Capacity Approximations of MIMO Interference Channels: Beyond Degrees of Freedom

by

Yimin Pang

B.A., Xidian University, 2008

M.S., Zhejiang University, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Electrical, Computer and Energy Engineering

2019

This thesis entitled:
Capacity Approximations of MIMO Interference Channels: Beyond Degrees of Freedom
written by Yimin Pang
has been approved for the Department of Electrical, Computer and Energy Engineering

---

Prof. Mahesh K. Varanasi

Prof. Youjian Liu

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Pang, Yimin (Ph.D., Electrical Engineering)

Capacity Approximations of MIMO Interference Channels: Beyond Degrees of Freedom

Thesis directed by Prof. Mahesh K. Varanasi

Spectrum sharing allows the coexistence of heterogeneous wireless networks on the same frequency band. Managing the interference between such networks is critically important to ensure high spectrum efficiency, thus motivating the study of multiple-input-multiple-output (MIMO) interference channels (IC) in information theory. This dissertation studies three classes of such interference channels, namely, the MIMO one-to-three IC, the MIMO IC-ZIC, and the MIMO MAC-IC-MAC.

The MIMO one-to-three IC is a partially connected three-user IC with multiple antenna terminals, where one transmitter that causes interference is heard at all three receivers, whereas the other two transmitters are heard only by their intended receivers. We present inner and outer bounds on the capacity region of the MIMO one-to-three IC, quantify the gap between the two bounds, and show that the gap is independent of the channel signal-to-noise ratios (SNRs) and interference-to-noise ratios (INRs). In particular, the achievable scheme at the interfering transmitter involves three-level superposition coding with linear precoding based on the generalized singular value decomposition (GSVD) whereas the non-interfering transmitters perform single-user coding with Gaussian codebooks and scaled identity covariances. The outer bound is obtained using genie-aided arguments with various combinations of genie information provided to the receivers. The generalized degrees of freedom (GDoF) region, which can be seen as a high SNR approximation of the capacity region, of the MIMO one-to-three IC is then fully characterized. We study the achievability of the GDoF region and the sum GDoF curve using an analysis tool developed in this dissertation, which we refer to as multidimensional signal-level partitioning. This tool is tailored for demonstrating the achievability of GDoF-tuples of a MIMO network that can be achieved via multi-level superposition coding.

The MIMO IC-ZIC is also a partially connected three-user IC consisting of three transmitter-receiver pairs. In the IC-ZIC, the first and second pairs form a two-user IC, the first and third pairs form a one-sided or Z interference channel (ZIC) and the second and third transmitter-receiver pairs taken by themselves are two non-interfering point-to-point links. In this thesis, an explicit inner bound is obtained via a coding scheme

is proposed in which the first transmitter employs three-level superposition coding (as in the MIMO one-to-three IC), the second one employs the previously proposed and well-known Karmakar-Varanasi coding scheme (which achieves a constant-gap-to-capacity region of the two-user MIMO IC), and the third transmitter employs single-user coding with a Gaussian codebook (with scaled identity covariance). An explicit single region outer bound based on genie-aided arguments is then obtained. The gap between the inner and outer bounds is then shown to be within a quantifiable gap to the capacity region and the gap is independent of channel SNRs and INRs. The GDoF region is then characterized and analyzed in a variety of channel settings. The difficulty in this part of the research lies in the quantification of the gap between the 28-inequality inner bound and the 33-inequality outer bound, which is characterized via a series of supporting lemmas that reveal the relationship between the entropy terms in the inner and outer bounds.

The MIMO MAC-IC-MAC consists of two interfering MACs in which there is interference only from one transmitter of each MAC to the receiver of the other MAC. Two achievable rate regions that are within a quantifiable gap of the capacity region for the discrete-memoryless semi-deterministic MAC-IC-MAC were obtained in a previous published work by Pang and Varanasi using inner and outer bounds that are unions of polytopes. In the dissertation, we obtain single region inner and outer bounds that characterize a constant-gap-to-capacity region of the MIMO MAC-IC-MAC. The inner bound is obtained by employing the Karmakar-Varanasi coding scheme at the interfering transmitters and single-user coding with Gaussian codebooks and scaled identity covariances at the non-interfering transmitters. Our work therefore unifies and generalizes the constant-gap-to-capacity regions of the MIMO MAC and the two-user MIMO IC. The GDoF region of the MIMO MAC-IC-MAC is also obtained and analyzed. The GDoF analysis reveals that, at high SNR, when the ratio of the INR to the SNR, both taken in dB, is within a certain range in any cell, non-interfering transmitters in that cell can fully occupy the receivers' signal partitions in one or more dimensions that cannot otherwise be utilized by the interfering transmitter alone. This phenomenon is a generalization of the one discovered by Pang and Varanasi in the previous published work on the scalar Gaussian MAC-IC-MAC.

Dedication

To my parents and friends.

Acknowledgements

I am deeply indebted to my dissertation advisor, Prof. Mahesh K. Varanasi, for innumerable discussions, insightful comments, his drive for precision and rigor, tireless efforts on writing papers, and continuous financial support throughout my research work, without any of which this dissertation would not have been possible. Having him as my advisor and mentor has been an privilege. I am also grateful to Profs. Youjian Liu, Fabio Somenzi, Sriram Sankaranarayanan and Lijun Chen for serving on the dissertation defense committee and their feedback. I appreciate my colleagues Mohamed Abdelaziz Salman and Henry Romero for many useful academic discussions. I also got invaluable writing advice from Elizabeth Cook and Alana Chen. Besides, I want to thank my friends Tim Groth, Chris Marriott, Kaniska Mohanty and Angelica Powell for their moral support. I am very much indebted to my parents.

Contents

Figures

Figure

Chapter 1

Overview

Due to the rapid increase of data demands in recent years, wireless co-band communication has drawn significant interest in both theory and practice. Some applications include the Bluetooth and Wi-Fi coexistence on 2.4 GHz band and the Wi-Fi and LTE-LAA coexistence on 5GHz band. Most recently, 3GPP agreed to start a work item which will define 5G New Radio on unlicensed spectrum. Such emerging technologies motivate the study of channel capacity and capacity achieving coding scheme for interference channels (ICs) in information theory. Besides, since the multiple-input-multiple-output (MIMO) technology has been widely adopted in modern wireless communication systems, it is practically important to study networks with vector input and output in such research.

A single cellular network usually consists of three basic component channels: the MIMO point-to-point (P2P) channel (direct link between a communication pair), the MIMO multiple access channel (MAC, uplink channel from multiple users to the base station) and the MIMO broadcast channel (BC, downlink channel from the base station to multiple users). The capacity regions of these three channels with constant (time/frequency invariant) channel realizations and full channel side information (CSI) known to all the transmitters and receivers have been fully determined. In particular, the capacity region of the MIMO P2P channel was reported by Telatar in [42]. The capacity region of MIMO MAC can be found in the works [9,48]. The MIMO BC capacity region was characterized by Weingarten, Steinberg and Shamai [46]. For results on capacity regions of fading MIMO P2P, MIMO MAC and MIMO BC, please refer to [33, 42].

When multiple heterogeneous networks coexist on the same frequency band, interference may occur between these networks. The three basic component channels introduced above are no longer sufficient to

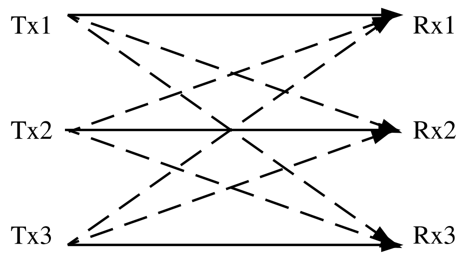

Figure 1.1: A three user IC

describe the character of the coexisted networks. Four additional component channels are needed: the K-user MIMO interference channel (see Fig. 1.1 for a three-user IC example, dashed lines represent interference links and solid lines direct links) where K P2P channels interfering with each other, the L-cell MIMO interfering multiple access channel (IMAC, see Fig. 1.2 for a two-cell IMAC example) where L MACs mutually interfering with each other, the L-cell MIMO interfering broadcast channel (IBC, see Fig. 1.3 for a two-cell IBC example) where L BCs interfering with each other, and the MIMO interfering multiple access broadcast channel (IMABC, see Fig. 1.4 for an IMABC consists of one MAC and one BC) where each MAC or BC interferes with other MACs and BCs. Note the term "interference channel" alone, as being used in the title and throughout the rest of this chapter, means a network contains interference links in general, whereas the term K-user interference channel or K-user IC refers to the network which contains K P2P channels interfering with each other.

The Shannon capacities of those four component channels are all open. Characterizing these capacities has not turned out to be easy. Information theorists have then been seeking for near optimal coding schemes which achieve approximate capacity regions of these channels. The research behind this dissertation is part of such efforts.

In the rest of this chapter, we first introduce three capacity approximations in Section 1.1, and then we summarize related previous milestone works at a high level in Section 1.2 (in Chapters 3-5, detailed related references will be reviewed with respect to the channel models to be studied therein). Lastly, we provide an overview of the research results in this dissertation in Section 1.3.

Throughout the dissertation, we always assume all transmitters and receivers know full global CSI in

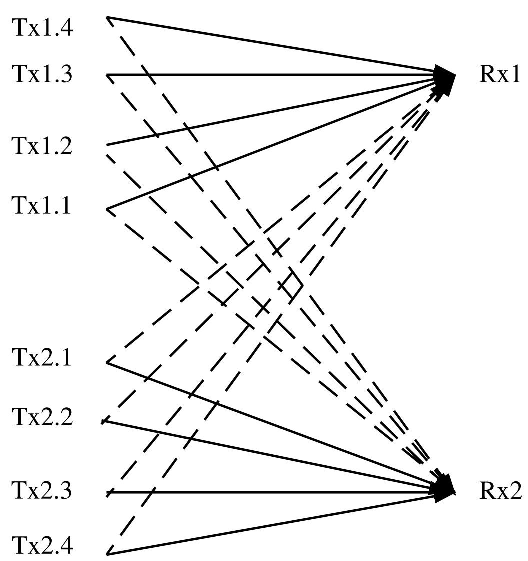

Figure 1.2: An IMAC consists of two uplink cells, each cell has four users.

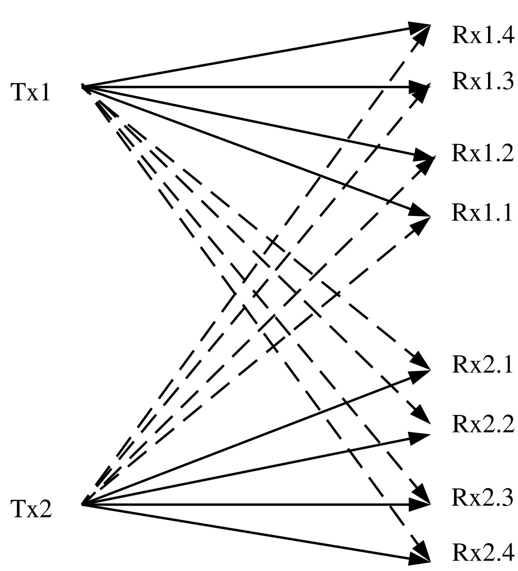

Figure 1.3: An IBC consists of two downlink cells where each cell has four users.

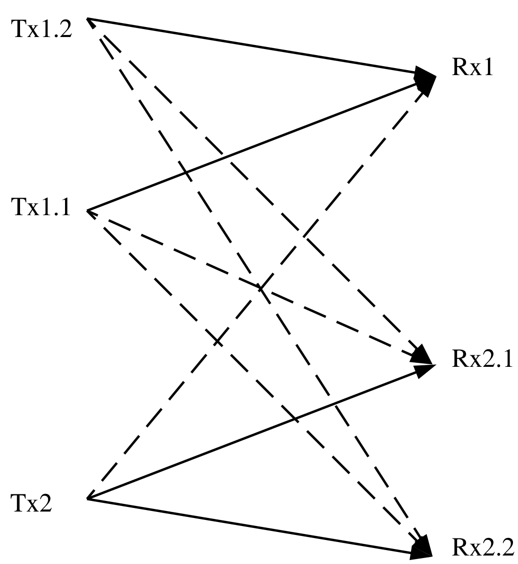

Figure 1.4: An IMABC consists of one downlink cell and one uplink cell, each cell. has two users

any network (also usually referred to as channel). Unless specified as fading or time/frequency varying, all channels have constant channel realizations.

1.1 Capacity Approximations

With decreasing accuracy levels, the following three different capacity approximations have been commonly used. They are constant-gap-to-capacity region, generalized degrees of freedom (GDoF) region and degrees of freedom region (DoF). We introduce these three approximate capacity regions in sequence in this section.

1.1.1 The Constant-gap-to-capacity Region

For a given channel model, a constant-gap-to-capacity region is an achievable region that lies within only constant gap to the Shannon capacity region regardless of the channel parameters, i.e., link signal-to-noise-ratios (SNRs), interference-to-noise-ratios (INRs) and transfer matrices. To obtain a constant-gap-to-capacity region for a channel with K involved rate tuples, we usually find a pair of inner and outer bounds $\mathcal{R}_{\text{in}}$ (which is an achievable region) and $\mathcal{R}_{\text{o}}$ such that for any rate tuple $(R_1, \cdots, R_K) \in \mathcal{R}_{\text{o}}$, we always have $(R_1 - n_1, \cdots, R_K - n_K) \in \mathcal{R}_{\text{in}}$ regardless of channel parameters. Since the capacity lies between $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$, we infer that the inner bound $\mathcal{R}_{\text{in}}$ lies within constant gap $(n_1, \cdots, n_K)$ to the capacity, and hence $\mathcal{R}_{\text{in}}$ is a constant-gap-to-capacity region of this channel. Let us take the constant-gap-to-capacity region of the two-user scalar Gaussian IC given in [15] as an example. Consider the two-user IC as shown in Fig. 1.5 whose parameters are given in the caption. We plot the inner and outer bounds ($\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$) provided by [15] in Fig. 1.6. According to [15], for any $(R_1, R_2) \in \mathcal{R}_{\text{o}}$, we always have $(R_1 - 1, R_2 - 1) \in \mathcal{R}_{\text{in}}$. Hence, the inner bound $\mathcal{R}_{\text{in}}$ is within $(n_1, n_2) = (1, 1)$ bit gap to the capacity region of this IC.

1.1.2 The GDoF Region

The next level of capacity approximation is the GDoF region. To explain the idea of the GDoF region, we start from the constant-gap-to-capacity region [15] of the two-user scalar Gaussian IC given in Fig. 1.5, but let us change the channel parameters a little. Let $\mathsf{SNR}_{11} = \rho^{\alpha_{11}}$, $\mathsf{SNR}_{22} = \rho^{\alpha_{22}}$, $\mathsf{INR}_{12} = \rho^{\alpha_{12}}$

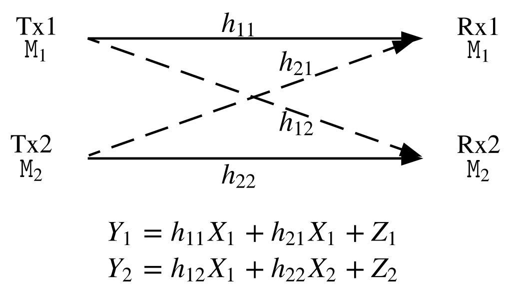

Figure 1.5: A two-user scalar Gaussian IC with symmetric channel parameters: $\mathsf{SNR}_{11} = P_1|h_{11}|^2/N_1 = 100$, $\mathsf{SNR}_{22} = P_2|h_{22}|^2/N_2 = 100$, $\mathsf{INR}_{12} = P_1|h_{12}|^2/N_2 = 50$ and $\mathsf{INR}_{21} = P_2|h_{21}|^2/N_1 = 50$, where P_1 and P_2 are the transmit power at Tx1 and Tx2, N_1 and N_2 are the power of the Gaussian noise Z_1 and Z_2 respectively. Tx1 (and Tx2 respectively) transmits message $\mathtt{M}_1$ ($\mathtt{M}_2$) to Rx1 (Rx2). Random variables X_1 and X_2 are the inputs of the channel, and Y_1 and Y_2 the outputs.

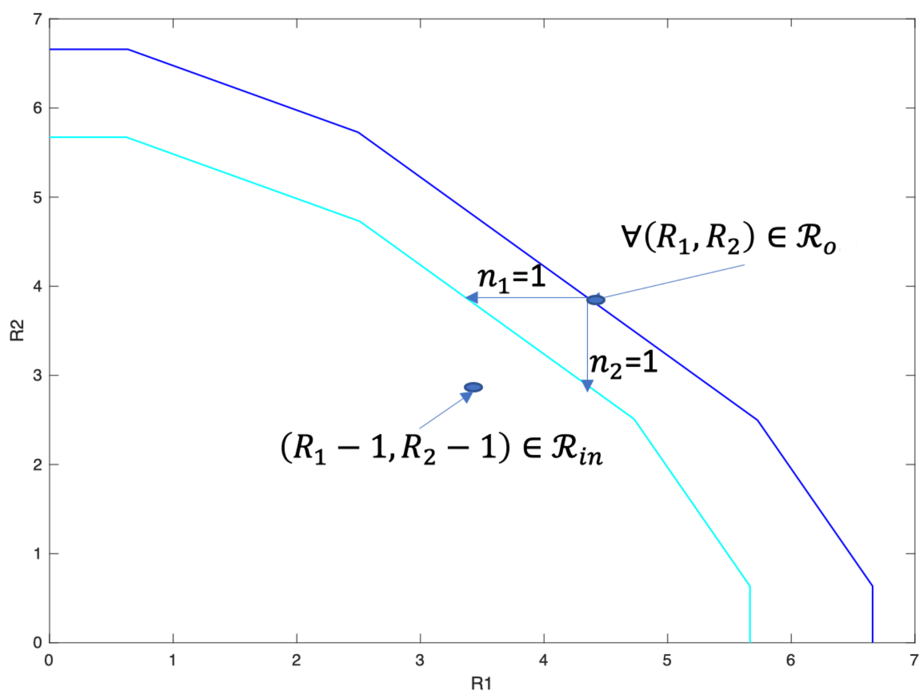

Figure 1.6: Inner and outer bounds and constant gap of the two-user IC given in Fig. 1.5

and $\mathsf{INR}_{21} = \rho^{\alpha_{21}}$, where ρ is a nominal power and α_{11}, α_{22}, α_{12} and α_{21} are the pre-log factors of the channel SNRs and INRs in dB scale. In the following numerical experiments, we keep $\alpha_{11} = \alpha_{22} = 3$ and $\alpha_{12} = \alpha_{21} = 2$ unchanged while increasing the nominal power ρ from 10 to 10^6 in several steps. We plot the inner and outer bounds as well as the normalized (scaled by $\log\rho$) inner and outer bounds in Fig. 1.7 ($\rho = 10$), Fig. 1.8 ($\rho = 100$), Fig. 1.9 ($\rho = 1000$) and Fig. 1.10 ($\rho = 10^6$). The scaling factor $\log\rho$ can be approximately viewed as the capacity of a reference AWGN (additive white Gaussian noise) channel with channel SNR ρ, i.e., $\log\rho \approx \log(1+\rho)$. It can be seen that both the inner and outer bounds expand with the increase of ρ. Since the gap between the inner and outer bounds is constant, it does not grow with the increase of the channel SNRs and INRs. The normalized inner and outer bounds do not expand with ρ, but they tend to coincide. At high SNR regime, some details of the shape of the normalized constant-gap-to-capacity tends to disappear, but the main sketch remains. When ρ goes to infinity, the gap completely disappears, and the normalized inner and outer bounds coincide. This coincided region is the so-called GDoF region, as shown in Fig. 1.11. Precisely, the GDoF region of the two-user scalar Gaussian IC is defined as

$$\mathcal{D}(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{21}) = \left\{(d_1, d_2) = \lim_{\rho\to\infty}\left(\frac{R_1}{\log\rho}, \frac{R_2}{\log\rho}\right) : (R_1, R_2) \in \mathcal{C}\right\}$$

where $\mathcal{C}$ is its capacity region. We do not know the capacity region $\mathcal{C}$, but we can get the exact GDoF region through a constant-gap-to-capacity region $\mathcal{R}_{\text{in}}$, i.e.,

$$\mathcal{D}(\alpha_{11}, \alpha_{22}, \alpha_{12}, \alpha_{21}) = \left\{(d_1, d_2) = \lim_{\rho\to\infty}\left(\frac{R_1}{\log\rho}, \frac{R_2}{\log\rho}\right) : (R_1, R_2) \in \mathcal{R}_{\text{in}}\right\},$$

as a finite number of bits is insignificant in the GDoF computation.

It is worth pointing out that even with a constant-gap-to-capacity region in hand, deriving the GDoF region is not always straightforward, because a constant-gap-to-capacity region could be a union of infinite regions each of which is contributed by one coding scheme. An ideal constant-gap-to-capacity region for GDoF computation contains single polytope (as the one shown in Fig. 1.6) from which the GDoF region can be easily obtained by definition. In this dissertation, we always obtain single region inner and outer bounds for the three channels to be discussed later. On the other hand, the derivation of GDoF region does not always require a constant-gap-to-capacity region. To get the GDoF region, it is sufficient to have an inner or outer bound which is within a ρ independent finite gap to the capacity. We shall adopt this

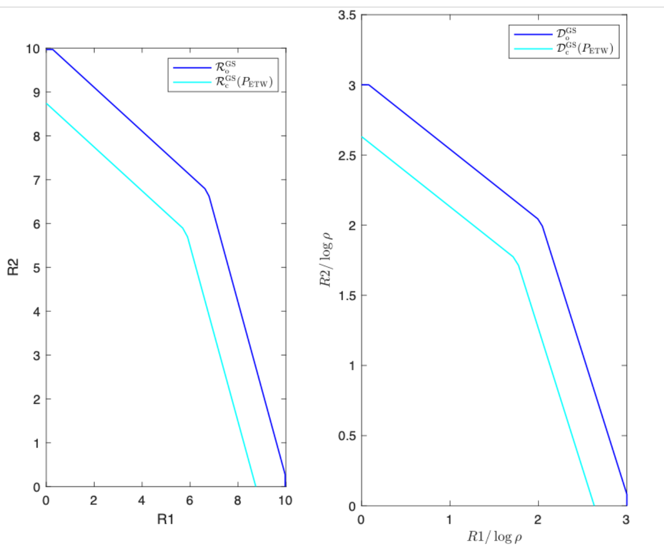

Figure 1.7: The left subplot shows the inner and outer bounds of the scalar Gaussian two-user IC with $\mathsf{SNR}_{11} = \mathsf{SNR}_{22} = 10^3$ and $\mathsf{INR}_{12} = \mathsf{INR}_{21} = 10^2$. The right subplot shows the normalized (scaled by $\log \rho$) inner and outer bounds.

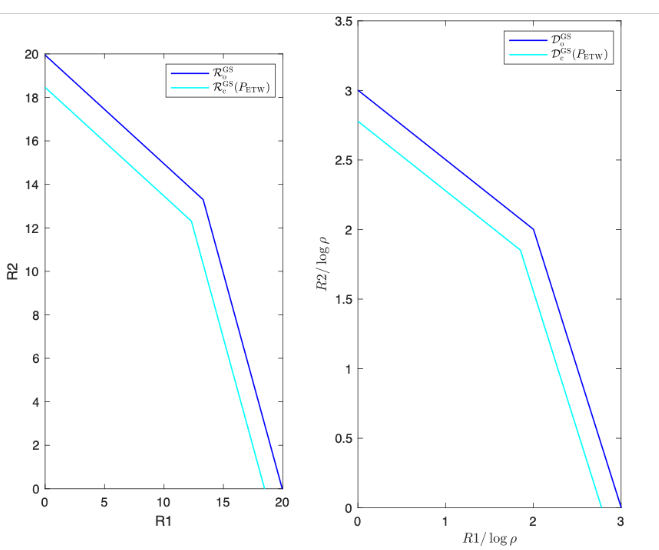

Figure 1.8: The left subplot shows the inner and outer bounds of the scalar Gaussian two-user IC with $\mathsf{SNR}_{11} = \mathsf{SNR}_{22} = 100^3$ and $\mathsf{INR}_{12} = \mathsf{INR}_{21} = 100^2$. The right subplot shows the normalized (scaled by $\log \rho$) inner and outer bounds.

