurable antennas and partial CSIT. Thus revealed synergistic DoF gains that cannot be seen through the study of each individual element by itself. A sufficient condition for synergistic DoF gains to exist was also presented. Then by introducing a new outer bound, the DoF region was completely characterized for the setting $N_1 < M_2 \leq N_2$. It was also shown that when transmitter has more antennas than its corresponding receiver, the synergistic benefit is limited by the dimension of the interference null space.

Chapter 6

Conclusion

In this dissertation, by exploring the DoF for MIMO interference networks under both perfect channel knowledge and channel uncertainty, we demonstrated the optimal use of multiple antennas from an information theoretic perspective. The contributions are summarized as follows:

- In Chapter 2, we proposed a new outer bound based on the idea of creating a replicated-network, i.e., creating copies (replicas) of certain users and choosing the connectivity of the replicated network in such a way that any achievable scheme in the original network translates into an achievable scheme for the replicated network. Based on this new bound, it was shown that for a K-user MIMO interference channel with arbitrarily rank-deficient cross-channels, where there are M_k antennas at the k^{th} user pair, half-the-cake DoF is optimal if the overall $M_\Sigma \times M_\Sigma$ channel matrix $\bar{\mathbf{H}}$ where all desired channels have been set to zero, has full rank.

- In Chapter 3, we exploreded the new achievability scheme for two-user MIMO interference channel with partial CSIT and arbitrary antenna configuration at each node. The ideas of signal space partitioning and elevated multiplexing, and how they work together were introduced.

- In Chapter 4, we extended the results for two-user MIMO IC with arbitrary antenna configurations and arbitrary partial CSIT levels in Chapter 3 into DoF region. The aligned image set approach cannot obtain the tight outer bound by itself, we utilized the sum-set inequalities in order to obtain the tight outer bound.
- In Chapter 5, the novel achievability scheme was proposed to jointly exploiting both reconfigurable antennas and partial CSIT in the two-user MIMO interference channel. The results demonstrated a synergistic DoF gains that cannot be seen through the study of each individual element by itself. A sufficient condition for synergistic DoF gains to exist was also presented. Then by introducing a new outer bound based on compound argument, the DoF region was completely characterized for the setting $N_1 < M_2 \leq N_2$.

An interesting future direction for the replication based bounds is to expand the class of networks where replication based bounds lead us to tight DoF characterizations and also starting to apply the bounds to multiple unicast settings. The problem of multiple unicast is a celebrated open problem in network coding, in conjunction with linear network coding. Some of the most useful bounds for the capacity of networks (corresponds to DoF in wireless networks) are min-cut bounds in general, and rank-based bounds for linear networks in particular. For instance, the rank of the overall channel between all the transmitters on one side and all the receivers on the other side, is the min-cut bound on the sum-capacity of the network. Similarly, replication based bounds will be quite useful to understand what kinds of bounds are implied by various rank constraints on parts of the network.

Another research direction under channel uncertainty is to characterize the DoF region for K-user MIMO IC. In addition to the ideas such as partial zero-forcing and elevated multiplexing introduced in this dissertation, one can expect that partial interference alignment approach will play an important role when the number of users are more than two. Currently outer bound approach, i.e., AIS and sum-set inequalities, also need to be improved, as change of basis operation cannot always transform a network with partial CSIT into a network with no CSIT if there are more than two users in this network. Further research followed by this direction is to also explore the Broadcast Channels, X channels and cellular networks, expand the scope to Generalize DoF, and incorporate other forms of channel knowledge, such as mixed, delayed and alternating channel knowledge in the context of reconfigurable antennas.

Bibliography

[1] G. Bresler, D. Cartwright, and D. Tse. Feasibility of interference alignment for the MIMO interference channel. *IEEE Transactions on Information Theory*, 60(9):5573–5586, 2014.

[2] V. Cadambe and S. Jafar. Interference Alignment and the Degrees of Freedom of the K user Interference Channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, Aug. 2008.

[3] A. B. Carleial. Outer bounds on the capacity of interference channels (corresp.). *IEEE Transactions on Information Theory*, 29(4):602–606, 1983.

[4] S. H. Chae and S.-Y. Chung. On the degrees of freedom of rank deficient interference channels. In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1367–1371, 2011.

[5] S. H. Chae, C. Suh, and S.-Y. Chung. Degrees of freedom of the rank-deficient interference channel with feedback. *IEEE Transactions on Information Theory*, 61(6):3326–3338, 2015.

[6] A. G. Davoodi and S. A. Jafar. Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT. *IEEE Transactions on Information Theory*, 62(10):5603–5618, 2016.

[7] A. G. Davoodi and S. A. Jafar. Transmitter Cooperation under Finite Precision CSIT:A GDoF Perspective. *IEEE Transactions on Information Theory*, 2016.

[8] A. G. Davoodi and S. A. Jafar. Sum-set inequalities from aligned image sets: Instruments for robust GDoF bounds. *arXiv preprint arXiv:1703.01168*, 2017.

[9] A. G. Davoodi, B. Yuan, and S. A. Jafar. GDoF of the MISO BC: Bridging the gap between finite precision and perfect CSIT. *arXiv preprint arXiv:1602.02203*, 2016.

[10] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory*, 54(12):5534–5562, 2008.

[11] C. Geng, H. Sun, and S. A. Jafar. Multilevel topological interference management. In *IEEE Information Theory Workshop*, pages 1–5, 2013.

[12] A. Ghasemi, A. S. Motahari, and A. K. Khandani. Interference alignment for the K user MIMO interference channel. *arXiv preprint arXiv:0909.4604v2*, 2011.

[13] A. Gholami Davoodi. Aligned Image Sets: Instruments for Robust GDoF Bounds. *PhD Dissertation*, 2018.

[14] O. Gonzalez, C. Beltran, and I. Santamaria. A feasibility test for linear interference alignment in MIMO channels with constant coefficients. *IEEE Transactions on Information Theory*, 60(3):1840–1856, 2014.

[15] T. Gou and S. A. Jafar. Degrees of freedom of the K user $M \times N$ MIMO interference channel. *IEEE Transactions on Information Theory*, 56(12):6040–6057, 2010.

[16] T. Gou and S. A. Jafar. Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT. *IEEE Communication Letters*, 16(7):1084 – 1087, July 2012.

[17] T. Gou, S. A. Jafar, C. Wang, S.-W. Jeon, and S.-Y. Chung. Aligned interference neutralization and the degrees of freedom of the $2 \times 2 \times 2$ interference channel. *IEEE Transactions on Information Theory*, 58(7):4381–4395, 2012.

[18] T. Gou, C. Wang, and S. A. Jafar. Aiming Perfectly in the Dark - Blind Interference Alignment through Staggered Antenna Switching. *IEEE Trans. on Signal Processing*, 59:2734–2744, June 2011.

[19] C. Hao and B. Clerckx. MISO networks with imperfect CSIT: A topological rate-splitting approach. *arXiv:1602.03768*, Feb. 2016.

[20] C. Hao, B. Rassouli, and B. Clerckx. Achievable DoF regions of MIMO networks with imperfect CSIT. *IEEE Trans. Inform. Theory.*, 63(10):6587–6606, Oct 2017.

[21] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath. On Degrees of Freedom Region of MIMO Networks without Channel State Information at Transmitters. *IEEE Transactions on Information Theory*, 58(2):849–857, Feb. 2012.

[22] S. Jafar. Interference Alignment: A New Look at Signal Dimensions in a Communication Network. In *Foundations and Trends in Communication and Information Theory*, pages 1–136, 2011.

[23] S. Jafar and M. Fakhereddin. Degrees of freedom for the MIMO interference channel. *IEEE Transactions on Information Theory*, 53(7):2637–2642, July 2007.

[24] S. A. Jafar. Blind Interference Alignment. *IEEE Journal of Selected Topics in Signal Processing*, 6(3):216–227, June 2012.

[25] S. A. Jafar and M. J. Fakhereddin. Degrees of freedom for the MIMO interference channel. *IEEE Transactions on Information Theory*, 53(7):2637–2642, 2007.

[26] H. Joudeh and B. Clerckx. Sum rate maximization for MU-MISO with partial CSIT using joint multicasting and broadcasting. *IEEE ICC*, 2015.