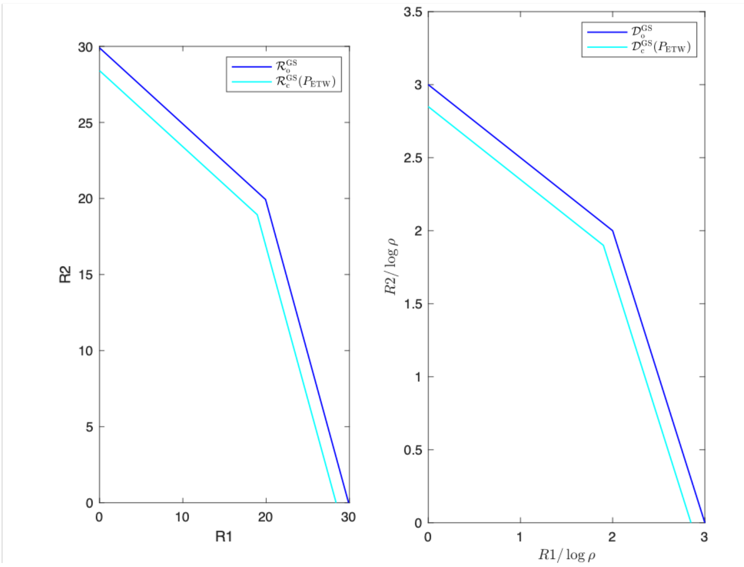

Figure 1.9: The left subplot shows the inner and outer bounds of the scalar Gaussian two-user IC with $\mathsf{SNR}_{11} = \mathsf{SNR}_{22} = 1000^3$ and $\mathsf{INR}_{12} = \mathsf{INR}_{21} = 1000^2$. The right subplot shows the normalized (scaled by $\log\rho$) inner and outer bounds.

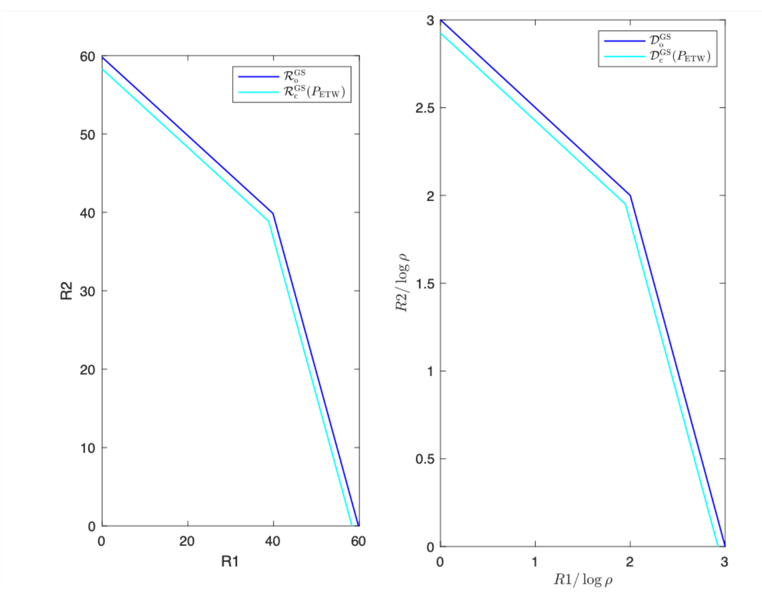

Figure 1.10: The left subplot shows the inner and outer bounds of the scalar Gaussian two-user IC with $\mathsf{SNR}_{11} = \mathsf{SNR}_{22} = 1000000^3$ and $\mathsf{INR}_{12} = \mathsf{INR}_{21} = 1000000^2$. The right subplot shows the normalized (scaled by $\log\rho$) inner and outer bounds.

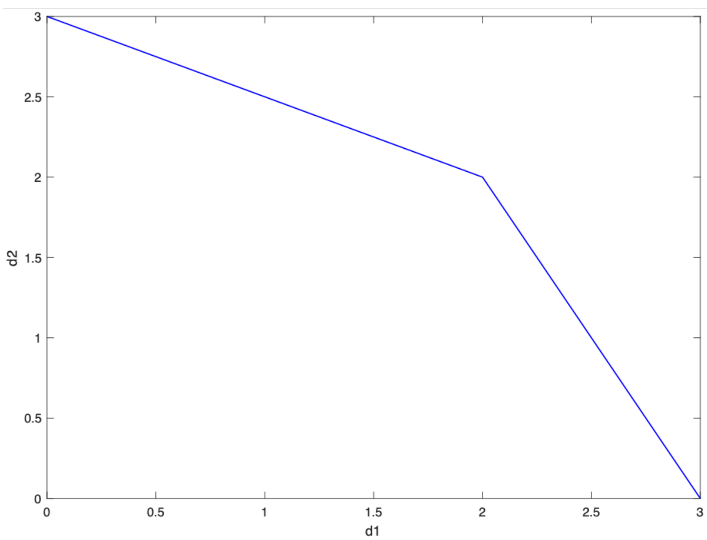

Figure 1.11: The GDoF region of the two-user scalar Gaussian IC with $\alpha_{11} = \alpha_{22} = 3$ and $\alpha_{12} = \alpha_{21} = 2$.

insight to characterize the GDoF regions of the channels to be discussed in Chapters 3 and 4, where the constant-gap-to-capacity regions are undetermined.

1.1.3 The DoF Region

Getting constant-gap-to-capacity or GDoF region for the aforementioned four component channels (the MIMO K-user IC, the MIMO IMAC, the MIMO IBC and the MIMO IMABC) is still difficult, if not as difficult as characterizing their Shannon capacity regions. Many information theory works have aimed down to the next level of channel capacity approximation–the DoF region. A DoF region is not an approximated GDoF region. For a given channel model, its DoF region is the exact GDoF region under a particular channel realization where all the channel SNRs and INRs have the same pre-log factor. In other words, for a given channel model, only one particular channel realization (pre-log factors of all channel SNRs and INRs are equal) has its DoF region defined.

1.2 Milestone Results on Capacity Approximations of MIMO Interference Channels

Previous works on the K-user MIMO interference channel are summarized in this paragraph. So far, the Shannon capacity region of even the simplest K-user MIMO IC, the two-user scalar Gaussian IC with constant channel realization, is still unknown. We first state known results on the K-user MIMO IC with constant channel realizations. Karmakar and Varanasi characterized a constant-gap-to-capacity region of the two-user MIMO IC. Etkin, Tse and Wang [15] characterized a constant-gap-to-capacity region of the two-user scalar Gaussian IC. Jafar and Vishwanath [24] obtained per-user GDoF of the K user symmetric scalar Gaussian IC where, in the notation of this paper, all direct links have the same SNR ρ and all interference links have the same INR ρ^{α}. Next, we state known results on the coarser sum-DoF metric on the K-user MIMO IC with time varying channel realizations. Gou and Jafar [22] provided inner and outer sum-DoF bounds for a class of MIMO K-user ICs with M antennas at each transmitter and N antennas at each receiver, and they showed these bounds are tight when the ratio $\max\{M, N\}/\min\{M, N\}$ is an integer. The sum-DoF of the scalar Gaussian K-user IC was reported in [6]. Interference alignment is considered in the achievable coding schemes in both [22] and [6].

For the MIMO IMAC, the MIMO IBC and the MIMO IMABC, relatively little is known compared to the K-user MIMO IC. Again, we first state results on those channels with constant channel realizations. Pang and Varanasi [37] obtained constant-gap-to-capacity region for the scalar Gaussian MAC-IC-MAC (a partially connected two-cell IMAC where only one interfering transmitter in each cell interferes the receiver in the other cell; it is the scalar version of the channel discussed in Chapter 5). Chaaban and Sezgin studied a fully connected two-cell channel in which a two-user MAC interferes with a point-to-point link [8]. The capacity region is found for very strong and some cases of strong interference, and upper and lower bounds on the sum-rate in the weak interference regime (with the lower bound achievable by treating interference as noise) are also obtained. Subsequently, in [20], they showed that when the interference is weak, treating interference as noise in their model is sub-optimal. Buhler and Wunder [5] derived upper bounds on the sum rate and an achievable scheme for the linear deterministic version of the model in [8]. Fritschek and Wunder obtained a result on the reciprocity between the two-cell deterministic IMAC and the two-cell deterministic IBC in [17], and obtained an achievable region under a weak interference condition for both those channels. In [18], the deterministic IMAC was revisited using the lower triangular deterministic model introduced by [34], and a constant-gap sum capacity was obtained. Fritschek and Wunder [16] closed the gap between the achievable sum rate regions for Gaussian IMAC and the deterministic IMAC. Their coding scheme employs signal scale alignment and lattice coding. For symmetric (with K-user per cell, M antennas per user and N antennas per base station) L-cell MIMO IMAC , Kim *et al* [29] derived an outer bound on its sum DoF and presented an achievable scheme that achieves this outer bound for the case $L = 2$. For symmetric L-cell MIMO IBC, Liu and Yang [32] determined per-user DoF for certain ranges of antenna configurations. For symmetric L-cell MIMO IMAC and IBC, Sridharan and Yu [40] investigated achievable schemes based on decomposition with asymptotic interference alignment and linear beamforming and showed that there are distinct regimes where one outperforms the other. In this work, the per-user DoF is determined for the case when each user is equipped with single antenna. Next, we state results on time/frequency varying channel realizations. For multi-subcarrier Gaussian IMAC and IBC, Suh and Tse [41] applied interference alignment and showed it achieves interference-free per-cell DoF when the number of users in each cell goes to infinity.

Suh, Ho and Tse introduced a downlink interference alignment scheme for multi-subcarrier cellular networks. The scheme only requires local channel state information at each base station, i.e., a base station only needs the CSI of its own cell. Jeon and Suh [25] investigated a two-cell IMABC where each user has single antenna and each base station has multiple antennas, and the sum DoF was characterized.

1.3 The Scope of This Dissertation

As seen from Section 1.2, most of the results on general K-user ($K > 2$) MIMO IC, MIMO IMAC, MIMO IBC or MIMO IMABC stay at DoF level. However, DoF (or sum DoF) approximation does not permit asymmetric scaling of the channel SNRs and INRs in dB scale, severely restricting its applicability to settings where the various received signals from the different transmitters at each receiver are of similar strength.

In this dissertation, we study capacity approximations for MIMO interference channels beyond DoF. We characterize GDoF regions of two particular cases of the three-user MIMO interference channels (the MIMO one-to-three IC and the MIMO IC-ZIC) and the constant-gap-to-capacity region of the MIMO MAC-IC-MAC (a class of partially connected MIMO IMACs). In all these channels to be investigated, we assume constant channel realization and full CSI known to all the involved transmitter and receivers. We always derive single region inner and outer bounds. In other words, each of our constant-gap-to-capacity or GDoF region is achievable by one coding scheme. In what follows, we briefly summarize the research results to be introduced in the next four chapters.

Before we investigate the three proposed channels, we first review the signal-level partitioning technique introduced by Pang and Varanasi in [37, Section III-F] and extended by the same authors to multidimensional signal-level partitioning in [36] (which was initially proposed informally by Karmakar and Varanasi in [26]) in Chapter 2. This technique will be an effective tool to demonstrate the achievability of any given GDoF-tuple that is achievable via multi-level superposition coding in a MIMO network.

Since the MIMO two-user IC already has a constant-gap-to-capacity region characterized in [27], the next goal towards the constant-gap-to-capacity or GDoF region of the K-user MIMIO IC is to find the constant-gap-to-capacity or GDoF region of the MIMO three-user IC, but it is a difficult job. In Chapter 3, we take one beginning step towards this goal by characterizing the GDoF region of the MIMO one-to-three IC, a special case of the three-user MIMO IC where only the first transmitter interferes two other unintended receivers (see Fig. 1.12). The coding scheme is enlightened by the Karmakar-Varanasi type coding scheme (KV coding scheme) [27] in the two-user MIMO IC as well as the multi-level superposition coding scheme used for the scalar Gaussian one-to-many IC [4]. More specifically, at Tx1 we perform three level superposition coding to encode four sub-messages m_{123}, m_{12}, m_{13} and m_{1p} which are intended to be decoded by Tx1-Tx3, Tx1-Tx2, Tx1 and Tx3, and Tx1 only. One challenge in the design of the coding scheme is to seek an appropriate mathematical tool to jointly decompose the two cross link transfer matrices so that the common and exclusive signal directions and levels between Tx1 and Rx2-Rx3 can be revealed. We adopt GSVD to fulfill the duty, but at the cost of losing the constant gap between the derived inner and outer bounds. The gap turns out to be independent of channel SNRs and INRs, but dependent on the channel transfer matrices. Nevertheless, such a pair of bounds is sufficient to characterize the GDoF region.

In Chapter 4, we take one further step by adding an additional interference link from Tx2 to Rx1 to the MIMO one-to-three IC, and the resulting channel is the MIMO IC-ZIC which contains a two-user MIMO IC between Tx1/Rx1 and Tx2/Rx2, and a two-user MIMO Z interference channel (ZIC) between Tx1/Rx1 and Tx3/Rx3 (see Fig. 1.13). With the knowledge of the GDoF optimal coding schemes for the MIMO one-to-three IC and the two-user MIMO IC, it is not hard to conjecture that a three-level superposition coding (as in the MIMO one-to-three IC) at Tx1, KV coding scheme (as in the two-user MIMO IC) at Tx2, and single user random coding at Tx3 (with Gaussian codebook and scaled identity covariance matrix) could be GDoF optimal. We prove that it is indeed the case. However, the mathematic process of quantifying the gap between the derived inner and outer bounds turns out to be challenging. The number of inequalities in the GDoF region of MIMO IC-ZIC increases significantly when compared to its sub-channels–the MIMO one-to-three IC and the two-user MIMO IC, which indicates high complexity of GDoF region of the fully connected three-user MIMO IC.

Lastly, we turn our attention to the MIMO MAC-IC-MAC, where two MACs interfere with each other through two marginal links (see Fig. 1.14). The constant-gap-to-capacity region of the scalar Gaussian MAC-IC-MAC has already been characterized in [37]. We extend it to the MIMO case in Chapter 5. We

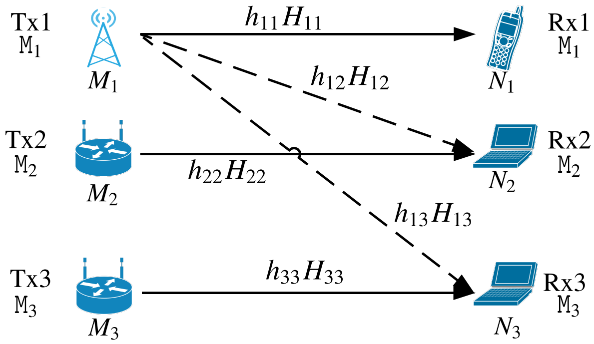

Figure 1.12: The MIMO one-to-three IC

perform KV coding scheme at the interfering transmitter (as in the two-user MIMO IC) and single user random coding (with Gaussian codebook and scaled identity covariance matrix, which is GDoF optimal for the MIMO MAC) at the non-interfering transmitters. The overall coding scheme yields a constant-gap-to-capacity region of the MIMO MAC-IC-MAC. The GDoF result of the MIMO MAC-IC-MAC shows that despite the existence of the interference links, each cell can achieve full sum symmetric GDoF as if it were interference free, as long as the INR to SNR ratio (in dB scale) is within a certain range (either weak enough or strong enough). This suggests that time or frequency sharing among cell users for the interfering uplink cellular network is not GDoF optimal.

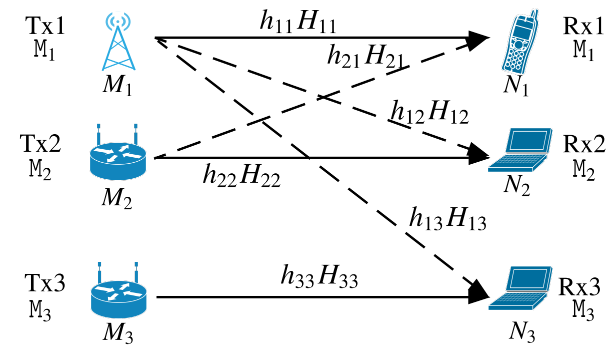

Figure 1.13: The MIMO IC-ZIC

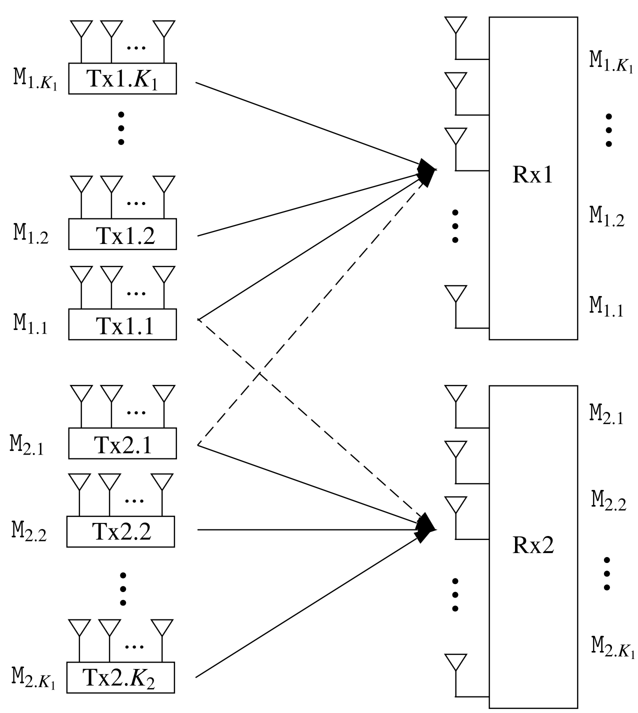

Figure 1.14: The MIMO MAC-IC-MAC

Chapter 2

Multidimensional Signal-Level Partitioning for GDoF Analysis

In this chapter, we generalize the scalar signal-level partitioning technique introduced in the paper [37, Section III-F] to the vector case, by developing a multidimensional signal-level partitioning technique that is suitable for demonstrating the achievability of boundary points (e.g., vertices) of the GDoF region of MIMO networks. This technique can be seen as a formalization of a similar idea introduced in [26, Section III-A] in the context of the 2-user MIMO interference channel, thereby widening its applicability beyond that context.

2.1 Signal-Level Partitioning in Scalar Gaussian Channels

Let us start from a complex-valued scalar AWGN received signal $Y = \sqrt{\rho^{\alpha}}X + Z$ in some link in a network with the additive noise Z being a zero-mean complex Gaussian random variable with unit-variance. Without loss of generality, we normalize the transmit power constraint to be unity by absorbing the signal amplitude into the channel gain, which in turn we denote as $\sqrt{\rho^{\alpha}}$, so that the received $\mathsf{SNR} = \rho^{\alpha}$.

Suppose the signal X is sent with full power, i.e. $E[|X|^2] = 1$. Since unit power can be expressed as

$$1 = (\rho^0 - \rho^{-\alpha}) + (\rho^{-\alpha} - \rho^{-2\alpha}) + (\rho^{-2\alpha} - \rho^{-3\alpha}) + \cdots \tag{2.1}$$

we accordingly let X be the result of linear superposition coding given as

$$X = \sqrt{\rho^0 - \rho^{-\alpha}}X_{p1} + \sqrt{\rho^{-\alpha} - \rho^{-2\alpha}}X_{p2} + \sqrt{\rho^{-2\alpha} - \rho^{-3\alpha}}X_{p3} + \cdots \tag{2.2}$$

where $X_{p1}, X_{p2}, X_{p3}, \cdots$ being mutually independent zero-mean, unit-variance complex Gaussian random variables. Henceforth, we will refer to $X_{p1}, X_{p2}, X_{p3}, \cdots$ as signal partitions since they are associated with powers that results from partitioning the total signal power into multiple levels as in (2.1). After going through a channel with gain $\sqrt{\rho^\alpha}$ (and additive noise with unit power) the part of the signal associated with X_{p1} has power ρ^α so that it can "carry" α GDoF, with the receiver decoding it by treating all other partitions as noise. Consider decoding the next lower signal partition X_{p2} by canceling the effect of the decoded X_{p1} from Y and treating all signal partitions below X_{p2} as noise. Since X_{p2} at the receiver has a power that is below the noise power it cannot carry positive GDoF. Similarly, all other lower partitions $X_{p3}, \cdots$ cannot carry positive GDoF either. Hence, ignoring these partitions, either by treating them as noise or by not transmitting them at all, is without loss of optimality when a capacity characterization that is accurate only up to GDoF is needed. Hence, the transmitted signal can be set to $X = \sqrt{\rho^0 - \rho^{-\alpha}} X_{p1}$ without loss of GDoF optimality. This is depicted in Fig. 2.1 as a vertical bar on the left with one partition X_{p1}. Its top is labeled ρ^0 which depicts the signal (power) level and the bottom $\rho^{-\alpha}$ which represents the next signal-level of X_{p2} if it exists, etc. The channel lifts the top of signal partition X_{p1} to level ρ^α at the receiver and this is depicted in the right hand side of Fig. 2.1. The resulting GDoF of α is hence achievable, which is exactly the GDoF of this AWGN link. In summary, in this link, we need one signal partition with power exponent resolution of α to achieve optimal GDoF. We henceforth refer to X_{p1} as a GDoF-effective partition to indicate it is sufficient to achieve the optimal GDoF.

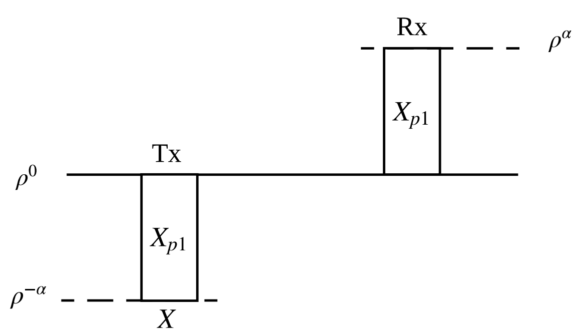

Figure 2.1: Signal-level partitioning at the transmitter and receiver of a Gaussian link $Y = \sqrt{\rho^\alpha} X + Z$ with $E[|X|^2] \leq 1$ and zero mean Gaussian noise $Z \in \mathcal{CN}(0, 1)$

In a network with multiple transmitters and/or receivers each link $i - j$ from transmitter i to receiver j can be modeled as having a channel gain of the form $\sqrt{\rho^{\alpha_{i \to j}}}$ with possibly distinct exponents $\alpha_{i \to j}$ (with ρ being some nominal SNR). In this case, multiple signal partitions are usually needed for each transmit signal to achieve a given GDoF tuple in general. The process of creating these multiple signal partitions by additively decomposing X into a series of ρ-ary components $X_{p1}, X_{p2}, \cdots$ as in (2.2) is called signal-level partitioning of X as explained previously. The number and resolution (i.e., size) of the signal partitions required for each transmit signal are not unique but depend in general on the network topology, all involved channel gain exponents $\alpha_{i \to j}$ of nominal SNR ρ, and the GDoF tuple to be achieved. There are several rules that must govern their selection which we shall reveal as needed in what follows. For instance, we can, without loss of generality, insist that the bottom signal partition (i.e., with the lowest power) should be set so that a partition below it would not be received at above the noise floor by **any** receiver in the network. Moreover, the number of partitions used for each transmit signal should provide sufficient resolution so that the they are aligned at each receiver. Moreover, not all signal partitions are "used" (so that a signal is not always transmitted with full available power) and what partitions to transmit is dictated by the requirement that the given achievable GDoF tuple must be achieved by successive cancellation decoding at each receiver. We illustrate these points in the next example.

Consider a complex Gaussian scalar MAC with received signal

$$Y = \sqrt{\rho^{0.8}} X_1 + \sqrt{\rho^{1.2}} X_2 + Z \tag{2.3}$$

with unit power constraints on X_1 and X_2 and zero-mean, unit-variance complex Gaussian noise. The GDoF region of this MAC is easily shown to be

$$0 \leq d_1 \leq 0.8$$

$$0 \leq d_2 \leq 1.2$$

$$d_1 + d_2 \leq 1.2$$

To achieve the corner point $(0.8, 0.4)$, we decompose the signals X_1 and X_2 into three signal partitions each at power exponent resolution 0.4 as depicted in the left hand side of Fig. 2.2. The resolution is chosen so it divides both the two channel gain exponents 0.8 and 1.2 so that the received signal partitions can be aligned at the receiver's grid as shown in the right hand side of Fig. 2.2. In particular, the top two signal partitions of X_1 and the top three partitions of X_2 can be heard by the receiver (i.e, while carrying positive GDoF), with the top two partitions of X_1 aligned with the bottom two partitions of X_2. To achieve the GDoF pair $(0.8, 0.4)$, we simply use (i.e., transmit) the top two partitions of X_1 which carry 0.8 GDoF together, and the top partition of X_2, which carries 0.4 GDoF. The decoder decodes the three signal partitions sequentially in decreasing order of signal strengths using successive cancellation (i.e., the top partition of X_2 first, the top partition of X_1 next and followed by its second partition). We say that the depth of signal partition is 3 in this scheme at each transmitter (even though the bottom partitions were unused) and the exponent resolution (i.e., the GDoF per partition) is 0.4. Note also that the the GDoF pair $(0.4, 0.8)$ can be easily achieved with the same depth-three signal partitioning and GDoF per partition of 0.4 by using the middle partition of X_1 and the top two partitions of X_2. Evidently, the GDoF pair $(0, 1.2)$ can be achieved by using all three signal partitions of X_2 and none of X_1.

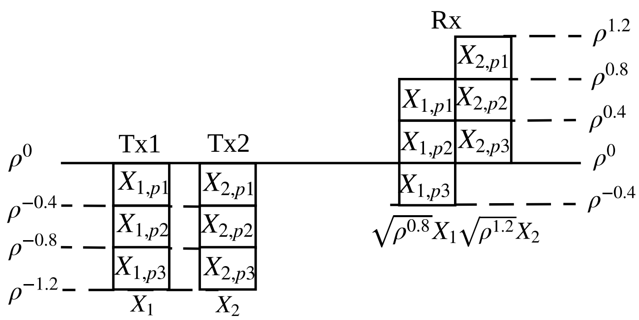

Figure 2.2: Signal-level partitioning of the transmitter and receiver of the MAC $Y = \sqrt{\rho^{0.8}}X_1 + \sqrt{\rho^{1.2}}X_2 + Z$

To summarize, the achievability analysis using signal partitioning restricts the encoding scheme to be multi-level superposition coding with signal-level alignment, and with each signal partition encoded independently with a different message, i.e. there is no cross-partition encoding. The decoding scheme is usually successive decoding at each receiver, except in certain cases when it is not sufficient, in which case joint decoding must be used, as illustrated in Section 2.4. In successive decoding, the decoder decodes signal partitions from top to bottom sequentially. Each signal partition is decoded by treating all the signal partitions below it as noise. The GDoF per partition should not only be set to be a common integer divisor of all the channel gain exponents $\alpha_{i\to j}$, but also of the individual values in the given GDoF tuple to be achieved (as was done in the achievable GDoF pairs considered for the MAC). We illustrate this point with another example.

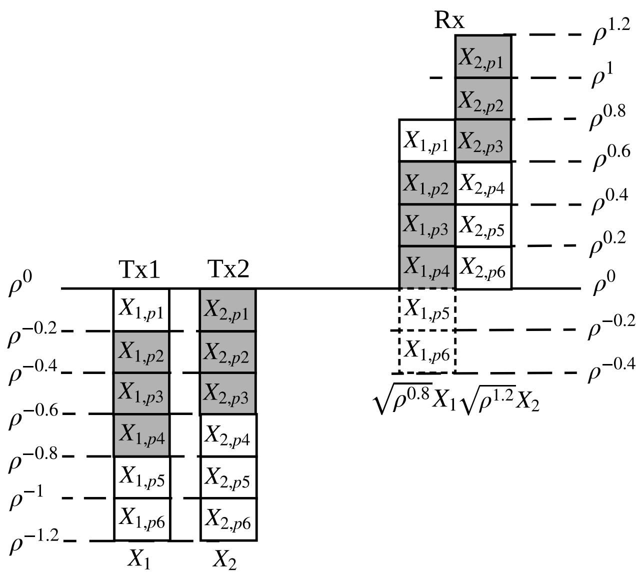

Figure 2.3: Coding scheme for $(d_1, d_2) = (0.6, 0.6)$ in MAC $Y = \sqrt{\rho^{0.8}}X_1 + \sqrt{\rho^{1.2}}X_2 + Z$

Continuing with the example of the MAC of (2.3), we next demonstrate the achievability of a maximum symmetric (and sum) GDoF pair $(d_1, d_2) = (0.6, 0.6)$. The depth-three partition with GDoF per partition of 0.4 of Fig. 2.2 does not suffice. Consider the depth-six partition of Fig. 2.3 in which the GDoF per partition is 0.2. The GDoF pair $(d_1, d_2) = (0.6, 0.6)$ can be achieved as follows: since the top two partitions of X_2 alone can reach power levels $\rho^{1.2}$ and ρ^{1} at the receiver, they **must** be utilized by Tx2 to get maximum sum GDoF. However, the receiver's power levels below $\rho^{0.8}$ can be shared by signal partitions of three partitions of Tx1 and the remaining one of Tx2. One such achievability scheme is the one shown in Fig. 2.2 where the shaded partitions of X_1 and X_2 (i.e., the second to the fourth of X_1 and the top three of X_2) are the ones that are used by the two transmitters. The receiver successively decodes the shaded partitions it sees from top to bottom using successive cancellation. Evidently, the higher resolution signal partitioning of Fig. 2.2 can be used to specify the partitions that must be used to achieve the three GDoF pairs that were achievable with the signal partitioning of Fig. 2.2 but also the new GDoF pair $(d_1, d_2) = (1.0, 0.2)$.

2.2 Multidimensional Signal-Level Partitioning

When an input signal is a vector, we can employ signal-level partition to each element or dimension of the vector individually at sufficient resolution. For convenience, we refer to each element of the transmit or receive signal simply as a "dimension" of the transmit and receive signal, respectively. For example, the input signal $X = \left(X^{(1)}\ X^{(2)}\right)^T$ is a two dimensional signal, with $X^{(1)}$ being its first dimension, and $X^{(2)}$ the second dimension. Similarly, in the output signal $Y = \left(Y^{(1)}\ Y^{(2)}\right)^T$, $Y^{(1)}$ is its first dimension and $Y^{(2)}$ is its second dimension.

Consider a 2×2 MIMO AWGN link $Y = \sqrt{\rho^{\alpha}} HX + Z$ where the channel matrix $H = \begin{pmatrix} [H]_{11} & [H]_{12} \\ [H]_{21} & [H]_{22} \end{pmatrix}$ satisfies the assumptions stated in Section 5.2 with $Z \sim \mathcal{CN}(0, I_2)$. The maximum achievable GDoF in this point-to-point link is clearly 2α. We show that 2α can be achieved by having each dimension of X carry α GDoF using a single signal partition. Let $X_{p1}^{(1)}$ and $X_{p1}^{(2)}$ be these GDoF effective partitions of $X^{(1)}$ and $X^{(2)}$. They are depicted on the left hand side of Fig. 2.4. The signal diagram on the right hand side depicts the two dimensions of the output signal Y. Each is a linear combination of the two input signals scaled by ρ^{α}, i.e. $Y^{(i)} = \sqrt{\rho^{\alpha} - \rho^{0}}[H]_{i1}X_{p1}^{(1)} + \sqrt{\rho^{\alpha} - \rho^{0}}[H]_{i2}X_{p1}^{(2)}$ for $i \in \{1, 2\}$. Due to the effect of entries of H, in a signal partition depiction, the tops of the signals $\sqrt{\rho^{\alpha} - \rho^{0}}[H]_{11}X_{p1}^{(1)}$, $\sqrt{\rho^{\alpha} - \rho^{0}}[H]_{12}X_{p1}^{(2)}$, $\sqrt{\rho^{\alpha} - \rho^{0}}[H]_{21}X_{p1}^{(1)}$ and $\sqrt{\rho^{\alpha} - \rho^{0}}[H]_{22}X_{p1}^{(2)}$ are almost never exactly aligned. However —and this is a crucial point —the misalignment will effectively disappear as ρ tends to infinity, and so we can depict the tops of these signal partitions as being aligned as shown in Fig. 2.4. On the other hand, the channel matrix has full rank w.p.1 according to its definition, which ensures that $Y^{(1)}$ and $Y^{(2)}$ are two linearly independent combinations of $X_{p1}^{(1)}$ and $X_{p1}^{(2)}$, both at power level ρ^{α}. Hence, decoding the received signal using zero-forcing will result in 2α GDoF, the exact GDoF of this link.

Remark 2.1. The deterministic model introduced in [2] in contrast involves a bit-level signal partitioning. It differs from the one discussed here in that it expresses a real-valued signal in its binary expansion $X = \sum_{i=1}^{\infty} X_{bi} 2^{-i}$ with $X_{bi} \in \{0, 1\}$. A link with gain $\sqrt{\rho}$ lifts the $\left\lceil \log_2 \sqrt{\rho} \right\rceil^+$ most significant bits of the input signal above the noise level. With channel gains approximated in this way and signals at or below noise level and the noise neglected, the capacity regions of the resulting deterministic P2P, MAC and broadcast

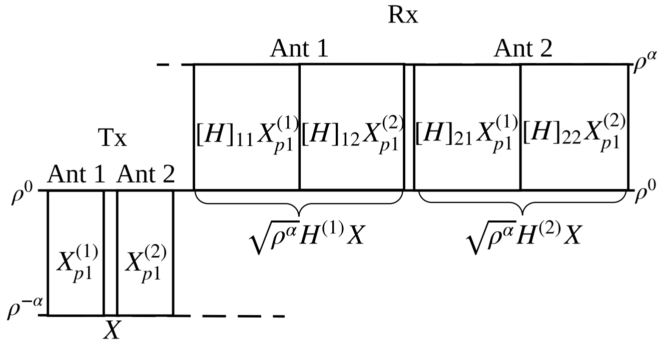

Figure 2.4: Signal-level partitioning in a 2×2 MIMO link $Y = \sqrt{\rho^\alpha}HX + Z$. The transmitted signal consists of one signal partition per dimension so that $X = [X_{p1}^{(1)}, X_{p1}^{(2)}]^T$ as depicted on the left (transmitter) side of the figure. The i^{th} row of H is denoted as $H^{(i)}$. The first two and last two signal partitions on the right (receiver) side denote the components due to $X_{p1}^{(1)}$ and $X_{p1}^{(2)}$ in the signals received at the first and second antennas, respectively. While the tops of these components wouldn't be aligned at finite SNR, they can be regarded as being effectively aligned, since it is the limit of high SNR that is relevant in GDoF analysis.

channels are seen to be constant-gap-to-capacity approximations of their underlying Gaussian scalar P2P, MAC, and broadcast channels, respectively [2, Section II-A-C]. However, a representative example of the MIMO P2P channel is given to illustrate the limitation of the deterministic model in [2, Section II-E]. In it, each element of the channel matrix is approximated in the deterministic model in the same way as channels gains in the scalar Gaussian channels (so that the misalignment of signal components arriving at the receiver would disappear). However, this leads to a reduction in the rank of the channel matrix in the deterministic model for the example considered, so that the gap between the capacity of the original Gaussian MIMO channel and that of its deterministic approximation becomes unbounded with increasing ρ, highlighting the shortcoming of the deterministic model of [2] for MIMO channels.

In contrast, our additive, linear, superposition coding-based ρ-ary signal-level partitioning of (2.2) employed in each dimension of the transmitted signal, with each partition in each dimension carrying α GDoF, leaves the entries of the channel matrix (and hence its rank) unaltered, and, since we are only interested in the limiting (GDoF) analysis as $\rho \to \infty$, we effectively have the alignment of signal partitions at the receiver as well.

Similar to the scalar signal-level partitioning, multidimensional signal-level partitioning also restricts the encoding scheme to be multi-level superposition coding with signal-level alignment and decoding to be based on successive cancellation (after channel inversion). Moreover, for simplicity, we preclude coding across signal dimensions in addition to disallowing coding across levels as in the scalar case. For example, in the case of the MIMO link, with two signal partitions per dimension, $X_{p1}^{(1)}, X_{p2}^{(1)}, X_{p1}^{(2)}$ and $X_{p2}^{(2)}$ are all encoded independently. Hence, the covariance matrix of the transmit signal X should be diagonal.

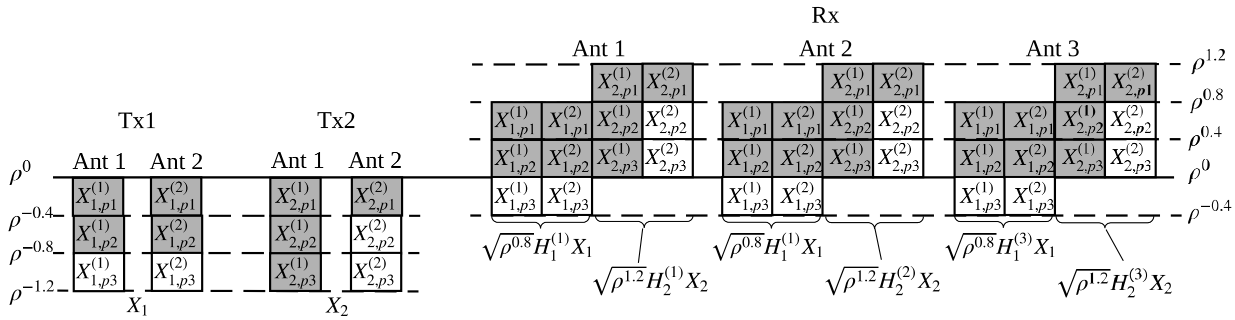

Figure 2.5: Multidimensional signal-level partitioning in a (2,2,3) MIMO MAC $Y = \sqrt{\rho^{0.8}}H_1X_1 + \sqrt{\rho^{1.2}}H_2X_2 + Z$ to achieve the GDoF pair (1.6, 1.6).

To be concrete, we explain the achievability of a GDoF pair in a (2,2,3) MIMO MAC (i.e., a two-user MAC in which the transmitters have two antennas each and the receiver has three antennas) using multidimensional signal-level partitioning next. Consider the input-output relationship of the (2,2,3) MIMO MAC to be

$$Y = \sqrt{\rho^{0.8}}H_1X_1 + \sqrt{\rho^{1.2}}H_2X_2 + Z$$

It is easily shown that the GDoF region of this MIMO MAC is given as the closure of $(d_1, d_2) \in \mathbb{R}_+^2$ which satisfies:

$$0 \leq d_1 \leq 1.6$$

$$0 \leq d_2 \leq 2.4$$

$$d_1 + d_2 \leq 3.2.$$

We will show the achievability of the corner point $(1.6, 1.6)$. Depth-three signal partitioning is employed for both dimensions of X_1 and X_2 as shown in Fig. 2.5 with GDoF per partition taken to be 0.4. The shaded signal partitions in Fig. 2.5 denote the ones that are used for transmission. Due to limitation of space, and since they can be naturally inferred, we omit the entries of the channel matrices H_1 and H_2

associated with the various signal components/partitions/dimensions in the figure. The particular choice of partitions to transmit in at Tx1 and Tx2 is made as follows: since the signal-level $\rho^{1.2}$ can only be reached by Tx2, we let Tx2 send messages on partitions $X_{2,p1}^{(1)}$ and $X_{2,p1}^{(2)}$ which together contain 0.8 GDoF. Moreover, since the receiver has only three antennas, we need to ensure that there are at most three different signal partitions that arrive at the receiver on levels $\rho^{0.8}$ or $\rho^{0.4}$ so that those partitions can be recovered via channel inversion. These levels are in turn both accessible by Tx1 and Tx2. Tx1 has no choice but to use both of the corresponding partitions (i.e., the top two partitions) in both dimensions to achieve a total of $d_1 = 1.6$ GDoF. Hence, Tx2 can only use one of the two partitions $X_{2,p2}^{(1)}$ or $X_{2,p2}^{(2)}$ and one of $X_{2,p3}^{(1)}$ or $X_{2,p3}^{(2)}$ to achieve an additional 0.8 GDoF for a total of 1.6 GDoF. This explains the choice of signal-level partitions and dimensions in Fig. 2.5 to achieve the GDoF corner point $(1.6, 1.6)$.

Thus far, we have seen the extension of scalar signal-level partitioning to vector or multidimensional signal-level partitioning is a straightforward way to demonstrate the achievability of a GDoF tuple of a MIMO network. However, when the numbers of antennas and/or transmitters/receivers increase, the complexity of the diagram grows significantly. We need to simplify the tool while still maintaining its usability and accuracy. This is what we do next.

Note that in the previous example of the MIMO MAC, if the signal diagram at the receiver is known, the transmit signal diagram at each transmitter can be uniquely determined using the channel gain exponents. Therefore, we can remove the transmit signal diagram altogether since it can be inferred. Also, note that the signal diagram for the receiver in Fig. 2.5 is a repetition of a per-receive antenna signal diagram as many times as there are number of receive antennas. This repetition is important since it can visually be verified that at each signal level there are no more partitions than the number of repetitions (or receive antennas) so that channel inversion can be performed to recover those partitions. Hence, with the understanding that the number of shaded (i.e., used) signal partitions is never made to exceed the number of dimensions of the receiver's signal space at a given signal level, we can make do with a signal diagram that captures just the per-receive-antenna diagram. Moreover, signal partitions at the same level can be moved as long as they are kept at the same level. This allows us to further simplify the picture at the receiver. Using these ideas, the signal diagram of Fig. 2.5 can be succinctly depicted as in Fig. 2.6. Note that in Fig. 2.6 we explicitly retain with each signal partition its provenance, i.e., the transmitter from when it was sent and the dimension number on which it was transmitted. Hence, Fig. 2.6 retains the critical information about how many GDoF each transmitter has at each level at the receiver (that it transmits and the receiver recovers). Note that the order (from left to right) of the three bars in Fig. 2.6 is immaterial.

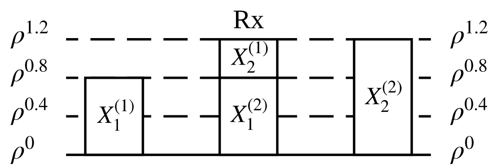

Figure 2.6: A succinct depiction of the multidimensional signal-level partitioning of Fig. 2.5 to achieve the GDoF pair (1.6, 1.6) in the (2,2,3) MIMO MAC $Y = \sqrt{\rho^{0.8}}H_1X_1 + \sqrt{\rho^{1.2}}H_2X_2 + Z$.

In summary, to use multidimensional signal-level partitioning for analysis, the following rules needs to be satisfied.

(1) All dimensions of a transmit signal will be amplified by the same channel gain.

(2) At a given receiver's signal level, the number of dimensions assigned to a particular transmitter cannot exceed the number of that transmitter's antennas.

(3) At a given receiver's signal level, the number of signal partitions assigned cannot exceed the number of receive antennas.

Recall we have used the notation $X_i^{(j)}$ as the j-th component (dimension) of the signal X_i. Hence, denote the total GDoF of signal partitions assigned to $X_i^{(j)}$ as $d_i^{(j)}$ so that the total GDoF carried by X_i is $d_i = \sum_j d_i^{(j)}$. Thus, all signals and the GDoF they carry can be depicted with the set $\{X_i^{(j)}\}_{i,j}$ and their associated partitions in the signal diagram from receiver's perspective, as shown in Fig. 2.6 for the achievement of the GDoF pair (1.6, 1.6) in the example of the (2,2,3) MIMO MAC. The two bottom signal-level partitions of Tx1 and Tx2 are assigned following the rules stated above. This simplified signal-level partitioning still clearly demonstrates the achievability of the corner point $(1.6, 1.6)$. Henceforth, we will adopt this simplified signal-level partitioning for vector signals. The final ingredient we must introduce is transmit beamforming.

2.3 Multidimensional Signal Partitioning with Beamforming

It remains to formulate the case when at least one transmitter employs beamforming. Consider a (2,2,1,1) MIMO Z interference channel where Tx1 and Rx1 have two transmit antennas each while Tx2 and Rx2 have one antenna each:

$$Y_1 = \sqrt{\rho} H_{11} X_1 + Z_1$$

$$Y_2 = \sqrt{\rho^{0.5}} H_{12} X_1 + \sqrt{\rho} X_2 + Z_2$$

According to [26, Theorem 1], we know the corner point $(1.5, 1)$ is achievable through superposition coding and transmit beamforming at Tx1. Its transmitted signal X_1 is the sum of two Gaussian random vectors X_{1c} and X_{1p} whose covariance matrices are

$$\begin{aligned} \mathrm{Cov}[X_{1c}] &= \frac{1}{2}\left(I_{M_1} - \left(I_{M_1} + \rho^{0.5} H_{12}^{\dagger} H_{12}\right)^{-1}\right) \\ \mathrm{Cov}[X_{1p}] &= \frac{1}{2}\left(I_{M_1} + \rho^{0.5} H_{12}^{\dagger} H_{12}\right)^{-1} \end{aligned}$$

Hence, X_1 can then be written as

$$\begin{aligned} X_1 &= X_{1c} + X_{1p} \\ &= V_{1.1\to 2} \begin{pmatrix} \sqrt{\frac{\rho^{0.5}\sigma^2_{1.1\to 2,k}}{2(1+\rho^{0.5}\sigma^2_{1.1\to 2,k})}} \mathtt{X}_{1c}^{(1)} \\ 0 \end{pmatrix} \\ &\quad + V_{1.1\to 2} \begin{pmatrix} \sqrt{\frac{1}{2(1+\rho^{0.5}\sigma^2_{1.1\to 2,k})}} \mathtt{X}_{1p}^{(1)} \\ \frac{1}{\sqrt{2}} \mathtt{X}_{1p}^{(2)} \end{pmatrix} \end{aligned}$$

where $\mathtt{X}_{1c}^{(i)}, \mathtt{X}_{1p}^{(i)} \sim \mathcal{CN}(0,1)$ describe the i-th independent data streams of public and private messages, respectively. The unitary beamforming matrix $V_{1.1\to 2}$ ensures $\mathtt{X}_{1p}^{(2)}$ will not interfere at Rx2 as this data stream will be sent in the null space of H_{12} (i.e., the second column of $V_{1.1\to 2}$ is the basis vector of the one-dimensional null space of H_{12} and its first column that of the orthogonal complement). To reflect beamforming in the signal-level partitioning, we partition and analyze the pre-beamforming signal

$$\mathtt{X}_1 = \begin{pmatrix} \mathtt{X}_1^{(1)} \\ \mathtt{X}_2^{(2)} \end{pmatrix}$$

$$\triangleq \begin{pmatrix} \sqrt{\frac{\rho^{0.5}\sigma^2_{1.1\to 2,k}}{2(1+\rho^{0.5}\sigma^2_{1.1\to 2,k})}}\mathtt{X}^{(1)}_{1c} + \sqrt{\frac{1}{2(1+\rho^{0.5}\sigma^2_{1.1\to 2,k})}}\mathtt{X}^{(1)}_{1p} \\ \frac{1}{\sqrt{2}}\mathtt{X}^{(2)}_{1p} \end{pmatrix}$$

instead of the direct transmit signal $X_1 = V_{1.1\to 2}\mathtt{X}_1$. When transmit beamforming is not needed, the transmitted signal can also be viewed as a pre-beamforming signal scaled by the reciprocal of the number of transmit antennas. In such an example, we would have $X_2 = \mathtt{X}_2$.

To achieve the GDoF corner point $(1.5, 1)$ in the (2,2,1,1) MIMO Z interference channel, we do not assign any part of user 1's message to $\mathtt{X}^{(1)}_{1c}$, but both the private data streams are fully utilized with $\mathtt{X}^{(1)}_{1p}$ assigned 0.5 GDoF and $\mathtt{X}^{(2)}_{1p}$ assigned 1 GDoF. The signal diagram that depicts this strategy at each of the two receivers is shown in Fig. 2.7. The key point here is that because of zero-forcing transmit beamforming, $\mathtt{X}^{(2)}_{1p}$ is not heard at receiver 2 and is hence not shown in Rx2's signal diagram in Fig. 2.7. We have to assign 1 GDoF to $\mathtt{X}_2$ to achieve the (1.5, 1) GDoF pair, but this is clearly decodable at Rx2 since $\mathtt{X}^{(1)}_{1p}$ arrives below the noise level at Rx2. Since Rx1 has two antennas it can decode $\mathtt{X}^{(1)}_{1p}$ and $\mathtt{X}^{(2)}_{1p}$ for a total of 1.5 GDoF. Note that any increase in the GDoF carried by $\mathtt{X}^{(1)}_{1p}$ will show up above the noise level at Rx2 so that it can only come at the price of a corresponding decrease in the GDoF carried by $\mathtt{X}_2$. For example, it is easily verified that the GDoF pair $(d_1, d_2) = (2, 0.5)$ is also achievable.