[27] L. Ke and Z. Wang. Degrees of freedom regions of two-user MIMO Z and full interference channels: The benefit of reconfigurable antennas. *Information Theory, IEEE Transactions on*, 58(6):3766–3779, 2012.

[28] M. Kiamari and A. S. Avestimehr. Are generalized cut-set bounds tight for the deterministic interference channel? *Proceedings of 53rd Allerton Conference (arXiv:1510.03951)*, 2015.

[29] S. R. Krishnamurthy, A. Ramakrishnan, and S. A. Jafar. Degrees of freedom of rank deficient MIMO interference channels. *IEEE Transactions on Information Theory*, 61(1):341–365, 2015.

[30] T. Liu, D. Tuninetti, and S. A. Jafar. The DoF of the asymmetric MIMO interference channel with square direct link channel matrices. *Proceedings of 52nd Allerton Conference (arXiv:1410.5107)*, 2014.

[31] T. Liu and C. Yang. Genie tree and degrees of freedom of the symmetric MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 64(22):5914–5929, 2016.

[32] A. Lozano, R. W. Heath Jr., and J. G. Andrews. Fundamental limits of cooperation. *IEEE Transactions on Information Theory*, 59(9):5213–5226, Sep. 2013.

[33] M. A. Maddah-Ali and D. 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.

[34] H. Maleki, S. A. Jafar, and S. Shamai. Retrospective Interference Alignment over Interference Networks. *IEEE Journal of Selected Topics in Signal Processing*, pages 228–240, June 2012.

[35] B. Nazer, M. Gastpar, S. A. Jafar, and S. Vishwanath. Ergodic interference alignment. *IEEE Transactions on Information Theory*, 58(10):6355–6371, 2012.

[36] L. Ruan, V. K. Lau, and M. Z. Win. The feasibility conditions for interference alignment in MIMO networks. *IEEE Transactions on Signal Processing*, 61(8):2066–2077, 2013.

[37] M. Torrellas, A. Agustin, J. Vidal, and O. Munoz-Medina. The DoF of the 3-user (p,p+1) MIMO interference channel. *IEEE Transactions on Communications*, 62(11):3842–3853, 2014.

[38] C. S. Vaze and M. K. Varanasi. The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT. *IEEE Transactions on Information Theory*, 58(8):5354–5374, 2012.

[39] C. Wang, T. Gou, and S. A. Jafar. Interference alignment through staggered antenna switching for MIMO BC with no CSIT. In *Asilomar Conference on Signals, Systems and Computers*, Nov. 2010.

[40] C. Wang, T. Gou, and S. A. Jafar. Subspace alignment chains and the degrees of freedom of the three-user MIMO interference channel. *IEEE Transactions on Information Theory*, 60(5):2432–2479, 2014.

[41] C. Wang, H. Sun, and S. A. Jafar. Genie chains: Exploring outer bounds on the degrees of freedom of MIMO interference networks. *IEEE Transactions on Information Theory*, 62(10):5573–5602, 2016.

[42] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi. Degrees of freedom of time correlated MISO broadcast channel with delayed csit. *IEEE Transactions on Information Theory*, To appear. 2013. Available online: http://dx.doi.org/10.1109/TIT.2012.2215953.

[43] C. Yetis, T. Gou, S. Jafar, and A. Kayran. On feasibility of interference alignment in MIMO interference networks. *IEEE Transactions on Signal Processing*, 58(9):4771–4782, 2010.

[44] B. Yuan and S. Jafar. Elevated multiplexing and signal space partitioning in the 2 User MIMO IC with partial CSIT. In *IEEE SPAWC*, 2016.

[45] Y. Zeng, X. Xu, Y. Guan, E. Gunawan, and C. Wang. Degrees of freedom of the three-user rank-deficient MIMO interference channel. *IEEE Transactions on Wireless Communications*, 13(8):41794192, 2014.

[46] Y. Zhu and D. Guo. The degrees of freedom of isotropic MIMO interference channels without state information at the transmitters. *IEEE Transactions on Information Theory*, 58(1):341–352, 2012.