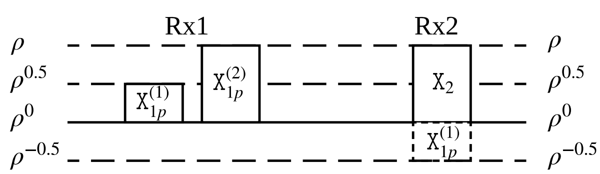

Figure 2.7: Signal-level partitioning of a (2,2,1,1) MIMO Z interference channel

With beamforming at Tx1, the pre-beamforming signal $\mathtt{X}_1$ still has an identity covariance matrix with independent encoding for $\mathtt{X}^{(i)}_{1,c}$ and $\mathtt{X}^{(i)}_{1,p}$. This is in accordance with the previous specification that each dimension and each signal level of the considered transmitted signal is encoded independently, only now, that specification applies to pre-beamformed signals.

2.4 Signal-Level Partitioning with Joint Decoding

From the previously discussed examples, it may seem as if successive cancellation decoding at each receiver would work for any coding scheme based on the signal-level partitioning method. This is however not true, as we illustrate with an example next. Within the framework of our partitioning method, the encoding scheme is restricted to be multi-level superposition coding with signal-level alignment. We show that such a restriction can produce cases when successive cancellation is insufficient to decode all the signal partitions at at least one receiver. Consider a Gaussian scalar two-user interference channel with input-output

$$Y_1 = \sqrt{\rho}X_1 + \sqrt{\rho^{1.5}}X_2 + Z_1$$

$$Y_2 = \sqrt{\rho^{1.5}}X_1 + \sqrt{\rho}X_2 + Z_2.$$

In this channel, each receiver receives interference that is stronger than its intended signal. According to the GDoF region from the work on the Gaussian two-user IC [15], it is known that the GDoF pair $(d_1, d_2) = (1, 0.5)$ is a corner point of the GDoF region. To demonstrate its achievability with a signal partitioning scheme shown in Fig. 2.8, we demonstrate that successive cancellation decoding is not sufficient at Rx 2. Joint decoding must be used as we explain next.

Since both the interferences are strong, the two transmitters do not send private sub-messages, i.e., $\mathtt{X}_1 = \mathtt{X}_{1c}$ and $\mathtt{X}_2 = \mathtt{X}_{2c}$. We let the GDoF per partition be 0.5 since the GDoF pair to be achieved in (1,0.5). At Rx1, we must have two signal partitions at levels $\rho^{0.5}$ and ρ^1 to receive $\mathtt{X}_{1c}$ in order to achieve $d_1 = 1$. Moreover, we have to ensure that the interference $\mathtt{X}_{2c}$ from Tx2 arrives at the partition level $\rho^{1.5}$ at Rx1 in order to achieve $d_2 = 0.5$ without causing GDoF reduction at Tx1. By adopting this GDoF allocation at Rx1, the signal diagram at Rx2 is then uniquely determined as shown in Fig. 2.8. Note that the bottom partition of $\mathtt{X}_{1c}$ overlaps with $\mathtt{X}_{2c}$ at Rx2. Hence, Rx2 cannot decode $\mathtt{X}_{1c}$ and $\mathtt{X}_{2c}$ sequentially using successive decoding. In Fig. 2.8, we slightly shift the signal partition of X_{2c} to show this overlap and to indicate the insufficiency of successive decoding. However, consider the MAC formed by (Tx1,Tx2)→Rx2 with the given transmit power setting. It has the following GDoF region

$$0 \le d_1 \le 1.5$$

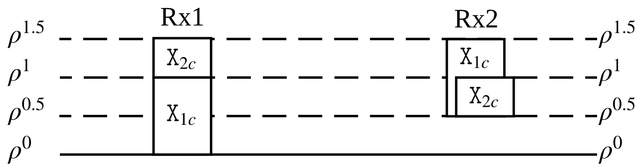

Figure 2.8: A strong interference channel where successive cancellation at Rx2 is not sufficient to demonstrate the GDoF achievability of $(1, 0.5)$ by signal-level partitioning

$$0 \leq d_2 \leq 1$$

$$d_1 + d_2 \leq 1.5$$

so that $(d_1, d_2) = (1, 0.5)$ lies on the boundary of this region, and is hence achievable by joint decoding. Note that the necessity of joint decoding is not predicated on choosing the GDoF per partition to be 0.5 (it could be smaller) nor on the assignment of both messages to be entirely common messages to be decoded at both receivers.

Chapter 3

Generalized Degrees of Freedom Region of the MIMO One-To-Three Interference Channel

3.1 Introduction

A Gaussian MIMO one-to-many interference channel (MIMO one-to-many IC) is a single-hop multi-terminal network which models spectrum sharing scenarios where there is only one communication system producing interference to all the others. One cause of one-sided interference is the disparity of transmission power among different communication systems that coexist on the same frequency band. For example, as shown in Fig. 3.1, the entire area in the figure is a macro cell covered by the radio tower Tx1, and two small cells operate on the same carrier frequency inside the macro cell. The transmit power used by the macro cell transmitter Tx1 is higher than the transmit power at Tx2 and Tx3 in the two small cells. We use solid lines to represent direct links and dashed lines interference links. The interference pattern shown in the figure is a consequence of the disparity of transmit power and network topology. One such application of this scenario is the cellular network range expansion by deploying multiple lower power pico eNBs (Tx2 and Tx3) under a macro cell centered with a macro eNB (Tx1) [3, Figure 1]. Due to the disparity of the transmit power, the interference from Tx1 to Rx2 or Rx3 is significantly stronger than the interference strength from either Tx2 or Tx3 to Rx1. Therefore, the interference from the small cell transmitters to Tx1 is negated. Also, as seen from Fig. 3.1, because the small cell 2 is located further from Tx1 than the small cell 1, the interference strength from Tx1 to Rx2 is stronger than from Tx1 to Rx3. Even without transmit power disparity, the topology of the network along could causes one-sided interference. For example, as shown in

Fig. 3.2, both Rx2 and Rx3 are located within the radio range of Tx1, but Rx1 is not in the radio range of Tx2 or Tx3. In this scenario, all the communication pairs are assumed to transmit at the same power level, but the locations of the transmitters and receivers result in one-sided interference from Tx1 to Rx2 and Rx3. The path-loss difference from Tx1 to Rx2 and Rx3 yields disparity of the interference strength at the two receivers. As most modem wireless systems have implemented FDM (frequency-division multiplexing), TDM (time-division multiplexing), or both to provide services to multiple users, there is effectively one user communicating with the infrastructure on a given sub-carrier and in a given time slot. Therefore, the one-to-many IC practically captures the essence of the one-sided interference issue in most modern spectrum sharing applications.

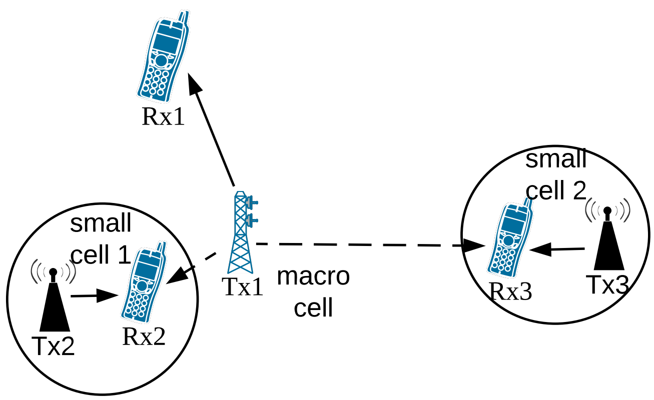

Figure 3.1: A MIMO one-to-three interference channel where the macro cell transmitter transmits at significantly higher power level than the small cell transmitters, causing one-sided interference

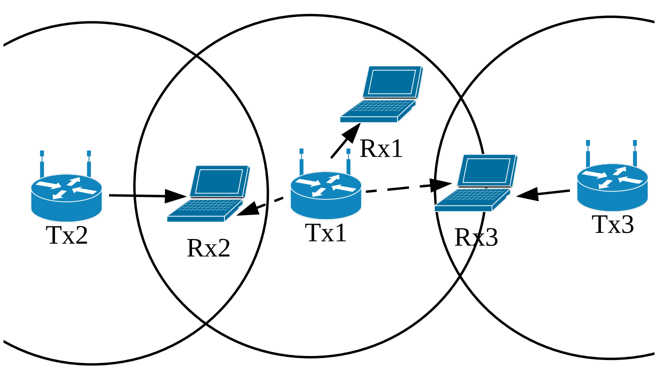

Figure 3.2: A one-to-many IC where Rx2 and Rx3 are located within the radio coverage of Tx1, causing one-sided interference

The simplest one-to-many IC is the Z interference channel which contains only two transmit-receiver pairs, with one of the transmitters interfering the unintended receiver. Constant-gap-to-capacity regions of scalar and vector Gaussian (or MIMO) Z interference channels can be inferred from the work of the scalar and MIMO two-user interference channels by [15] and [27], respectively. On the other hand, Bresler **et al** characterized a constant-gap-to-capacity region for general **scalar** Gaussian one-to-many interference channel in [4]. To the best of our knowledge, the constant-gap-to-capacity region of the vector Gaussian or MIMO Gaussian one-to-many IC remains an open problem. Since the multiple-antenna transmission and reception (also known as multi-input, multi-output or MIMO) has become popular in modern wireless networks, the study of coding scheme for MIMO one-to-many interference channel has important practical value in supervising the design of the next generation mobile network with spectrum sharing. In this chapter, we take the first step towards this problem by tackling the fundamental generalized degrees of freedom (GDoF) region of the three-user case. As an outcome of this research, the sum GDoF curve of the scenario shown in Fig. 3.1 is plotted in Fig. 3.3 with a practical set of channel parameters given in the caption. The strict definition channel model and parameters will be given in Section 3.2.2, and the GDoF region and sum GDoF curve with respect to Fig. 3.2 will be defined and studied in Section 3.5.

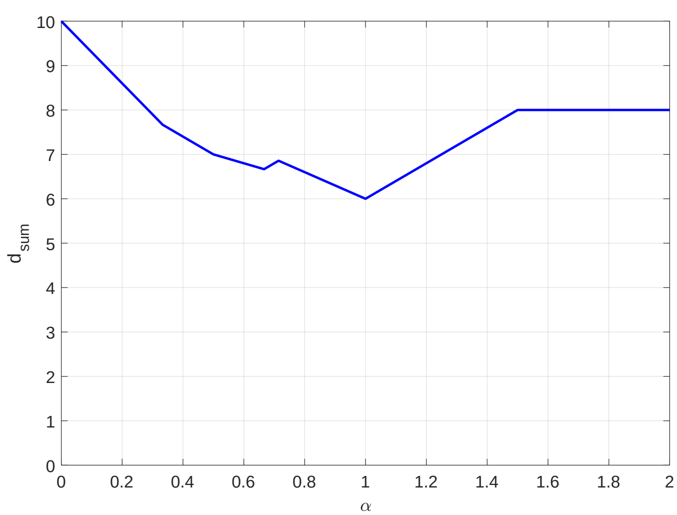

Figure 3.3: Sum GDoF curve of the scenario shown in Fig. 3.1. The parameters are chosen as follows. Tx1 and Rx1 are equipped with 3 antennas each; Tx2, Rx2, Tx3 and Rx3 are equipped with 2 antennas each; Tx1, Tx2 and Tx3 transmit at power $\rho^{2\alpha}$, ρ^{α} and ρ^{α} (to reflect the transmit power disparity); the interference strength from Tx1 to Rx2 and Rx3 are ρ^{α} and $\rho^{\alpha/2}$, respectively.

3.1.1 Main Contributions

We obtain single region inner and outer bounds for the MIMO one-to-three IC that are within SNR and INR independent gap. The achievable scheme for the inner bound involves three-level additive superposition coding with four sub-messages at the interfering transmitter and single-user coding (without water-filling) at the non-interfering transmitters. The message $\mathtt{M}_1$ at the interfering transmitter 1 is split into four sub-messages, namely $\mathtt{M}_{1p}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{123}$. As their subscripts indicate, they are to be decoded at Rx1 only, Rx1 and Rx2, Rx1 and Rx3, and Rx1-Rx3, respectively. The four sub-messages are coded independently according to a vector Gaussian distribution with **explicitly** specified covariance matrices, and they are additively superposed and transmitted. In particular, those covariance matrices are specified via the generalized singular value decomposition (GSVD) of the cross channel matrices. Consequently, a single and explicit polyhedral inner bound is obtained. As a by product, a per-distribution inner bound is also obtained for the discrete-memoryless one-to-three IC. The outer bound is obtained by providing various combinations of genie information to the receivers. The gap between the inner and outer bounds is quantified and shown to be independent of SNRs and INRs (with increasing nominal SNR). Hence, such a gap is tight enough to characterize the fundamental generalized degrees of freedom (GDoF) region. In the end, we analyze the GDoF and sum GDoF achievability of several channel examples with multi-dimensional signal level partitioning introduced in Chapter 2.

3.1.2 Previous Related Work

For the two user MIMO interference channels, Karmakar and Varanasi characterized a constant-gap-to-capacity region in [27] and GDoF region in [26], which lays the foundation of the coding scheme of the MIMIO one-to-three IC in this chapter. A constant-gap-to-capacity region of the Gaussian scalar one-to-many IC was reported by [4], and the resulting GDoF region will be reinforced in this chapter. We shall demonstrate a smaller gap for the SISO one-to-three IC than the one derived in [4]. The idea of genie-aided argument in the proof of the outer was first introduced in the work of the semi-deterministic two-user interference by Telatar and Tse [43]. The GSVD has also been used in earlier works on other vector channels. For example, Ekrem and Ulukus [14] introduced a GSVD based coding scheme for the broadcast channel with common and private messages.

3.1.3 Notation

Throughout, the i-th transmitter/receiver is denoted as Txi/Rxi for $i \in \{1,2,3\}$, and its message, transmit symbol, rate and degrees of freedom (GDoF) are denoted as $\mathtt{M}_i$, X_i, R_i and d_i, respectively. The number of antenna at Txi and Rxi is denoted as M_i and N_i, respectively.

We use capital letters to denote random vectors such as X_i. The underlying alphabets are denoted by $\mathcal{X}_i$, and specific values by x_i. We use the usual short hand notation for (conditional) probability distributions where the lower case arguments also denote the random variables whose (conditional) distribution is being considered. For example, $p(y_i|x_i)$ denotes $p_{\mathrm{Y}_i|\mathrm{X}_i}(y_i|x_i)$.

We use $\mathbb{C}$ to denote the set of complex numbers and $Z \sim \mathcal{CN}(0, I_N)$ to denote a N-dimensional random vector Z that obeys the complex circularly symmetric Gaussian distribution with zero mean and covariance matrix I_N (the $N \times N$ identity matrix). The note either $\det(\cdot)$ or $|\cdot|$ is used to represent the determinant of a matrix. The number of antennas at Tx$i.j$ and Rxi are denoted as $M_{i.j}$ and N_i. The Frobenius norm of a matrix H is denoted by $\|H\|_{\mathrm{F}}^2$, i.e., $\|H\|_{\mathrm{F}}^2 = \mathrm{Tr}(HH^\dagger)$, where $\mathrm{Tr}(\cdot)$ returns the trace of a given matrix. We use $\mathbb{U}^{N\times N}$ to represent the set of $N \times N$ unitary matrices. The k-th row and column of the matrix H are denoted as $H^{(k)}$ and $H^{[k]}$ respectively. A sub-matrix obtained by taking the rows k_1 to k_2 of the matrix H is written as $H^{(k_1:k_2)}$. A sub-matrix obtained by taking the columns k_1 to k_2 of the matrix H is written as $H^{[k_1:k_2]}$. The linear span of matrix H is denoted as $\langle H \rangle$. For two matrices A and B, if $(A - B)$ is positive definite (p.d.) or positive semi-definite (p.s.d), we write the relationship as $A \succ B$ or $A \succeq B$, respectively. We use $o(1)$ to represent a term which approaches zero asymptotically and $\mathcal{O}(1)$ to represent a term which is bounded above by some constant. The function $(M)^+$ returns the maximum value of M and 0, i.e., $(M)^+ = \max\{M, 0\}$. The minimum and maximum singular value of a matrix H are denoted as $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$, respectively. We refer rectangular diagonal matrix as any matrix whose nonzero entries only appear on one particular diagonal (not necessarily the main diagonal). The diagonal values of a rectangular diagonal matrix are the entries on that diagonal which contains nonzero values. The minimum and maximum nonzero diagonal values of a rectangular diagonal matrix Σ are denoted as $\sigma_{\min}(\Sigma)$

and $\sigma_{\max}(\Sigma)$, respectively.

The rest of the chapter is organized as follows. Section 3.2 defines the DM and MIMO one-to-three IC models, and discuss the channel structure of the MIMO one-to-three IC. Section 3.4 presents the inner and outer bounds for MIMO one-to-three IC, with an inner bound for DM one-to-three IC as a byproduct. GDoF region will be characterized in Section 3.5. Section 3.6 concludes the chapter. Some proofs are relegated to Appendices.

3.2 Channel Models

In this section, we first introduce the general discrete-memoryless one-to-three interference channel and the MIMO one-to-three interference channel. Then we explain the channel structure of a motivating example, in particular the structure of the two interference signals. Finally, the channel structure of general MIMO one-to-three interference channel is demonstrated using the generalized singular value decomposition of the two cross channel matrices.

3.2.1 Discrete-memoryless One-to-three Interference Channel (DM one-to-three IC)

In a DM one-to-three IC, as shown in Fig. 3.4, there are three direct point-to-point links, namely Tx1→Rx1, Tx2→Rx2 and Tx3→Rx3. Two interference links exist from Tx1 to Rx2 and Rx3, respectively, as shown in Fig. 3.4. The DM one-to-three IC is defined in Definition 3.1.

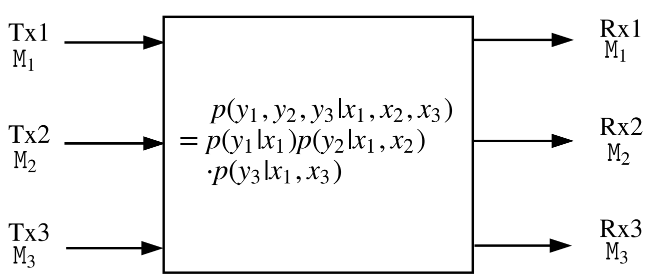

Figure 3.4: DM one-to-three IC

Definition 3.1. A discrete memoryless one-to-three interference channel is a three-transmitter and three-receiver network $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3,\ p(y_1, y_2, y_3|x_1, x_2, x_3),\ \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3)$ with transition probability satisfying

$$\begin{aligned} &p(y_1^n, y_2^n, y_3^n|x_1^n, x_2^n, x_3^n) \\ &\quad = \prod_{t=1}^{n} \left(p(y_{1t}|x_{1t})p(y_{2t}|x_{1t}, x_{2t})p(y_{3t}|x_{1t,}x_{3t})\right). \end{aligned} \tag{3.1}$$

The input and output symbols X_i and Y_i are taken from discrete alphabets $\mathcal{X}_i$ and $\mathcal{Y}_i$, respectively, where $i \in \{1, 2, 3\}$. Message $\mathtt{M}_i$ is generated from set $\mathcal{M}_i$ uniformly at random, and encoded at transmitter Txi. Receiver Rxi decodes $\mathtt{M}_i$ as $\hat{\mathtt{M}}_i$.

Given the channel as defined in Definition 3.1, a $(n, R_1, R_2, R_3, P_e^{(n)})$ coding scheme for a DM one-to-three IC consists of

- $\mathtt{M}_i$, the message to transmit at Txi, assumed to be uniformly distributed over $\mathcal{M}_i \in \{1, \cdots, 2^{nR_i}\}$, for each $i \in \{1, 2, 3\}$;
- Encoding functions $f_i(\cdot)$ such that

$$\mathtt{f}_i(\cdot): \qquad \mathcal{M}_i \longmapsto \mathcal{X}_i^n,\ m_i \longmapsto x_i^n(m_i).$$

- Decoding functions $g_i(\cdot)$ such that

$$\mathtt{g}_i(\cdot): \ \mathcal{Y}_i^n \longmapsto \mathcal{M}_i,\ y_i^n \longmapsto \hat{m}_i(y_i^n).$$

The probability of error $P_e^{(n)}$ is defined to be

$$P_e^{(n)} = P\left\{\mathtt{M}_1 \neq \hat{\mathtt{M}}_1, \mathtt{M}_2 \neq \hat{\mathtt{M}}_2 \text{ or } \mathtt{M}_3 \neq \hat{\mathtt{M}}_3\right\}.$$

A rate-tuple (R_1, R_2, R_3) is said to be achievable if there exists a sequence of $(n, R_1, R_2, R_3, P_e^{(n)})$ coding schemes for which $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region of the DM one-to-three IC is the closure of all achievable rate tuples of this channel, denoted as $\mathcal{C}^{\mathrm{DM}}$.

3.2.2 Gaussian MIMO One-to-three Interference Channel (MIMO One-to-three IC)

A $(M_1, N_1, M_2, N_2, M_3, N_3)$ Gaussian MIMO (multiple-input-multiple-output) one-to-three IC, as shown in Fig. 3.2.2, has M_i antennas at Txi and N_i antennas at Rxi for each $i \in \{1, 2, 3\}$. Assuming the channel transfer matrices and path attenuations (also referred to as channel gains) are time-invariant during the transmission, the input-output relationship of this channel is described by

$$Y_1 = h_{11}H_{11}X_1 + Z_1 \tag{3.2}$$

$$Y_2 = h_{12}H_{12}X_1 + h_{22}H_{22}X_2 + Z_2 \tag{3.3}$$

$$Y_3 = h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3, \tag{3.4}$$

where $X_i \in \mathbb{C}^{M_i \times 1}$ and $Y_i \in \mathbb{C}^{N_i \times 1}$ are complex input and output vectors, and$H_{ij} \in \mathbb{C}^{N_j \times M_i}$ is the channel transfer matrix from Txi to Rxj whose Frobenius norm satisfies $\|H_{ij}\|_F^2 = 1$. We assume the entries of the transfer matrices H_{ij} are drawn i.i.d. from a continuous and unitarily invariant distribution [45], i.e., $UH_{ij}V$ is identically distributed to H_{ij} for any $U \in \mathbb{U}^{N_i \times N_j}$ and $V \in \mathbb{U}^{M_i \times M_i}$, so that H_{ij} has full rank with probability one (w.p.1). The path attenuation h_{ij} from Txi to Rxj is a complex number. The Gaussian noise Z_i are i.i.d. $\mathcal{CN}(\mathbf{0}, I_{N_i})$ across i. Let Cov$[x_{it}]$ be the covariance of the t-th symbol of the transmitted codeword $x_i^n \in \mathcal{X}_i^n$ at Txi. The codeword x_i^n should meet the average per-codeword power constraint,

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{Tr}(x_{it}x_{it}^\dagger) \leq P_i. \tag{3.5}$$

The SNR and INR at receiver Rxi are defined to be

$$\mathsf{SNR}_{ii} = P_i|h_{ii}|^2 \triangleq \rho^{\alpha_{ii}},\ i \in \{1,2,3\} \tag{3.6}$$

$$\mathsf{INR}_{1i} = P_1|h_{1i}|^2 \triangleq \rho^{\alpha_{1i}},\ i \in \{2,3\} \tag{3.7}$$

where ρ is the nominal SNR based on which the direct channel SNRs and the two cross-channel INRs are defined. The distinct SNR and INR exponents allow us to express the disparities in power levels observed across the direct and cross links as multiplicative terms associated with the nominal SNR in the dB scale. Without loss of generality, we assume $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$ and denote the capacity region of the MIMO one-to-three IC as $\mathcal{C}$.

3.2.3 A Motiving Example

Understanding the relationship of the two vector interference signals $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$ is crucial to designing a coding scheme which adapts the channel parameters. In this subsection, we investigate

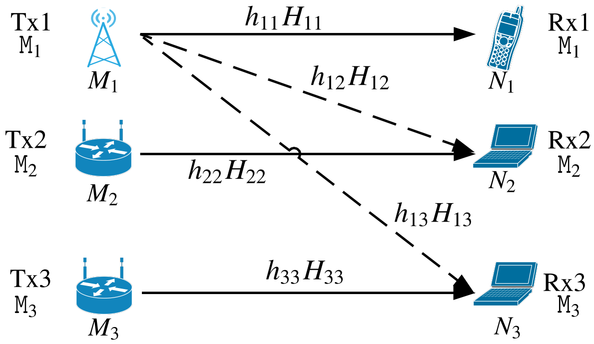

Figure 3.5: The MIMO one-to-three IC

a simple and intuitive MIMO one-to-three IC and demonstrate that the interferences and noise received at Rx2 and Rx3 can be written as disjointed channel side informations which carry the common and exclusive parts of $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$.

Consider a $(3,3,2,2,2,2)$ MIMO one-to-three IC with cross channel matrices H_{12} and H_{13} expressed as

$$H_{12} = U_{12}\Sigma_{12}V_{12}^{\dagger} = U_{12}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \tag{3.8}$$

$$H_{13} = U_{13}\Sigma_{13}V_{13}^{\dagger} = U_{13}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}. \tag{3.9}$$

Other channel parameters can be arbitrarily chosen. Remarkably, the equation (3.9) is not a singular value decomposition of H_{13} since the nonzero values are not on the main diagonal of the matrix Σ_{13}.

The earlier work [4, Section VI-II] pointed out that multi-level superposition coding is sufficient to achieve a constant-gap-to-capacity region (and hence is GDoF optimal) for the Gaussian scalar K-user one-to-many IC. More specifically, the coding scheme therein assumes the K interfered receivers are ordered by increasing interference. The interfering transmitter splits the transmitted signal into multiple partitions so that every interfering receiver decodes the received partitions above its noise floor. In the case of MIMO one-to-three IC, we cannot sort interferences simply because the two interferences are vectors. Thus, we cannot directly apply the coding scheme in [4, Section VI-II] for the MIMO one-to-many IC. Rather, we need a deeper understanding of the structure of this channel, especially the structure of the two interferences $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$ in order to design a coding scheme that adapts the channel parameters.

With the chosen channel matrices (3.8) and (3.9), the two interference signals received by Rx2 and Rx3 can then be written as

$$h_{12}H_{12}X_1 = h_{12}U_{12}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \\ X_1^{(3)} \end{pmatrix}$$

and

$$h_{13}H_{13}X_1 = h_{13}U_{13}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \\ X_1^{(3)} \end{pmatrix}.$$

Observing the positions of nonzero values in the matrices Σ_{12} and Σ_{13}, we readily see that the first component of the X_1 (denoted as $X_1^{(1)}$) is hearable at Rx2 but not at Rx3, the third component of X_1 ($X_1^{(3)}$) is hearable at Rx3 but not at Rx2, and the second component of X_1 ($X_1^{(2)}$) can be heard by both Rx2 and Rx3, but Rx2 hears a stronger version of $X_1^{(2)}$. Given the common and exclusive parts seen between the interferences $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$, we then construct channel side informations S_{123}, S_{12} and S_{13} as

$$S_{123} = h_{13}U_{13}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}X_1 + U_{13}\begin{pmatrix} \left(U_{13}^{-1}\right)^{(1)} Z_3 \\ 0 \end{pmatrix}$$

$$S_{12} = h_{12}U_{12}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}X_1 + Z_2$$

$$S_{13} = h_{13}U_{13}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}X_1 + U_{13}\begin{pmatrix} 0 \\ \left(U_{13}^{-1}\right)^{(2)} Z_3 \end{pmatrix}.$$

Now the channel input-output relationship can be rewritten as

$$Y_1 = h_{11}H_{11}X_1 + Z_1 \tag{3.10}$$

$$Y_2 = (S_{12} - S_{123}) + S_{123} + h_{22}H_{22}X_2 \tag{3.11}$$

$$Y_3 = S_{123} + S_{13} + h_{33}H_{33}X_3. \tag{3.12}$$

The channel side information S_{123} contains the interference contributed by $X_1^{(2)}$ with the lower channel gain h_{13}, it is the common part of the interference which could be received by both Rx2 and Rx3. The difference of $S_{12} - S_{123}$ is the interference that is hearable by Rx2 but not Rx3, and it consists of two parts: the contribution by $X_1^{(1)}$ with channel gain h_{12} and also the contribution of $X_1^{(2)}$ amplified by the higher channel gain h_{12}. The channel side information S_{13} contains the interference contributed by $X_1^{(3)}$ with channel gain h_{13} that is hearable by Rx3 but not Rx2. With the channel structure reflected by the channel side informations S_{123}, S_{12} and S_{13} by (3.10)-(3.12), a coding scheme arises naturally. The message $\mathtt{M}_1$ is split into four parts as $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$. The first three sub-messages are carried by S_{123}, S_{12} and S_{13} are decoded by Rx1-Rx3, Rx1 and Rx2, and Rx1 and Rx3, respectively. The sub-message $\mathtt{M}_{1p}$ is the private sub-message to be decoded by Rx1 only. We shall complete the discussion of the coding scheme for this channel in Section 3.4.1.

In this example, it is critical to find out the common and exclusive parts of the two interference signals and formulate the disjointed channel side informations in determining the coding scheme. Since we have $V_{12} = V_{13} = I_3$ here, the common and exclusive parts of $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$ can be directly obtained by observing the positions of the nonzero entries in Σ_{12} and Σ_{13}. For arbitrary channel matrices H_{12} and H_{13}, a more sophisticated matrix decomposition technique is needed to reveal the relationship between the two interferences, which will be introduced in the next subsection.

3.2.4 Channel Structure of the MIMO One-to-three IC

Next, we demonstrate the channel structure of the MIMO one-to-three IC in general. Since channel matrices H_{12} and H_{13} have full rank w.p.1, we have

$$r_{12} \triangleq \text{rank}(H_{12}) = \min\{M_1, N_2\} \tag{3.13}$$

$$r_{13} \triangleq \text{rank}(H_{13}) = \min\{M_1, N_3\} \tag{3.14}$$

$$r \triangleq \text{rank}\begin{pmatrix} H_{12} \\ H_{13} \end{pmatrix} = \min\{M_1, N_2 + N_3\}. \tag{3.15}$$

When we linearly spanned the row vectors of H_{12} and H_{13}, the intersection of the two resulting spaces have dimension

$$r_{123} \triangleq r_{12} + r_{13} - r. \tag{3.16}$$

We jointly decompose the two channel matrices H_{12} and H_{13} via the generalized singular value decomposition (GSVD) [35], which is

$$H_{12} = U_{12}\Sigma_{12} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix}^{\dagger} U^{\dagger} \triangleq U_{12}\Sigma_{12}V^{\dagger} \tag{3.17}$$

$$H_{13} = U_{13}\Sigma_{13} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix}^{\dagger} U^{\dagger} \triangleq U_{13}\Sigma_{13}V^{\dagger}. \tag{3.18}$$

$U_{1i} \in \mathbb{U}^{N_i\times N_i}$ and $U \in \mathbb{U}^{M_1\times M_1}$ are unitary matrices. $\Sigma_{1i} \in \mathbb{R}^{N_i\times r}$ is a real and rectangular diagonal matrix. $V_r \in \mathbb{C}^{r\times r}$ is a non-singular upper triangular matrix and $V \triangleq U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix} \in \mathbb{C}^{M_1\times r}$. Matrices Σ_{12} and Σ_{13} have the following structure

$$\Sigma_{12} = \begin{array}{c} \\ r-r_{13} \\ r_{123} \\ N_2-r_{12} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array} \tag{3.19}$$

$$\Sigma_{13} = \begin{array}{c} \\ r_{123} \\ r-r_{12} \\ N_3-r_{13} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{pmatrix} \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}, \tag{3.20}$$

where C and S are both non-negative real diagonal matrices satisfying $C^2 + S^2 = I$. The acronym N.E. means "never exists". Note $N_2 - r_{12}$ and $r - r_{12}$ cannot be simultaneously positive according to (3.13) and

(3.15), and the matrix Σ_{12} is in form of either

$$\Sigma_{12} = \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r-r_{13} \\ r_{123} \end{array} & \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \end{pmatrix} \end{array}$$

or

$$\Sigma_{12} = \begin{array}{c c} & \begin{array}{cc} r-r_{13} & r_{123} \end{array} \\ \begin{array}{c} r-r_{13} \\ r_{123} \\ N_2-r_{12} \end{array} & \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & C \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}.$$

These two different forms of the matrix Σ_{12} can be unified as the one given by (3.19), where the acronym N.E. comes from the fact that Σ_{12} can only be a 2×3 or 3×2 rectangular diagonal matrices. Similarly, $N_3 - r_{13}$ and $r - r_{13}$ cannot be simultaneously positive because of (3.14) and (3.15), and Σ_{13} also has two different forms which can be unified as (3.20).

Remark 3.1. GSVD decomposes H_{12} and H_{13} jointly and ensures both (3.17) and (3.18) have the same right hand side matrix V but at the cost of permitting V to be non-unitary and rank deficient. GSVD renders similar decomposition forms for H_{12} and H_{13} as single value decomposition (SVD); however, the rectangular diagonal matrix Σ_{1i} is usually not the singular value matrix of H_{1i}. Also, the matrix Σ_{1i} is of size $N_i \times r$ instead of $N_i \times M_1$ as in SVD. Since V is generally not unitary, the column vectors of V do not form a orthonormal basis of the transmit signal space at Tx1.

Remark 3.2. Note we have assumed INR_{12} and INR_{13} are sufficiently large so that we can disregard the gains contributed by C and S which is justifiable for analysis up to GDoF accuracy.

Remark 3.3. Through GSVD, the reception of the two interferences sent to Rx2 and Rx3, i.e. $h_{12}H_{12}X_1$ and $h_{13}H_{13}X_1$, respectively, can be understood as follows. First, the input vector signal $X_1 \in \mathbb{C}^{M_1}$ is transformed into a r dimensional column vector $V^\dagger X_1$, leaving the remaining $(M_1 - r)^+$ dimensions unheard by both Rx2 and Rx3. When $r < M_1$, the matrix V is row rank deficient which reflects the case when the total number of receive antennas at Rx2 and Rx3 is less the number of transmit antenna at Tx1. Next, the matrix Σ_{1i} determines which components in $V^\dagger X_1$ are transferred to Rxi. The first $r - r_{13}$ transmit directions of $V^\dagger X_1$

are heard by Rx2 but not by Rx3 because Σ_{12} has identity block matrix in the upper left, whereas Σ_{13} has all-zero block matrix in the upper left. For a similar reason, the last $r - r_{12}$ transmit directions of $V^\dagger X_1$ are received at Rx3 but not Rx2. The middle r_{123} transmit directions of $V^\dagger X_1$ are received by both Rx2 and Rx3, but with different interference strengths (INR_{12} and INR_{13}). Finally, the left hand side unitary matrix U_{1i} produces N_i linear combinations of the r_{1i} components in $\Sigma_{1i} V^\dagger X_1$ as the interference signals, which are the N_i signal received by the N_i antennas at Rxi.

Now that the GSVD provides a joint decomposition with the same right hand matrix V, the common and exclusive parts of the two interference signals can be determined according to matrices Σ_{1i} and channel gain h_{1i}. Define two matrices I_{12} and Λ_{12} to be

$$I_{12} \triangleq \begin{array}{c} \\ r - r_{13} \\ r_{123} \\ N_2 - r_{12} \end{array} \begin{array}{c} \begin{array}{ccc} r - r_{13} & r_{123} & r - r_{12} \end{array} \\ \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array} \tag{3.21}$$

and

$$\Lambda_{12} \triangleq \begin{array}{c} \\ r - r_{13} \\ r_{123} \\ N_2 - r_{12} \end{array} \begin{array}{c} \begin{array}{ccc} r - r_{13} & r_{123} & r - r_{12} \end{array} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array}, \tag{3.22}$$

respectively, and the matrix Σ_{12} can be written as

$$\Sigma_{12} = I_{12} + \Lambda_{12}.$$

Similarly we define

$$I_{13} \triangleq \begin{array}{c} \\ r_{123} \\ r - r_{12} \\ N_3 - r_{13} \end{array} \begin{array}{c} \begin{array}{ccc} r - r_{13} & r_{123} & r - r_{12} \end{array} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array} \tag{3.23}$$

and

$$\Lambda_{13} \triangleq \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r_{123} \\ r-r_{12} \\ N_3-r_{13} \end{array} & \begin{pmatrix} \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}, \tag{3.24}$$

so the matrix Σ_{13} can be written as

$$\Sigma_{13} = I_{13} + \Lambda_{13}.$$

Let the matrices G_{12} and J_{12} be

$$G_{12} \triangleq U_{12}\Lambda_{12}V^{\dagger} \text{ and } J_{12} \triangleq U_{12}I_{12}V^{\dagger}.$$

Hence, $H_{12} = G_{12} + J_{12}$. Let the matrices G_{13} and J_{13} be

$$G_{13} \triangleq U_{13}\Lambda_{13}V^{\dagger} \text{ and } J_{13} \triangleq U_{13}I_{13}V^{\dagger}.$$

Hence, $H_{13} = G_{13} + J_{13}$.

The common part of the two interferences should be the r_{123} elements in the "middle" of column vector $V^{\dagger}X_1$ which are both transferred to Rx2 and Rx3. In light of the interference power disparity to Rx2 and Rx3, we define S_{123} to be the weaker version of the middle r_{123} transmit directions of $V^{\dagger}X_1$, and it could be either $h_{13}G_{12}X_1$ or $h_{13}G_{13}X_1$. Both these two signals contain the same information about X_1, albeit received as different signals. The power difference in $h_{13}G_{12}X_1$ and $h_{13}G_{13}X_1$ (caused by different diagonal values in C and S) is ignorable with sufficiently large INR_{12} and INR_{13}. $h_{12}H_{12}X_1 - h_{13}G_{12}X_1$ is the exclusive part of the interference to Rx2. It consists of the interference sent along the first $r - r_{13}$ directions which are hearable at Rx2 but not at Rx3, and the interference sent along the middle r_{123} directions with extra power gain. $h_{13}J_{13}X_3$ is the exclusive part of the interference to Rx3 which is the interference sent along the bottom $r - r_{12}$ directions which are hearable at Rx3 but not at Rx2.

So far we have figured out the structure of the two interference signals received at Rx2 and Rx3. The structure of the channel can thus be expressed in terms of three channel side information S_{123}, S_{12} and S_{123}

which are defined as follows.

$$S_{123} \triangleq h_{13}G_{13}X_1 + U_{13}\begin{pmatrix} U_{13}^{-1(1:r_{123})}Z_3 \\ \mathbf{0}_{(N_3-r_{123})\times 1}\end{pmatrix} \tag{3.25}$$

$$S_{12} \triangleq h_{12}H_{12}X_1 + Z_2 \tag{3.26}$$

$$S_{13} \triangleq h_{13}J_{13}X_1 + U_{13}\begin{pmatrix} \mathbf{0}_{r_{123}\times 1} \\ U_{13}^{-1(r_{123}+1:N_3)}Z_3\end{pmatrix} \tag{3.27}$$

We then can rewrite the channel output Y_2 and Y_3 in terms of their intended signal and channel side information as

$$Y_2 = S_{12} + h_{22}H_{22}X_2 \tag{3.28}$$

$$Y_3 = S_{123} + S_{13} + h_{33}H_{33}X_3 = U_{13}\begin{pmatrix} U_{13}^{-1(1:r_{123})}S_{123} \\ U_{13}^{-1(r_{123}+1:N_3)}S_{13}\end{pmatrix} + h_{33}H_{33}X_3. \tag{3.29}$$

The side informations S_{123}, S_{12} and S_{123} carry the parts of the interferences that is hearable by both Rx2 and Rx3, Rx2 but not Rx3, and Rx3 but not Rx2. Each side information not only contains certain part of the interference signal, but also the associated Gaussian noise elements along the corresponding directions. There is no S_{123} explicitly in (3.28) because $S_{12} = h_{12}U_{12}I_{12}V^{\dagger}X_1 + h_{12}U_{12}\Lambda_{12}V^{\dagger}X_1 + Z_2$ already contains a scaled and linearly transformed version of the interference signal in S_{123} in the term $h_{12}U_{12}\Lambda_{12}V^{\dagger}X_1$. As will be seen in Section 3.4, the GDoF optimal coding scheme at Tx1 incorporates signal direction alignment to utilize the exclusive transmit directions from Tx1 to Rx2 and Rx3 respectively, as well as signal level alignment to adapt the disparity of the interference strength along the common transmit directions from Tx1 to Rx2-Rx3.

Remark 3.4. Note the bottom $(N_3 - r_{123})$ rows of S_{123} and the upper r_{123} rows of S_{13} are all zeros. Hence the side information S_{123} and S_{13} to Rx3 are disjointed in signal directions, and we have

$$h(S_{123} + S_{13}) = h(S_{123}^{(1:r_{123})}, S_{13}^{(r_{123}+1:N_3)}) = h(S_{123}, S_{13}). \tag{3.30}$$

This setup plays an important role in deriving the outer bound in Section 3.4.3 when $h(S_{123}^n + S_{13}^n)$ needs to be processed, as it is easier to process joint differential entropy $h(S_{123}^n, S_{13}^n)$ than to process the entropy of the sum $h(S_{123}^n + S_{13}^n)$.

The relationship between $V_r V_r^\dagger$ and the two scaled identity matrices $\lambda^2_{\min}(V_r)I_r$ and $\lambda^2_{\max}(V_r)I_r$ will be frequently used in the rest of the chapter. We present it in Fact 3.1.

Fact 3.1. *Let V_r be a $r \times r$ full rank square matrix. The following relationship holds between the matrices $V_r V_r^\dagger$, $\lambda^2_{\min}(V_r)I_r$ and $\lambda^2_{\max}(V_r)I_r$*

$$\lambda^2_{\min}(V_r)I_r \preceq V_r V_r^\dagger \preceq \lambda^2_{\max}(V_r)I_r, \tag{3.31}$$

which is equivalent to

$$\frac{V_r V_r^\dagger}{\lambda^2_{\max}(V_r)} \preceq I_r \preceq \frac{V_r V_r^\dagger}{\lambda^2_{\min}(V_r)} \tag{3.32}$$

and

$$\lambda^{-2}_{\max}(V_r)I_r \preceq V_r^{-1} V_r^{\dagger -1} \preceq \lambda^{-2}_{\min}(V_r)I_r. \tag{3.33}$$

3.3 Multi-level Superposition Coding and Inner Bound for DM One-to-three IC

The structure of the channel (c.f. Definition 3.1) suggests a natural coding scheme for DM one-to-three IC: superposition coding at Tx1 and independent single user random coding at Tx2 and Tx3. Since Rx2 and Rx3 receive different versions of the interference from Tx1, the coding scheme should adapt this difference. Therefore, we split the message $\mathtt{M}_1$ at Tx1 into four parts ($\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$) and perform three level superposition coding to let those four messages be decodable at Rx1-Rx3, Rx1 and Rx2, Rx1 and Rx3 and Rx1 only, respectively. The set of coding distributions is given in Definition 3.2.

Definition 3.2. Let $\mathcal{P}_{\text{in}}$ be the set of distributions of joint random variables $(Q, W_{123}, W_{12}, W_{13}, X_1, X_2, X_3)$ that can be factored as

$$p(q, w_{123}, w_{12}, w_{13}, x_1, x_2, x_3) = p(q)p(w_{123}|q)p(w_{12}|w_{123})p(w_{13}|w_{123})p(x_1|w_{123}, w_{12}, w_{13}) \prod_{i \in \{2,3\}} p(x_i|q). \tag{3.34}$$

An inner bound can be obtained for any fixed coding distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$ through a detailed joint typicality analysis. We state it in Theorem 3.1.

Theorem 3.1. *For a DM one-to-three IC and some fixed distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$, the following region $\mathcal{R}_{\text{in}}^{\text{DM}}(P_{\text{in}})$ given by*

$$\mathcal{R}_{\text{in}}^{\text{DM}}(P_{\text{in}}) \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq I(X_1; Y_1|Q) \tag{3.35}$$

$$R_2 \leq I(X_2; Y_2|W_{123}, W_{12}, Q) \tag{3.36}$$

$$R_3 \leq I(X_3; Y_3|W_{123}, W_{13}, Q) \tag{3.37}$$

$$R_1 + R_2 \leq I(X_1; Y_1|W_{123}, W_{12}, Q) + I(X_2, W_{123}, W_{12}; Y_2|Q) \tag{3.38}$$

$$R_1 + R_3 \leq I(X_1; Y_1|W_{123}, W_{13}, Q) + I(X_3, W_{123}, W_{13}; Y_3|Q) \tag{3.39}$$

$$R_1 + R_2 + R_3 \leq I(X_1; Y_1|W_{123}, W_{12}, W_{13}, Q) + I(X_2, W_{12}; Y_2|W_{123}, Q) + I(X_3, W_{123}, W_{13}; Y_3|Q) \tag{3.40}$$

$$R_1 + R_2 + R_3 \leq I(X_1; Y_1|W_{123}, W_{12}, W_{13}, Q) + I(X_2, W_{123}, W_{12}; Y_2|Q) + I(X_3, W_{13}; Y_3|W_{123}, Q) \tag{3.41}$$

$$\begin{aligned} 2R_1 + R_2 + R_3 \leq\ & I(X_1; Y_1|W_{123}, W_{12}, W_{13}, Q) + I(X_1; Y_1|W_{123}, Q) + I(X_2, W_{123}, W_{12}; Y_2|Q) \\ & + I(X_3, W_{123}, W_{13}; Y_3|Q)\big\} \end{aligned} \tag{3.42}$$

is achievable, i.e., $\mathcal{R}_{\text{in}}^{\text{DM}} \subseteq \mathcal{C}^{\text{DM}}$.

Proof Outline. We outline the proof here and relegate the full proof to Appendix A.1. As previously stated, Tx1 performs three level of superposition coding, and Tx2 and Tx3 perform independent single user random coding. More specifically, Tx1 splits a message m_1 into four parts: m_{123}, m_{12}, m_{13} and m_{1p}. The sub-message m_{123}, which needs to be decoded by Rx1-Rx3, is first encoded to the first level codeword $w_{123}^n(m_{123})$. Then the multicast sub-message m_{1i} is encoded to $w_{1i}^n(m_{1i}, w_{123}^n(m_{123}))$, which needs to be decoded by Rx1 and Rxi for $i \in \{2, 3\}$. This is the second level superposition coding. Finally, based on m_{1p}, which is the private message to be decoded by Rx1, the entire message is encoded to the codeword $x_1(m_{1p}, w_{12}^n(m_{12}, w_{123}^n(m_{123})), w_{13}^n(m_{13}, w_{123}^n(m_{123})))$ for transmission. Txi, $i = \{2, 3\}$, sends information m_i via some codeword $x_i^n(m_i)$ using a single-user random codebook and Rxi decodes the intended message m_i. Fourier-Motzkin elimination is used to eliminate the three rate variables associated with the auxiliary random variables W_{123}, W_{12} and W_{13} to obtain the achievable region. □

3.4 Bounds on the Capacity Region of the MIMO One-to-three IC

We present single region inner and outer bounds for the MIMO one-to-three IC within quantifiable (and channel SNR/INR independent) gap in this section. In section 3.4.1, we demonstrate an intuitive coding scheme for the motivating MIMO one-to-three IC introduced in Section 3.2.3. In Section 3.4.2, we present an **explicit** additive superposition coding scheme for the general MIMO one-to-three IC with Gaussian codebooks and specified covariance matrices. We obtain a single region inner bound which has the form of a single polytope. In Section 3.4.3, we characterize a **single** region outer bound by genie aided argument. In Section 3.4.4, the gap between the inner and outer bounds is then quantified and shown to be dependent only on the entries of the channel matrices H_{12} and H_{13}, leading to the characterization of the fundamental GDoF region.

3.4.1 Coding Scheme for the Motivating Example

We explicitly specify one coding scheme for the motivating example given in Section 3.2.3. Following the discussion in Section 3.3, we split the message at Tx1 into four parts, namely $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$. Also, according to the coding scheme for the DM one-to-three IC, these four sub-messages are encoded to W_{123}^n, W_{12}^n, W_{13}^n and W_{1p}^n respectively, and Tx2 and Tx3 use single user random coding. In this subsection, we demonstrate one coding scheme and therefore yields one coding distribution for W_{123}, W_{12}, W_{13} and W_{1p}, as well as X_2 and X_3. This coding scheme adapts the channel structure we have explored in Section 3.2.3. The basic idea is given below.

Recall the sub-message $\mathtt{M}_{123}$ is the common message for all three receivers, so it should be sent at the highest possible power level along the signal directions that all three receivers could hear. Hence, we encode $\mathtt{M}_{123}$ into the codeword W_{123}^n and transmit W_{123}^n along the transmit direction $I_3^{(2)}$ at power level ρ^0. The signal sub-message $\mathtt{M}_{12}$ is only intended for Rx1 and Rx2. Therefore, we encode it to W_{12}^n and transmit W_{12}^n along the transmit direction $I_3^{(1)}$ at power level ρ^0 (hence is not heard by Rx3) and also along the transmit direction $I_3^{(2)}$ at power level $\rho^{-\alpha_{13}}$ (hence is received under the noise floor at Rx3). The sub-message $\mathtt{M}_{13}$ should be received by Rx1 and Rx3, so we encode it into W_{13}^n and transmit it along the direction $I_3^{(3)}$ at power level ρ^0 (hence is not heard by Rx2). In case SNR_{11} is greater than both INR_{12} and INR_{13}, Rx1 can decode a private message $\mathtt{M}_{1p}$. We encode $\mathtt{M}_{1p}$ to W_{1p}^n and transmit W_{1p}^n on all three transmit directions at power level $\rho^{-\alpha_{12}}$ (hence not heard by either Rx2 or Rx3).

With the basic idea we have developed, we let the transmitted signal X_1 be a sum of four independent Gaussian sub-signals

$$X_1 = \sqrt{P_1}(W_{123} + W_{12} + W_{13} + W_{1p}),$$

where $W_{123} \sim \mathcal{CN}(\mathbf{0}, Q_{123})$, $W_{12} \sim \mathcal{CN}(\mathbf{0}, Q_{12})$, $W_{13} \sim \mathcal{CN}(\mathbf{0}, Q_{1p})$ and $W_{1p} \sim \mathcal{CN}(\mathbf{0}, Q_{1p})$ are the auxiliary Gaussian random variables to encode $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$, respectively. The entire signal is transmitted with full power, and the covariance matrix of X_1 is chosen to be the identity matrix scaled by the reciprocity of the number of transmit antennas at Tx1, i.e.,

$$Q_1 \triangleq \mathrm{Cov}[X_1] = \frac{P_1}{3} I_3. \tag{3.43}$$

The codeword W_{1p}^n should be sent along all transmit directions and should be received under the noise floor at both Rx2 and Rx3. Therefore, we choose the covariance of the W_{1p} as

$$Q_{1p} = \frac{1}{3}\left(I_3 + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} + \rho^{\alpha_{13}}\Sigma_{13}^\dagger\Sigma_{13}\right)^{-1} = \frac{1}{3}\begin{pmatrix} \frac{1}{1+\rho^{\alpha_{12}}} & & \\ & \frac{1}{1+\rho^{\alpha_{12}}+\rho^{\alpha_{13}}} & \\ & & \frac{1}{1+\rho^{\alpha_{13}}} \end{pmatrix}. \tag{3.44}$$

It ensure that the contributions of W_{1p} at Rx2 and Rx3 have covariances

$$\rho^{\alpha_{12}} H_{12} Q_{1p} H_{12}^\dagger \preceq I_{N_2} \text{ and } \rho^{\alpha_{13}} H_{13} Q_{1p} H_{13}^\dagger \preceq I_{N_3}.$$

Therefore, with this covariance matrix Q_{1p}, W_{1p} indeed arrives at the unintended receivers under the noise floor. The choice of Q_{1p} can be seen as an extension of the selection of the covariance matrix for private sub-messages for the two-user MIMO IC by [27]. The difference here is that Q_{1p} should be chosen so that the interference W_{1p} arrives under the noise floor at both Rx2 and Rx3.

Next, we determine the covariance of W_{12}. W_{12} carries the sub-message $\mathtt{M}_{12}$ to be decoded by Rx1 and Rx2, but not Rx3, so a direct idea of Q_{12} is to make sure that W_{12} arrives under the noise floor at Rx3. However, we already require W_{1p} to be received under the noise floor at Rx3; therefore, the covariance of

W_{12} should be selected to ensure that the sum signal of W_{1p} and W_{12} are received under the noise floor at Rx3, which implies

$$Q_{12} + Q_{1p} = \frac{1}{3}(I_3 + \rho^{\alpha_{13}}\Sigma_{13}^{\dagger}\Sigma_{13})^{-1} = \frac{1}{3}\begin{pmatrix} 1 & & \\ & \frac{1}{1+\rho^{\alpha_{13}}} & \\ & & \frac{1}{1+\rho^{\alpha_{13}}} \end{pmatrix}. \tag{3.45}$$

Subtracting (3.44) from (3.45), we have

$$Q_{12} = \frac{1}{3}\begin{pmatrix} \frac{\rho^{\alpha_{12}}}{1+\rho^{\alpha_{12}}} & & \\ & \frac{\rho^{\alpha_{12}}}{(1+\rho^{\alpha_{12}}+\rho^{\alpha_{13}})(1+\rho^{\alpha_{13}})} & \\ & & 0 \end{pmatrix}.$$

This is consistent with the basic idea we have developed that the W_{12} should be sent along the first transmit direction at power level ρ^0 (as Rx3 cannot hear this direction) and along the second transmit direction at power level $\rho^{-\alpha_{13}}$ (as it arrives under the noise floor at Rx3 in this direction). W_{12} does not transmit on the third transmit direction at all.

The covariance of W_{13} can be determined in a similar fashion. We let the sum covariance $Q_{13} + Q_{1p}$ satisfy

$$Q_{13} + Q_{1p} = \frac{1}{3}(I_3 + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12} + \rho^{\alpha_{13}}\Lambda_{13}^{\dagger}\Lambda_{13})^{-1} = \frac{1}{3}\begin{pmatrix} \frac{1}{1+\rho^{\alpha_{12}}} & & \\ & \frac{1}{1+\rho^{\alpha_{12}}+\rho^{\alpha_{13}}} & \\ & & 1 \end{pmatrix} \tag{3.46}$$

so we have

$$Q_{13} = \frac{1}{3}\begin{pmatrix} 0 & & \\ & 0 & \\ & & \frac{\rho^{\alpha_{13}}}{1+\rho^{\alpha_{13}}} \end{pmatrix}.$$

Such choices of Q_{1p}, Q_{12} and Q_{13} automatically guarantee the sum of W_{1p}, W_{12} and W_{13} be received under the noise floor along the second transmit direction at Rx2 and Rx3, because

$$Q_{12} + Q_{13} + Q_{1p} = \frac{1}{3}\begin{pmatrix} 1 & & \\ & \frac{1}{1+\rho^{\alpha_{13}}} & \\ & & 1 \end{pmatrix}.$$

The covariance of W_{123} has to be

$$Q_{123} = Q_1 - Q_{1p} - Q_{12} - Q_{13} = \frac{1}{3}\begin{pmatrix} 0 & & \\ & \frac{\rho^{\alpha_{13}}}{1+\rho^{\alpha_{13}}} & \\ & & 0 \end{pmatrix}.$$

It is also consistent with our basic idea that W_{123} should be transmitted along the second dimension at power level ρ^0.

Tx2 and Tx3 merely transmit X_2 and X_3 using full power, with scaled identity covariance matrices, i.e.,

$$Q_2 = \frac{P_2}{2} I_2 \text{ and } Q_3 = \frac{P_3}{2} I_2.$$

Remarkably, we do not perform water-filling at these two transmitters, since we have explained in 5.3.1 that the scaled identity matrix is sufficient to achieve a rate region within constant gap to the capacity for MIMO MAC (hence also MIMO P2P channel).

Now a coding scheme has been uniquely determined, and it will be clear at the end of this section that this coding scheme is GDoF optimal.

3.4.2 Inner Bound for the MIMO One-to-three IC

We have derived an achievable region $\mathcal{R}_{\text{in}}(P_{\text{in}})$ for the DM one-to-three in Theorem 3.1 for any coding distribution P_{in}. In this subsection, we explicitly specify one coding distribution for the MIMO one-to-three IC, and then we substitute the coding distribution of this particular coding scheme in the DM one-to-three inner bound to get an explicit and single region inner bound for MIMO one-to-three IC.

Starting from the coding scheme proposed in Section 3.3, first of all we disable time sharing among the three transmitters. A non-interfering transmitter encodes its entire message using single user Gaussian codebook, and the total transmit power is uniformly and independently allocated among all transmit antennas, i.e.,

$$\text{Cov}[X_2] = \frac{P_2}{M_2} I_{M_2} \text{ and } \text{Cov}[X_3] = \frac{P_3}{M_3} I_{M_3}. \tag{3.47}$$

At Tx1, following the coding scheme developed from the motivating example, we let the transmitted signal X_1 be the direct sum of four independent Gaussian random vectors W_{123}, W_{12}, W_{13} and W_{1p}, i.e.,

$$X_1 = \sqrt{P_1}(W_{123} + W_{12} + W_{13} + W_{1p}).$$

Next, we specify the covariance matrices Q_{123}, Q_{12}, Q_{13} and Q_{1p} of W_{123}, W_{12}, W_{13} and W_{1p}, respectively. Define

$$V_p \triangleq U^{\dagger -1} \begin{pmatrix} V_r^{\dagger -1} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+} \end{pmatrix}, \tag{3.48}$$

as a linear precoding matrix. The covariance matrices Q_{123}, Q_{12}, Q_{13} and Q_{1p} satisfy the restrictions given by (3.49)-(3.53). It can be readily seen that

$$Q_{12} + Q_{1p} \succeq Q_{1p} \text{ and } Q_{12} + Q_{1p} \succeq Q_{1p},$$

which implies both Q_{12} and Q_{13} are positive semi-definite and therefore are valid covariance matrices. Note that even though there are four restrictions on Q_{1p}, Q_{12} and Q_{13}, i.e., (3.49)-(3.52), it can be shown that these covariances exist and they can be uniquely determined. In particular, it is not difficult to see that (3.52) results from adding the left and right hand sides of (3.50) and (3.51) and subtracting from that result the left and right hand sides of (3.49).

$$Q_{1p} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} + \rho^{\alpha_{13}}\Sigma_{13}^\dagger\Sigma_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \tag{3.49}$$

$$Q_{13} + Q_{1p} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \tag{3.50}$$

$$Q_{12} + Q_{1p} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}}\Sigma_{13}^\dagger\Sigma_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \tag{3.51}$$

$$Q_{12} + Q_{13} + Q_{1p} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \tag{3.52}$$

$$Q_1 \triangleq Q_{12} + Q_{13} + Q_{1p} + Q_{123} = \frac{V_pV_p^\dagger}{\mathrm{Tr}(V_pV_p^\dagger)}. \tag{3.53}$$

The scaling by $\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}$ is required to satisfy the power constraint given by (3.5). Using the inequality (3.31), the trace $\mathrm{Tr}(V_p V_p^\dagger)$ can be upper and lower bounded as

$$
\begin{aligned}
\mathrm{Tr}(V_p V_p^\dagger) = \mathrm{Tr}(V_p^\dagger V_p) &= \mathrm{Tr}\begin{pmatrix} V_r^{-1} V_r^{\dagger-1} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+} \end{pmatrix} \\
&= \mathrm{Tr}(V_r^{-1} V_r^{\dagger-1}) + (M_1 - r)^+ \\
&\geq \mathrm{Tr}\left(\frac{1}{\lambda_{\max}^2} I_r\right) + (M_1 - r)^+ \\
&= \frac{r}{\lambda_{\max}^2(V_r)} + (M_1 - r)^+ \qquad (3.54)\\
&\triangleq \zeta_{\min} \qquad (3.55)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Tr}(V_p V_p^\dagger) &\leq \frac{r}{\lambda_{\min}^2(V_r)} + (M_1 - r)^+, \qquad (3.56)\\
&\triangleq \zeta_{\max} \qquad (3.57)
\end{aligned}
$$

respectively.

The covariance matrices Q_{123}, Q_{12}, Q_{13} and Q_{1p} and the precoding matrix V_p requires the GSVD of H_{12} and H_{13} (c.f. (3.17) and (3.18)). This GSVD based coding scheme for the MIMO one-to-three IC generalizes the coding scheme for the motivating one-to-three IC in Section 3.4.1. In the motivating example, we have V as an identity matrix and each component of X_1 transmit along a particular direction; therefore, no precoding is needed. However, when V is not unitary, each transmit direction transmits certain linear combination of the components of X_1. The inverse matrix $V_r^{\dagger-1}$ on the right hand side of (3.48) compensates the non-unitarity of V_r, so the first r components in the post-precoding signal $V_r^{\dagger-1} X_1^{(1:r)}$ transmit exactly along the r transmit directions independently. On the other hand, the precoding matrix V_p preserves the rest $M_1 - r$ transmit directions for only Rx1 to receiver, and these $M_1 - r$ transmit directions will be sent along the null space of $\left\langle \begin{matrix} H_{12} \\ H_{13} \end{matrix} \right\rangle$ (and henceforth will not be heard by either Rx2 or Rx3). Since the matrix V in GSVD is not unitary, the precoding matrix V_p is not unitary either.

Applying Theorem 3.1 to the MIMO settings and evaluating for the coding distribution resulting from

(3.49)-(3.53) and (3.47), we get the following achievable region for the MIMO one-to-three IC in Theorem 3.2.

Theorem 3.2. *For the MIMO one-to-three IC, let*

$$\beta_2 \triangleq \log\left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| + r_{123}\log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{12})}\right) + (r - r_{13}) \tag{3.58}$$

$$\beta_3 \triangleq \log\left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| + r_{123}\log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{13})}{\sigma_{\min}^2(\Lambda_{13})}\right) + (r - r_{12}) \tag{3.59}$$

then the following region $\mathcal{R}_{\text{in}}$ *given by* (3.60)-(3.67) *is achievable.*

$$\mathcal{R}_{\text{in}} \triangleq \Bigg\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_1 H_{11}^\dagger\right| \tag{3.60}$$

$$R_2 \leq \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{12}^\dagger + \frac{\rho^{\alpha_{22}}}{M_2} H_{22} H_{22}^\dagger\right| - \beta_2 \tag{3.61}$$

$$R_3 \leq \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13}(Q_{12} + Q_{1p}) H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33} H_{33}^\dagger\right| - \beta_3 \tag{3.62}$$

$$R_1 + R_2 \leq \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}(Q_{13} + Q_{1p}) H_{11}^\dagger\right| + \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} Q_1 H_{12}^\dagger + \frac{\rho^{\alpha_{22}}}{M_2} H_{22} H_{22}^\dagger\right| - \beta_2 \tag{3.63}$$

$$R_1 + R_3 \leq \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}(Q_{12} + Q_{1p}) H_{11}^\dagger\right| + \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} Q_1 H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33} H_{33}^\dagger\right| - \beta_3 \tag{3.64}$$

$$\begin{aligned} R_1 + R_2 + R_3 \leq\; & \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger\right| + \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{12} + Q_{13} + Q_{1p}) H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} H_{22}^\dagger\right| \\ & + \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} Q_1 H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33} H_{33}^\dagger\right| - \beta_2 - \beta_3 \end{aligned} \tag{3.65}$$

$$\begin{aligned} R_1 + R_2 + R_3 \leq\; & \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger\right| + \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} Q_1 H_{12}^\dagger + \frac{\rho^{\alpha_{22}}}{M_2} H_{22} H_{22}^\dagger\right| \\ & + \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13}(Q_{12} + Q_{13} + Q_{1p}) H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33} H_{33}^\dagger\right| - \beta_2 - \beta_3 \end{aligned} \tag{3.66}$$

$$\begin{aligned} 2R_1 + R_2 + R_3 \leq\; & \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger\right| + \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}(Q_{12} + Q_{13} + Q_{1p}) H_{11}^\dagger\right| \\ & + \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} Q_1 H_{12}^\dagger + \frac{\rho^{\alpha_{22}}}{M_2} H_{22} H_{22}^\dagger\right| + \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} Q_1 H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33} H_{33}^\dagger\right| \\ & - \beta_2 - \beta_3 \Bigg\} \end{aligned} \tag{3.67}$$

Proof. The evaluation of the inner bound in Theorem 3.1 under the MIMO setting and the coding scheme introduced above is relegated to Appendix A.2. □

Let us take a deeper look at the coding scheme. From the restrictions (3.49)-(3.52), the individual covariance matrices Q_{12}, Q_{13} and Q_{123} can be obtained as (3.68)-(3.70).

$$Q_{12} = \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} V_p \cdot \begin{pmatrix} \frac{\rho^{\alpha_{12}}}{1+\rho^{\alpha_{12}}} I_{12}^\dagger I_{12} + (I_r + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13})^{-1} - (I_r + \rho^{\alpha_{12}} \Lambda_{12}^\dagger \Lambda_{12} + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13})^{-1} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} \cdot V_p^\dagger \tag{3.68}$$

$$Q_{13} = \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} V_p \begin{pmatrix} \frac{\rho^{\alpha_{13}}}{1+\rho^{\alpha_{13}}} I_{13}^\dagger I_{13} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} V_p^\dagger \tag{3.69}$$

$$\begin{aligned} Q_{123} &= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} V_p V_p^\dagger - Q_{1p} - Q_{12} - Q_{13} \\ &= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} V_p \begin{pmatrix} I_r - (I_r + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13})^{-1} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} V_p^\dagger \end{aligned} \tag{3.70}$$

Let $\mathtt{X}_{123}$, $\mathtt{X}_{12}$, $\mathtt{X}_{13}$ and $\mathtt{X}_{1p}$ be zero mean Gaussian vectors with identity covariance matrices, of length r_{123}, r_{12}, r_{13} and M_1 respectively. With their chosen covariance matrices, the auxiliary random vectors W_{123}, W_{12}, W_{13} and W_{1p} can be alternatively written as follows.

$$W_{123} = \sum_{k=1}^{r_{123}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{13}} \lambda_{13,r-r_{13}+k}^2}} V_p^{[r-r_{13}+k]} \mathtt{X}_{123}^{(k)} \tag{3.71}$$

$$\begin{aligned} W_{12} = &\sum_{k=1}^{r-r_{13}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} V_p^{[k]} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{12}}}} \mathtt{X}_{12}^{(k)} \\ &+ \sum_{k=r-r_{13}+1}^{r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{13}}\lambda_{13,k}^2} - \frac{1}{1+\rho^{\alpha_{12}}\lambda_{12,k}^2 + \rho^{\alpha_{13}}\lambda_{13,k}^2}} V_p^{[k]} \mathtt{X}_{12}^{(k)} \end{aligned} \tag{3.72}$$

$$W_{13} = \sum_{k=1}^{r-r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} V_p^{[r_{12}+k]} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{13}}}} \mathtt{X}_{13}^{(k)} \tag{3.73}$$

$$\begin{aligned} W_{1p} = &\sum_{k=1}^{r-r_{13}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{12}}}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} + \sum_{k=r-r_{13}+1}^{r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{12}}\lambda_{12,k}^2 + \rho^{\alpha_{13}}\lambda_{13,k}^2}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} \\ &+ \sum_{k=r_{12}+1}^{r} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{13}}}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} + \sum_{k=r+1}^{M} \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} \end{aligned} \tag{3.74}$$

In the aforementioned equation, $\mathtt{X}_{123}^{(k)}$ denotes the k-th data stream to Rx1-Rx3 that carries (part of) the public sub-message $\mathtt{M}_{123}$ along the transmit direction $V_p^{[r-r_{13}+k]}$, and we call $\mathtt{X}_{123}^{(k)}$ the k-th public data stream to Rx1-Rx3. It needs to be decoded by all three receivers. Similarly, $\mathtt{X}_{12}^{(k)}$ is the k-th public data stream to Rx1 and Rx2 along the transmit direction $V_p^{[k]}$. The data streams $\mathtt{X}_{12}$ can be divided into two groups. The first $r - r_{13}$ data streams are sent with approximately full power, and they can be received only by Rx1 and Rx2 since they are sent along the null space of H_{13}. The rest r_{123} data streams are sent with power $\rho^{-\alpha_{13}}$ so they are decodable by Rx1 and Rx2, but not Rx3, since they arrive under the noise floor of Rx3. The public data streams $\mathtt{X}_{13}$ for Rx1 and Rx3 are only received by Rx1 and Rx3 as they are sent through the null space of H_{12}. The data streams $\mathtt{X}_{1p}$ are received by Rx1 only; we call them private data streams. The first $r - r_{13}$ private data streams are hearable by Rx2, but not Rx3; therefore, they are transmitted at power level $\rho^{-\alpha_{12}}$ so they arrive at Rx2 under the noise floor. The data streams $\mathtt{X}_{1p}^{(r-r_{13}+1)}$, $\cdots$, $\mathtt{X}_{1p}^{(r_{12})}$ are hearable by both Rx2 and Rx3, and they are sent at power level $\rho^{-\alpha_{12}}$ so they arrive at Rx2 and Rx3 under the noise floor. The next $r - r_{12}$ private data streams $\mathtt{X}_{1p}^{(r_{12}+1)}$, $\cdots$, $\mathtt{X}_{1p}^{(r)}$ are hearable by Rx3, but not Rx2, and they are sent at power level $\rho^{-\alpha_{13}}$ so they arrive under the noise floor at Rx3. When there are more transmit antennas at Tx1 than the total receiver antennas at both Rx2 and Rx3, the precoding matrix V_p lets $M_1 - r$ private data streams (the last part on the right hand of (3.74)) transmit along the null space of $\langle V \rangle$, which are exclusively hearable by Rx1. Thus, these $M_1 - r$ private data streams are sent at power level ρ^0.

To keep our statement consistent in the rest of the chapter, we also write the signals X_2 and X_3 in terms of independent data streams, i.e.,

$$X_2 = \sqrt{\frac{P_2}{M_2}}\mathtt{X}_2 \; X_3 = \sqrt{\frac{P_3}{M_2}}\mathtt{X}_3. \tag{3.75}$$

They are only to be received and decoded by their intended receivers.

Remark 3.5. In the SISO (single-input-single-output) one-to-three IC, there is only one transmit direction at Tx1 which can be heard by all three receivers, but there is still interference power disparity at Rx2 and Rx3. Therefore the coding scheme discussed in this subsection is specialized as follows. Tx1 split the message in three parts: $\mathtt{M}_{123}$, $\mathtt{M}_{12}$ and $\mathtt{M}_{1p}$. All three sub-messages are transmitted along the only signal direction the channel has. The $\mathtt{M}_{123}$ will be transmitted at power level ρ^0 so every receiver could decode it. The sub-message $\mathtt{M}_{12}$ is transmitted at power level $\rho^{-\alpha_{13}}$ so it is received below the noise floor at Rx3, but decodable at Rx2 due to extra interference power. The private sub-message $\mathtt{M}_{1p}$ is transmitted at power level $\rho^{-\alpha_{12}}$ so it is only decodable by Rx1. This coding scheme is the same as the coding scheme presented in Section VII of [4].

3.4.3 Outer Bound

We derive a **single** region outer bound $\mathcal{R}_{\text{o}}$ for MIMO one-to-three IC. We provide various combinations of genie informations to Rxi to produce upper bounds on R_i in several different forms, and then linearly combine those upper bounds across $i \in \{1,2,3\}$ to obtain sum rate upper bounds. The outer bound is stated in Theorem 3.3. To present the outer bound, we define the following relevant terms by which the outer bound can be stated in a short form.

$$K_{1p} \triangleq \left(I_{M_1} + \rho^{\alpha_{12}} H_{12}^{\dagger} H_{12} + \rho^{\alpha_{13}} H_{13}^{\dagger} H_{13}\right)^{-1} \tag{3.76}$$

$$K_{12,1p} \triangleq \left(I_{M_1} + \rho^{\alpha_{13}} H_{13}^{\dagger} H_{13}\right)^{-1} \tag{3.77}$$

$$K_{13,1p} \triangleq \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^{\dagger} G_{13} + \rho^{\alpha_{12}} H_{12}^{\dagger} H_{12}\right)^{-1} \tag{3.78}$$

$$K_{12,13,1p} \triangleq \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^{\dagger} G_{13}\right)^{-1} \tag{3.79}$$

Theorem 3.3. *Define*

$$\eta \triangleq \log\left|\max\left\{\lambda_{\max}^2(V_r), 1\right\}\right| + r_{123}\log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{13})}{\sigma_{\min}^2(\Lambda_{12})}\right), \tag{3.80}$$

and a region $\mathcal{R}_{\text{o}}$ *as given by* (3.81)-(3.88),

$$\mathcal{R}_{\text{o}} \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^{\dagger}\right| \tag{3.81}$$

$$R_2 \le \log\left|I_{N_2} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}\right| \tag{3.82}$$

$$R_3 \le \log\left|I_{N_3} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}\right| \tag{3.83}$$

$$R_1 + R_2 \le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^{\dagger}\right| + \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}\right| + \eta \tag{3.84}$$

$$R_1 + R_3 \le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,1p} H_{11}^{\dagger}\right| + \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}\right| \tag{3.85}$$

$$R_1 + R_2 + R_3 \leq \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger} \right| + \log \left| I_{N_2} + \rho^{\alpha_{12}} H_{12} K_{12,13,1p} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger} \right|$$
$$+ \log \left| I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger} \right| \quad (3.86)$$
$$R_1 + R_2 + R_3 \leq \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger} \right| + \log \left| I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger} \right|$$
$$+ \log \left| I_{N_3} + \rho^{\alpha_{13}} H_{13} K_{12,13,1p} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger} \right| + \eta \quad (3.87)$$
$$2R_1 + R_2 + R_3 \leq \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger} \right| + \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,13,1p} H_{11}^{\dagger} \right|$$
$$+ \log \left| I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger} \right| + \log \left| I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger} \right| + \eta \}. \quad (3.88)$$

We have

$$\mathcal{C} \subseteq \mathcal{R}_{\mathrm{o}}.$$

Proof Outline. The fundamental idea in the proof of the outer bound is to construct a virtual channel whose output is then regarded as genie-aided side information to help each receiver to decode its intended signal (therefore making the receiver more interference-resilient). We construct genie informations T_{123}, T_{12} and T_{13} which are identically distributed as the channel side informations S_{123}, S_{12} and S_{13}, respectively, but each pair of corresponding "T" and "S" random vectors are independent conditioned on X_1. The upper bound is proved in three steps. First, by providing one or more of those genie informations to Rxi, $i \in \{1, 2, 3\}$ we derive a series of individual upper bounds on R_i. Some of the bounds may contain entropy terms which can not be single-letterized. Secondly, we linearly combine those individual upper bounds across $i \in \{1, 2, 3\}$ to obtain sum rate upper bounds with unsingle-letterized entropy terms eliminated. At this step, we get an intermediate outer bound in terms of channel side and genie information symbols. This outer bound is a union of polytopes over all admissible input distributions. Finally, we optimize the input distribution to be Gaussian and plug in the optimized distribution to obtain a single region output bound only in terms of the channel parameters. This genie aided argument was first introduced in [43] in characterizing an outer bound for the semi-deterministic two-user interference channel. In [43], there is only one genie information T_i regarding the interference that comes from a particular transmitter Txi, $i \in \{1, 2\}$, and this genie information T_i is given to the intended receiver Rxi only. In the case of the MIMO one-to-three IC, we have various combinations of T_{123}, T_{12} and T_{13} to feed Tx1 to produce a variety of outer bounds on R_1. Besides, the genie information T_{123} is not only fed to Rx1, but to Rx2 and Rx3 as well to produce certain upper bounds on R_2 and R_3. Please refer to Appendix A.3 for details. $\square$

3.4.4 Quantifiable Gap

An achievable rate region of a MIMO one-to-three IC is within gap (n_1, n_2, n_3) to its capacity if for any given rate tuple $(R_1, R_2, R_3) \in \mathcal{C}$, the rate tuple $(R_1 - n_1, R_2 - n_2, R_3 - n_3)$ is within that achievable region. We call the tuple n_i the individual gap on R_i. Since we do not know the capacity region $\mathcal{C}$, we quantify the gap between the inner bound $\mathcal{R}_{\text{in}}$ and the outer bound $\mathcal{R}_{\text{o}}$, and the resulting gap will be an upper bound of the gap (henceforth also a gap) between $\mathcal{R}_{\text{in}}$ and $\mathcal{C}$. The main result in this subsection is stated in Theorem 3.4.

Theorem 3.4. *For any* $(R_1, R_2, R_3) \in \mathcal{R}_{\text{o}}$*, let*

$$(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3) = \left((R_1 - \eta - \delta_1)^+, (R_2 - \beta_2 - \delta_2)^+, (R_3 - \beta_3 - \delta_3)^+\right)$$

, where

$$\delta_1 \triangleq \min\{M_1, N_1\} \left(\log\left(\zeta_{\max} \max\left\{\lambda^2_{\max}(V_r), 1\right\}\right)\right)^+ \tag{3.89}$$

$$\delta_2 \triangleq \min\{M_1 + M_2, N_2\} log \max\left\{\zeta_{\max} \max\left\{\lambda^2_{\max}(V_r), 1\right\}, M_2\right\} \tag{3.90}$$

$$\delta_3 \triangleq \min\{M_1 + M_3, N_3\} \log \max\left\{\zeta_{\max} \max\left\{\lambda^2_{\max}(V_r), 1\right\}, M_3\right\}. \tag{3.91}$$

Then we have

$$(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3) \in \mathcal{R}_{\text{in}}.$$

Proof Outline. There is a one-to-one correspondence between a rate variable R_i on the left hand side and a positive entropy term (in the form of $\log|I_{N_i} + \cdots|$) on the right side of inequalities in both inner and outer bounds. The difference between each pair of corresponding positive entropy terms in the inner and outer bound is upper bounded by δ_i, which contributes to individual gap n_i. Also note there is a one-to-one correspondence between R_2 (and R_3) in the left hand side and β_2 (β_3) on the right hand side of each sum rate restriction in the inner bound, which contributes to the partial individual gap β_2 and β_3 to n_2 and n_3.

Finally, we let the term η in the inequalities (3.84), (3.87) and (3.88) of $\mathcal{R}_{\text{o}}$ be absorbed in n_1. The details of the proof is relegated to Appendix A.4. □

Remark 3.6. Let us compute the gap for the SISO one-to-three IC, where only one antenna is equipped at each transmitter and each receiver. We have $\beta_1 = \beta_2 = 1$, $\eta = 1$ and $\delta_1 = \delta_2 = \delta_3 = 0$. Theorem 3.4.4 tells the coding scheme achieves an achievable region within $(1, 1, 1)$ bit gap to the capacity. It is a smaller gap then the one achieved in [4, Theorem 23] with $K = 3$ therein.

3.5 The GDoF Region of MIMO One-to-three IC

The generalized degrees of freedom (GDoF) is an information-theoretic performance metric that characterizes the number of independent data streams a network could support simultaneously among all users at high SNR regime. In this section, we first compute the GDoF region of the MIMO one-to-three IC, and then focus the achievability of the key corner points in the GDoF region and the sum GDoF curve. In what follows, we define $\bar{\alpha} = \{\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}\}$.

3.5.1 The GDoF Region

The definition of GDoF region of the MIMO one-to-three IC is given in Definition 3.3.

Definition 3.3. The *generalized degrees of freedom* region of a MIMO one-to-three $\mathcal{D}(\bar{\alpha}) \in \mathbb{R}_+^3$ with the capacity region $\mathcal{C}(\bar{\alpha})$ is defined as

$$\left\{(d_1, d_2, d_3) : d_i = \lim_{\rho\to\infty} \frac{R_i}{\log\rho}, i \in \{1, 2, 3\} \text{and} \, (R_1, R_2, R_3) \in \mathcal{C}(\bar{\alpha})\right\}. \tag{3.92}$$

In the rest of the chapter, we call (d_1, d_2, d_3) a GDoF tuple. To compute the GDoF region in this section, we need a slight different version of Lemma 5.1 which is stated in Fact 3.2. They differ in that the matrices H_1, H_2, $\cdots$, H_n only need to be full rank w.p.1 here, whereas the entries of the matrices in Lemma 5.1 are drawn i.i.d. from a continuous unitarily invariant distribution. Fact 3.2 can be proved with similar mathematical induction as in the proof of Lemma 5.1.

Fact 3.2. *Let $H_1 \in \mathbb{C}^{u\times u_i}$, $H_2 \in \mathbb{C}^{u\times u_2}$, $\cdots$, $H_n \in \mathbb{C}^{u\times u_n}$ be n full rank matrices (w.p.1) such that $H = [H_1, H_2, \cdots, H_n]$ is also full rank w.p.1. Then, for asymptotic ρ*

$$\log\det\left(I_u + \sum_{i=1}^{n} \rho^{a_i} H_i H_i^{\dagger}\right) = g(u, (a_1, u_1), \cdots, (a_n, u_n)) \log(\rho) + \mathcal{O}(1), \tag{3.93}$$

where for any $(u, u_1, \cdots, u_n) \in \mathbb{Z}^{+(n+1)}$ and $(a_1, \cdots, a_n) \in \mathbb{R}^n$, the function $g(u, (a_1, u_1), \cdots, (a_n, u_n))$ is defined as

$$\begin{aligned} &g(u, (a_1, u_1), (a_2, u_2), \cdots, (a_n, u_n)) \\ &= \sum_{i=i_1}^{i_n} \left\{ \min\{u, u_{i_1}\} a_{i_1}^+ + \min\left\{(u - u_{i_1})^+, u_{i_2}\right\} a_{i_2}^+ + \cdots + \min\left\{\left(u - \sum_{j=1}^{i_{n-1}} u_j\right)^+, u_{i_n}\right\} a_{i_n}^+ \right\} \end{aligned}$$

for $i_1 \neq i_2 \neq \cdots \neq i_n \in \{1, \cdots, n\}$ such that $a_{i_1} \geq a_{i_2} \geq \cdots \geq a_{i_n}$.

Theorem 3.5. *The GDoF region $D(\bar{\alpha})$ of the MIMO one-to-three IC with $\bar{\alpha} = \{\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}\}$ is given by* (3.94)-(3.101).

$$D(\bar{\alpha}) \triangleq \big\{(d_1, d_2, d_3) \in \mathbb{R}_+^3 :$$

$$d_1 \leq \min\{M_1, N_1\}\alpha_{11} \tag{3.94}$$

$$d_2 \leq \min\{M_2, N_2\}\alpha_{22} \tag{3.95}$$

$$d_3 \leq \min\{M_3, N_3\}\alpha_{33} \tag{3.96}$$

$$d_1 + d_2 \leq g\left(N_1, \left((\alpha_{11} - \alpha_{12})^+, r_{12}\right), (\alpha_{11}, M_1 - r_{12})\right) + g\left(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)\right) \tag{3.97}$$

$$d_1 + d_3 \leq g\left(N_1, \left((\alpha_{11} - \alpha_{13})^+, r_{13}\right), (\alpha_{11}, M_1 - r_{13})\right) + g\left(N_3, (\alpha_{13}, M_1), (\alpha_{33}, M_3)\right) \tag{3.98}$$

$$\begin{aligned} d_1 + d_2 + d_3 \leq\ & g\left(N_1, \left((\alpha_{11} - \alpha_{12})^+, r_{12}\right), \left((\alpha_{11} - \alpha_{13})^+, r - r_{12}\right), (\alpha_{11}, M_1 - r)\right) \\ &+ g\left(N_2, \left((\alpha_{12} - \alpha_{13})^+, r_{123}\right), (\alpha_{12}, r_{12} - r_{123}), (\alpha_{22}, M_2)\right) + g\left(N_3, (\alpha_{13}, M_1), (\alpha_{33}, M_3)\right) \end{aligned} \tag{3.99}$$

$$\begin{aligned} d_1 + d_2 + d_3 \leq\ & g\left(N_1, \left((\alpha_{11} - \alpha_{12})^+, r_{12}\right), \left((\alpha_{11} - \alpha_{13})^+, r - r_{12}\right), (\alpha_{11}, M_1 - r)\right) \\ &+ g\left(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)\right) + g\left(N_3, (\alpha_{13}, r_{13} - r_{123}), (\alpha_{33}, M_3)\right) \end{aligned} \tag{3.100}$$

$$\begin{aligned} 2d_1 + d_2 + d_3 \leq\ & g\left(N_1, \left((\alpha_{11} - \alpha_{13})^+, r_{123}\right), (\alpha_{11}, M_1 - r_{123})\right) \\ &+ g\left(N_1, \left((\alpha_{11} - \alpha_{12})^+, r_{12}\right), \left((\alpha_{11} - \alpha_{13})^+, r - r_{12}\right), (\alpha_{11}, M_1 - r)\right) \end{aligned}$$

$$+ g\left(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)\right) + g\left(N_3, (\alpha_{13}, M_1), (\alpha_{33}, M_3)\right) \Big\}. \tag{3.101}$$

Proof Outline. The GDoF region of the MIMO one-to-three IC is presented in the theorem below. In Definition 3.3, the GDoF region is defined via the capacity region $C(\bar{\alpha})$. We do not have the exact capacity region determined for the MIMO one-to-three IC, but Theorem 3.4 has shown that both $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$ are within quantifiable gap to the capacity region and that the gap is independent of the channel SNR and INR. Because a finite number of bits are insignificant in the GDoF computation, the GDoF region can be obtained from either $\mathcal{R}_{\text{in}}$ or $\mathcal{R}_{\text{o}}$. Please refer to Appendix A.5 for the complete proof. □

Example 3.1. Consider the MIMO one-to-three IC with the following parameters: $\alpha_{11} = \alpha_{22} = \alpha_{33} = 1$, $\alpha_{12} = 0.6$, $\alpha_{13} = 0.3$, $M_1 = N_1 = 3$ and $M_2 = M_3 = N_2 = N_3 = 2$. Given this setting, we have $r = 3$, $r_{123} = 1$ and $r_{12} = r_{13} = 2$. The GDoF region is plotted in Fig. 3.6.

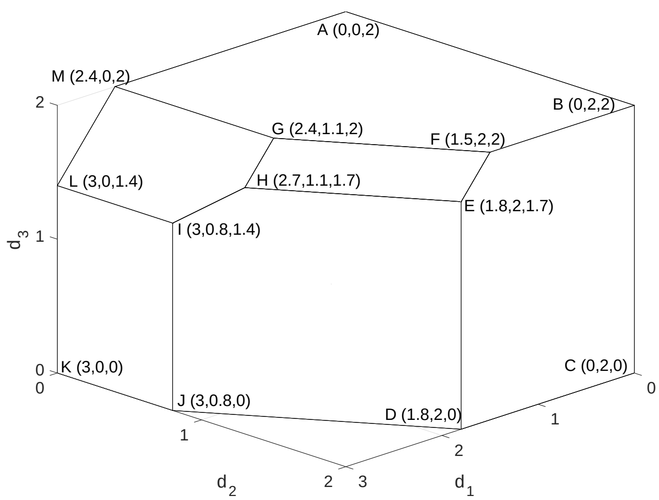

Figure 3.6: GDoF region of a MIMO one-to-three IC with $M_1 = N_1 = 3$, $M_2 = M_3 = N_2 = N_3 = 2$, $\alpha_{12} = 0.6$ and $\alpha_{13} = 0.3$

We provide an overview of the GDoF region in Example 3.1. The MIMO one-to-three IC consists of two two-user Z ICs as its sub-channels. The tuples on the $(d_1, d_2, 0)$ form the GDoF region of the two-user Z IC with INR $\rho^{\alpha_{12}}$ and the tuples on the $(d_1, 0, d_3)$ plane form the GDoF region of the other two-user Z IC with INR $\rho^{\alpha_{13}}$. The rate tuples on d_3 vs d_2 plane when $d_1 = 0$ reflect the GDoF region of a parallel channel with Tx2/Rx2 and Tx3/Rx3 while Tx1 is off. The sum GDoF plane is E-F-G-H, and any GDoF tuple on this plane achieves the max sum GDoF 5.5.

Example 3.2. Continue with the MIMO one-to-three IC in Example 3.1. We describe the structure of the transmitted signals from the three transmitters in terms of independent data streams according to (3.71)-(3.74) in Section 3.4.2. Our coding scheme suggests we send the following data streams at Tx1.

$$
\begin{aligned}
W_{123} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} V_p^{(2)} \sqrt{1 - \frac{1}{1+\rho^{0.3}\lambda_{13,r-r_{13}+k}^2}} \mathtt{X}_{123}^{(1)} \\
W_{12} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \left(V_p^{(1)} \sqrt{1 - \frac{1}{1+\rho^{0.6}}} \mathtt{X}_{12}^{(1)} + V_p^{(2)} \sqrt{\frac{1}{1+\rho^{0.3}\lambda_{13,2}^2} - \frac{1}{1+\rho^{0.6}\lambda_{12,2}^2 + \rho^{0.3}\lambda_{13,2}^2}} \mathtt{X}_{12}^{(2)} \right) \\
W_{13} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} V_p^{(3)} \sqrt{1 - \frac{1}{1+\rho^{0.3}}} \mathtt{X}_{13}^{(1)} \\
W_{1p} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^\dagger)}} \left(V_p^{(1)} \sqrt{\frac{1}{1+\rho^{0.6}}} \mathtt{X}_{1p}^{(1)} + V_p^{(2)} \sqrt{\frac{1}{1+\rho^{0.6}\lambda_{12,k}^2 + \rho^{0.3}\lambda_{13,k}^2}} \mathtt{X}_{1p}^{(2)} + V_p^{(3)} \sqrt{\frac{1}{1+\rho^{0.3}}} \mathtt{X}_{1p}^{(3)} \right).
\end{aligned}
$$

More specifically, there is one public data stream $\mathtt{X}_{123}^{(1)}$ for all receivers, two public data streams $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{12}^{(2)}$ for Rx1 and Rx2, one public data stream $\mathtt{X}_{13}^{(1)}$ for Rx1 and Rx3, and three private data streams for Rx1 only. The data streams $\mathtt{X}_{123}^{(1)}$, $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{13}^{(1)}$ are sent at power level ρ^0. The data stream $\mathtt{X}_{12}^{(2)}$ is sent at power level $\rho^{-0.3}$ as this is the part to be decoded by Rx2, but treated as noise by Rx3. The first private data stream $\mathtt{X}_{1p}^{(1)}$ is sent at power level $\rho^{-0.6}$ so that Rx2 can treat it as noise. The second private data stream $\mathtt{X}_{1p}^{(2)}$ is sent at power level $\rho^{-0.6}$ so that both Rx2 and Rx3 can treat it as noise. The third private data stream $\mathtt{X}_{1p}^{(3)}$ is sent at power level $\rho^{-0.3}$ so Rx3 can treat is as noise.

In what follows, we analyze the achievability of the four corner points on the max sum GDoF plane in Example (3.1). For each corner point, we provide GDoF distribution among the data streams revealed in Example (3.2). The detailed GDoF allocation of each data stream will be illustrated via multi-dimensional signal partitioning introduced in Chapter 2. Each GDoF allocation will be plotted in a signal diagram with each of the received independent (transmit) signal directions (from the receiver's perspective) plotted as a multi-leveled bar whose top level marks its signal strength and the vertical height of each partition is proportional to the DoFs carried by it. The underlying coding scheme can be directly read from the GDoF allocation. A transmitter encodes all data streams on a (transmit) signal direction by multi-level superposition coding from bottom to top (refer to the figure for the position of the data streams on each signal direction), and the receiver decodes the signal by either successive cancellation or joint decoding. No cross signal level is employed, and each data stream is encoded independently. The underlying coding scheme can be different from the coding scheme we used to derive the inner bound in Section 3.4.2. In all the GDoF analysis figures in the rest of the chapter, the transmit signal directions $V_p^{[1]}$, $V_p^{[2]}$, and $V_p^{[3]}$ are sorted from left to right in sequence at Rx1, $V_p^{[1]}$ and $V_p^{[2]}$ are shown from left to right at Rx2, and $V_p^{[2]}$ and $V_p^{[3]}$ are shown from left to right at Rx3.

3.5.1.1 The achievability of point E (1.8,2,1.7)

We choose the GDoF distribution $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0$, $d_{12}^{(2)} = 0$, $d_{13}^{(1)} = 0.3$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_2^{(1)} = d_2^{(2)} = 1$, $d_3^{(1)} = 1$ and $d_3^{(2)} = 0.7$. The GDoF allocation among the three transmitters are illustrated in Fig. 3.7. This allocation guarantees an interference free channel between Tx2 and Rx2. Due to the precoding (by matrix V_p), the second and third transmit directions $V_p^{[2]}$ and $V_p^{[3]}$ do not appear at Rx2 and Rx3, respectively. All the signal levels at both Rx2 and Rx3 are fully utilized.

Rx1 first removes the effect of both $\mathtt{X}_{1p}^{(1)}$ and $\mathtt{X}_{1p}^{(2)}$ from Y_1 by zero forcing, i.e. projecting the received signal onto the 1-D plane which is perpendicular to $H_{11}V_p^{[1:2]}$. $\mathtt{X}_{13}^{(1)}$ can be recovered by treating the contribution of $\mathtt{X}_{1p}^{(3)}$ as noise (with the equivalent noise floor raised to $\rho^{0.7}$), which gives $d_{13}^{(1)} = 0.3$. Since $\mathtt{X}_{13}^{(1)}$ should also be decoded by Rx3, we shall verify the achievability of $d_{13}^{(1)} = 0.3$ at Rx3 later. Subtracting the effect of $\mathtt{X}_{13}^{(1)}$, $\mathtt{X}_{1p}^{(3)}$ can be recovered with GDoF $d_{1p}^{(3)} = 0.7$. After both $\mathtt{X}_{13}^{(1)}$ and $\mathtt{X}_{1p}^{(3)}$ are recovered, we remove their effects on the received signal Y_1. Then we see a 2×2 MIMO P2P channel between Tx1 and Rx1, and data streams $\mathtt{X}_{1p}^{(1)}$ and $\mathtt{X}_{1p}^{(2)}$ can then be recovered, which gives $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$. Rx2 directly decodes its intended signals as the interference from Tx1 will be under the noise floor. Rx3 decodes its intended signal by treating the interference from Tx1 as noise. It can be seen that the data stream $\mathtt{X}_{13}^{(1)}$ can indeed achieve GDoF 0.3 at Rx3 as well as at Rx1.

3.5.1.2 The achievability of point F (1.5,2,2)

The GDoF distribution among data streams could be $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0$, $d_{12}^{(2)} = 0$, $d_{13}^{(1)} = 0$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_2^{(1)} = d_2^{(2)} = 1$ and $d_3^{(1)} = d_3^{(2)} = 1$. The GDoF allocation among the three transmitters are illustrated in Fig. 3.8. This coding scheme guarantees interference free GDoF for the entire network.

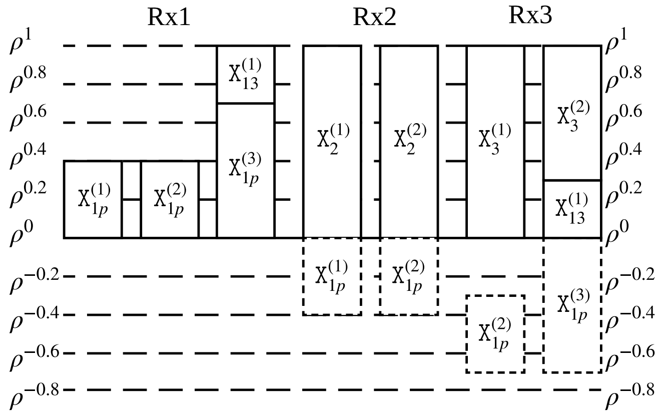

Figure 3.7: GDoF allocation at corner point E

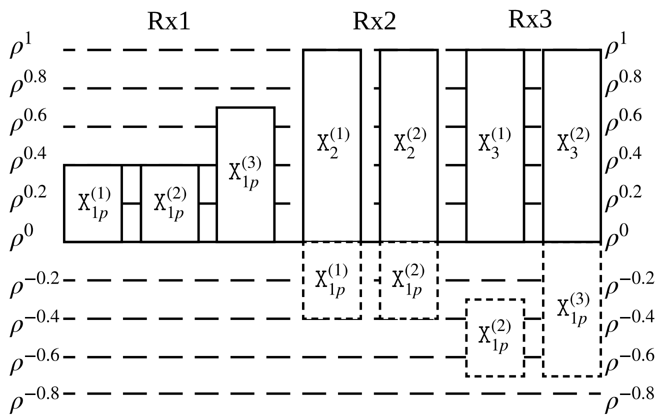

Figure 3.8: GDoF allocation at corner point F

3.5.1.3 The achievability of point G (2.4,1.1,2)

We choose the GDoF distribution among data streams as follows: $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0.3$, $d_{12}^{(2)} = 0.3$, $d_{13}^{(1)} = 0$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_{2}^{(1)} = 0.7$, $d_{2}^{(2)} = 0.4$ and $d_{3}^{(1)} = d_{3}^{(2)} = 1$. The GDoF allocation among the three transmitters are illustrated in Fig. 3.9. This coding scheme brings an interference free channel between Tx3 and Rx3. As stated previously, the data stream $\mathtt{X}_{12}^{(1)}$ is sent onto the null space of $\langle H_{13} \rangle$, so it does not appear at Rx3; however, the $\mathtt{X}_{12}^{(2)}$ is exclusive to Rx1 and Rx2 due to the difference $\alpha_{12} \leq \alpha_{13}$; therefore, $\mathtt{X}_{12}^{(2)}$ appears at Rx3 under the noise floor.

Rx1 first removes the effects of $\mathtt{X}_{12}^{(1)}$, $\mathtt{X}_{1p}^{(1)}$, $\mathtt{X}_{12}^{(2)}$ and $\mathtt{X}_{1p}^{(2)}$ from Y_1 by zero forcing, i.e. projecting the received signal onto the 1-D plane which is perpendicular to $H_{11}V_p^{[1:2]}$. $\mathtt{X}_{1p}^{(3)}$ can be recovered with $d_{1p}^{(3)} = 0.7$.

Subtracting the effect of $\mathtt{X}_{1p}^{(3)}$ from Y_1 , we see a 2×2 MIMO P2P channel between Tx1 and Rx1, and data streams $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{12}^{(2)}$ can be recovered by treating $\mathtt{X}_{1p}^{(1)}$ and $\mathtt{X}_{1p}^{(2)}$ as noise, which gives $d_{12}^{(1)} = 0.6$ and $d_{12}^{(2)} = 0.3$. Subtracting the effect of $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{12}^{(2)}$, $\mathtt{X}_{1p}^{(1)}$ and $\mathtt{X}_{1p}^{(2)}$ can be decoded, resulting $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$. The power level assignments of $\mathtt{X}_{2}^{(1)}$, $\mathtt{X}_{2}^{(2)}$, $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{12}^{(2)}$ permit these data streams to be jointly decoded, resulting $d_{2}^{(1)} = 0.4$, $d_{2}^{(2)} = 0.7$, $d_{12}^{(1)} = 0.6$ and $d_{12}^{(2)} = 0.3$. Rx3 decodes its intended signal only, leading $d_{3}^{(1)} = d_{3}^{(2)} = 1$.

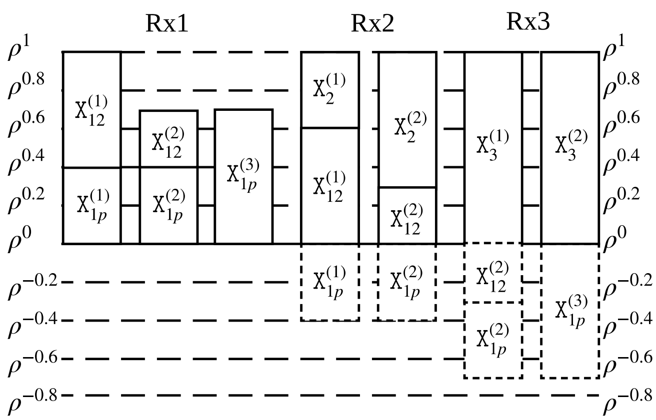

Figure 3.9: GDoF allocation at corner point G

3.5.1.4 The achievability of point H (2.7,1.1,1.7)

The GDoF distribution among data streams could be $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0.3$, $d_{12}^{(2)} = 0.7$, $d_{13}^{(1)} = 0$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_{2}^{(1)} = 0.7$, $d_{2}^{(2)} = 0.4$ and $d_{3}^{(1)} = d_{3}^{(2)} = 1$. The GDoF allocation among the three transmitters are illustrated in Fig. 3.10.

Remark 3.7. In Example 3.1, the interference strength is moderate ($\alpha_{13} = 0.6$) from Tx1 to Rx2 and weak ($\alpha_{13} = 0.3$) from Tx1 to Rx3. In the four achievable schemes discussed above, we keep data streams $\mathtt{X}_{123}$ null because $\mathtt{X}_{123}$ is received above the noise floor at both Rx2 and Rx3 and decoding $\mathtt{X}_{123}$ will cause GDoF reduction at both Rx2 and Rx3. More specifically, in the considered channel setting, carrying messages on data stream $\mathtt{X}_{12}$ or $\mathtt{X}_{13}$ brings the same amount of GDoF to Rx1 as the reduction of GDoF at Rx2 or Rx3, respectively. However, carrying message on data stream $\mathtt{X}_{123}$ reduces twice the GDoF at Rx2 and Rx3 in total than the GDoF obtained by Rx1. We illustrate a GDoF distribution when $\mathtt{X}_{123}$ is active in Fig. 3.11.

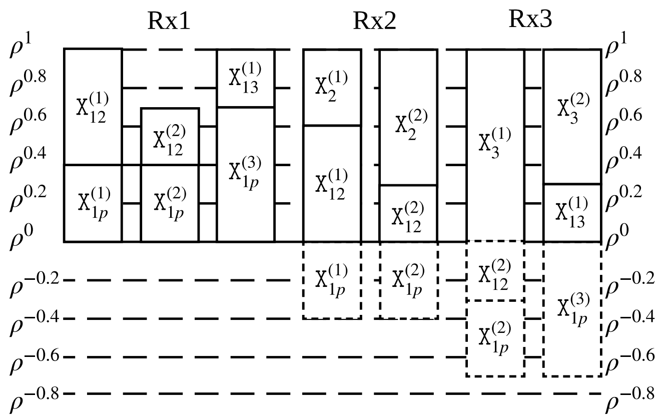

Figure 3.10: GDoF distribution at corner point H

It is a coding scheme for corner point I in Fig. 3.6, which results sum GDoF 5.2.

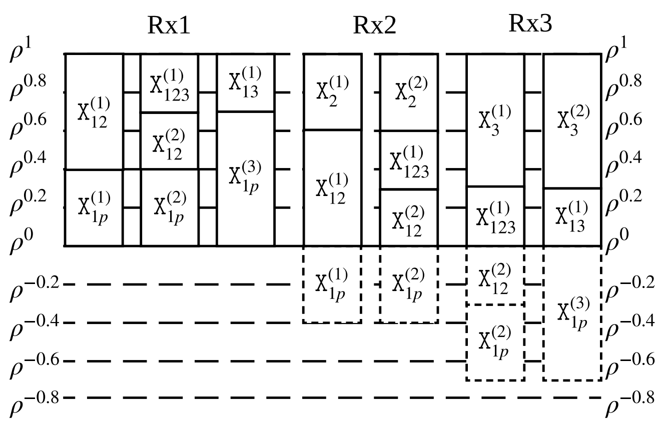

Figure 3.11: GDoF allocation at corner point I

3.5.2 The Sum GDoF Curve

Next, we keep the number of transmit and receive antennas unchanged in Example (3.1), and let α run through the internal $[0, 2]$ to see the variation of the sum GDoF. The sum GDoF vs α curve is plotted in Fig. 3.12. There are two corner points on the curve. At the first corner point the interference to Rx2 becomes strong interference, i.e. $\alpha_{12} = 1$. At the second corner point the interference to Rx3 becomes strong interference, i.e. $\alpha_{13} = 1$. In between the two corner points, one of the interference channels is moderate, i.e. $\alpha_{13} \in [0.5, 1]$, and the other one is strong, i.e. $\alpha_{12} = [1, 2]$. We focus a sum GDoF optimal corner point

$(\alpha, d_{sum}) = (0.8, 4.8)$ on the curve. By Theorem 3.5, we the equal GDoF tuple $(1.6, 1.6, 1.6)$ is achievable. Coding scheme to achieve this tuple is not unique, and we presents two different coding schemes through the multi-dimensional signal level partitioning in Fig. 3.13 and Fig. 3.14. In the first coding scheme, data stream $\mathtt{X}_{123}$ is active, contrary to the case when both interferences are not strong in the previous example, activating data stream $\mathtt{X}_{123}$ can be sum GDoF optimal because Rx2 receives strong interference, and $\mathtt{X}_{123}^{(1)}$ is received above ρ^0, therefore it does not deduct GDoF at Rx2. The second coding scheme achieves the same GDoF tuple, but only uses two antennas at Tx1 to transmit the signal. It is simpler and more energy efficient than the first coding scheme.

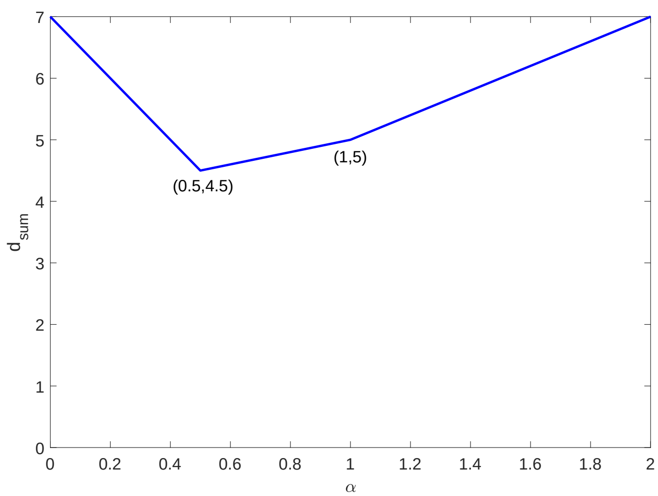

Figure 3.12: Sum GDoF curve of MIMO one-to-three IC with $M_1 = N_1 = 3$, $M_2 = M_3 = N_2 = N_3 = 2$, $\alpha_{12} = 2\alpha$ and $\alpha_{13} = \alpha$

Lastly, we plot the sum GDoF curve of a SISO one-to-three IC in Fig. 3.15. There is only one antenna at each transmitter and each receiver, and again we choose $\alpha_{13} = \alpha$ and $\alpha_{12} = 2\alpha$.

3.6 Conclusion

We delve into the channel structure of the MIMO one-to-three IC with the aid of GSVD, and designed a explicit coding scheme which adapts the channel structure. A pair of single region inner and outer bounds are derived and shown to be within a SNR/INR independent gap. The GDoF region of the MIMO one-to-three IC is then fully characterized. We also numerically studied achievability of GDoF region and sum

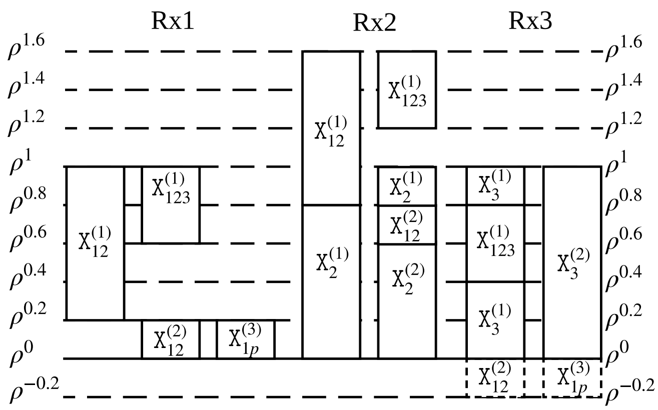

Figure 3.13: Coding scheme 1 for GDoF tuple $(1.6, 1.6, 1.6)$ of the considered example

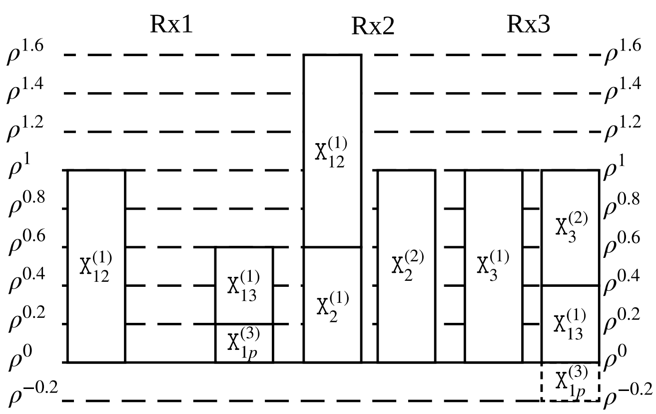

Figure 3.14: Coding scheme 2 for GDoF tuple $(1.6, 1.6, 1.6)$ of the considered example

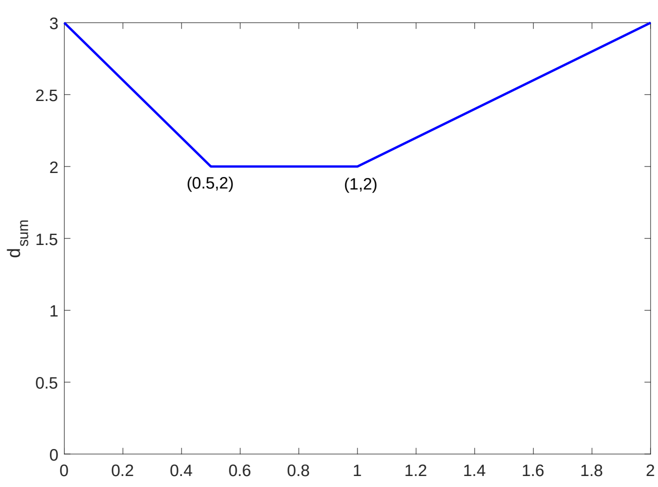

Figure 3.15: Sum GDoF curve of SISO one-to-three IC with $\alpha_{12} = 2\alpha$ and $\alpha_{13} = \alpha$

GDoF curve of several channel examples.

Chapter 4

Generalized Degrees of Freedom Region of the MIMO IC-ZIC

4.1 Introduction

An IC-ZIC is a single-hop multi-terminal network with three transmitters (Tx1-Tx3) and three receivers (Rx1-Rx3). Tx1/Rx1 and Tx2/Rx2 form a two-user interference channel (IC). Tx1/Rx1 and Tx3/Rx3 form a two-user Z interference channel (ZIC) where interference only comes from Tx1 to Rx3 (refer to Fig. 4.4 for a discrete memoryless IC-ZIC or Fig. 4.2.2 for a MIMO IC-ZIC). An IC-ZIC can also be formed from the one-to-three interference channel by adding one more interference link from one of the non-interfering transmitters (Tx2 or Tx3) to Rx1 therein. Without loss of generality, we let this interference link come from Tx2 to Rx1. To illustrate the practical scenarios which IC-ZIC models, we borrow and modify the two practical scenarios of the one-to-three IC introduced in Fig. 3.1 and Fig. 3.2. In Fig. 4.1, the entire area is a macro cell served by the radio tower Tx1, and two small cells operate on the same carrier frequency inside the macro cell. The transmit power used by the macro cell transmitter (Tx1) is higher than the transmit power at Tx2 and Tx3 in the two small cells. We use solid lines to represent direct links and dashed lines interference links. The interference pattern shown in the figure is a consequence of the transmit power disparity and channel topology. One such application of this scenario is the cellular network range expansion by deploying multiple lower power pico eNBs (Tx2 and Tx3) under a macro cell centered with a macro eNB (Tx1) [3, Figure 1]. As Tx1 transmits at significantly higher power level than Tx2 or Tx3 (to cover the entire area), there are interferences from Tx1 to Rx2 and Rx3; the interference from Tx2 to Rx1 exists because Rx1 is located on the margin of the small cell 1; the interference from Rx3 to Tx1 is negated since the small cell 2 is located far

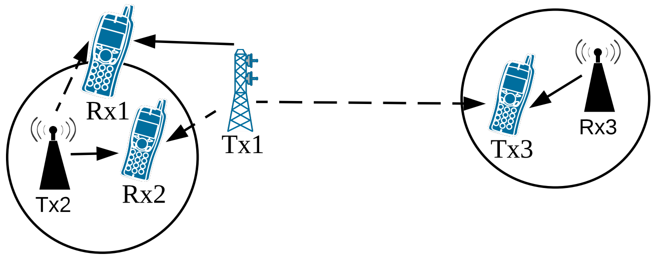

Figure 4.1: A MIMO IC-ZIC where the macro cell transmitter transmits at significantly higher power level than the small cell transmitters

enough from Rx1. Also, as seen from Fig. 4.1, because the small cell 2 is located further from Tx1 than the small cell 1, the interference strength from Tx1 to Rx2 is stronger than from Tx1 to Rx3. In the one-to-three IC, Rx2 can be assumed to receive stronger interference than Rx3 without loss of generality. However, since we have fixed Tx2 to be another interfering transmitter in IC-ZIC, we do no assume the interference strength at Rx2 is stronger than Rx3. For example, in Fig. 4.2, all transmitters transmit at the same power level, and Tx1/Rx1 and Tx2/Rx2 have mutual interference. The path-loss difference from Tx1 to Rx2 and Rx3 yields disparity of the interference strength at the two receivers. But as Rx3 is located closer to Tx1, it receives stronger interference from Tx1 than Rx2.

The main goal of this chapter is to characterize the fundamental generalized degrees of freedom (GDoF) region of the MIMO IC-ZIC. As an outcome of this research, the sum GDoF curve of the scenario shown in Fig. 4.1 is plotted in Fig. 4.3 with a practical set of channel parameters given in the caption. The channel model and parameters are defined in Section 4.2. The GDoF region and sum GDoF curve with respect to Fig. 4.2 will be investigated in Section 4.5.

4.1.1 Main Contributions

We obtain single region inner and outer bounds for the MIMO IC-ZIC. Since Tx1 produces interferences to two receivers, we employ the same three level superposition coding scheme at Tx1 as in the MIMO one-to-three IC (c.f. Section 3.4.2). More specifically, the message $\mathtt{M}_1$ at Tx1 is split into four sub-messages: $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$. As their subscripts indicate, they are to be decoded by Rx1 only, Rx1 and Rx2, Rx1 and Rx3, and Rx1-Rx3, respectively. The four sub-messages are coded independently according to a vector

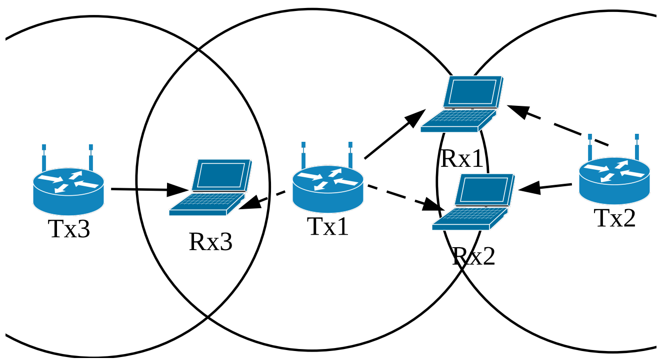

Figure 4.2: A MIMO IC-ZIC all transmitter transmit at the same power level

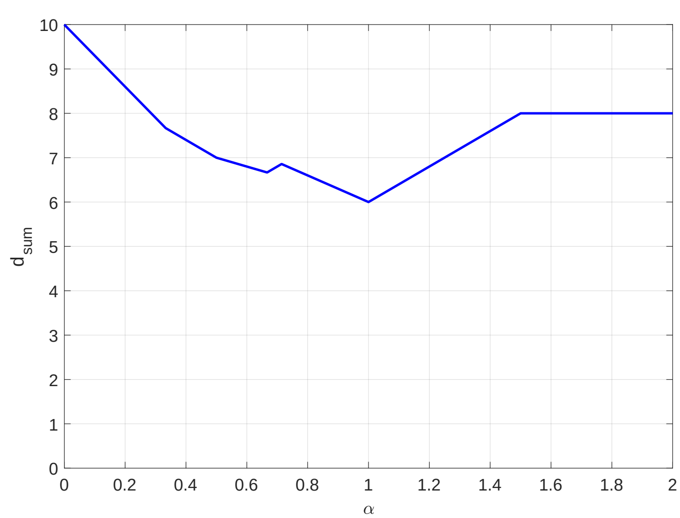

Figure 4.3: Sum GDoF curve of the scenario shown in Fig. 4.1. The parameters are chosen as follows. Tx1 and Rx1 are equipped with 3 antennas each; Tx2, Rx2, Tx3 and Rx3 are equipped with 2 antennas each; Tx1, Tx2 and Tx3 transmit at power $\rho^{2\alpha}$, ρ^{α} and ρ^{α} (to reflect the transmit power disparity); the interference strength from Tx1 to Rx2, Tx2 to Rx1 and Tx1 to Rx3 are ρ^{α}, ρ^{α} and $\rho^{\alpha/2}$, respectively.

Gaussian distribution with **explicitly** specified covariance matrices, and they are additively superposed and transmitted. In particular, those covariance matrices are specified via the GSVD of the two cross channel matrices from Tx1 to Rx2 and Rx3. Since Tx2 only produces interference to Rx1, we employ Karmakar-Varanasi type [27] coding scheme at Tx2. The transmitted message $\mathtt{M}_2$ is split into common sub-message $\mathtt{M}_{21}$ and private message $\mathtt{M}_{2p}$ to be coded by Rx1-Rx2 and Rx2 only, respectively. Tx3 simply encodes its entire message using single user Gaussian codebook without water-filling. Consequently, a single and explicit polyhedral inner bound is obtained. As a byproduct, a per-distribution inner bound is also obtained for the discrete-memoryless IC-ZIC. The outer bound is obtained by providing various combinations of genie information to the receivers. The difficulty lies within the quantification of the gap from the obtained inner and outer bounds which contain 33 and 28 inequalities respectively. It is done with the aid of a series of supporting lemmas which reveals the relationship between the set functions in the inner and outer bounds (to be defined later). The gap between the inner and outer bounds is quantified and shown to be independent of SNRs and INRs (with increasing nominal SNR). Hence, such a gap is tight enough to characterize the fundamental generalized degrees of freedom (GDoF) region. In the end, we analyze the GDoF and sum GDoF achievability of several channel examples with multi-dimensional signal level partitioning introduced in Chapter 2.

4.1.2 Previous Related Works

The two-user IC and the one-to-three IC are two sub-channels embedded in the IC-ZIC. We summarize the known results regarding these two sub-channels. For the general DM two-user IC, the Han-Kobayashi achievable scheme (HK scheme) in [23], as well as its alternative, the CMG scheme of [11], give the (same) best inner bound to the capacity region known to date. Telatar and Tse [43] found an outer bound for the class of semi-deterministic interference channels and quantified the gap to the CMG inner bound. The idea of genie-aided argument in the proof of the outer bound of MIMO IC-ZIC was first introduced in [43]. Etkin et al [15] and Karmakar and Varanasi [27] characterized constant-gap-to-capacity regions for the Gaussian scalar and vector two-user ICs, respectively. The GDoF region of the MIMO two-user IC was characterized and studied in [26]. A constant-gap-to-capacity region of the Gaussian scalar one-to-many IC was reported by [4]. The GDoF region of the MIMO one-to-three IC was characterized previously in Section 3.5.

4.1.3 Notations

Throughout, the i-th transmitter/receiver is denoted as Txi/Rxi for $i \in \{1,2,3\}$, and its message, transmit symbol, rate and degrees of freedom (GDoF) are denoted as $\mathtt{M}_i$, X_i, R_i and d_i, respectively. The number of antenna at Txi and Rxi is denoted as M_i and N_i, respectively.

We use capital letters to denote random vectors such as X_i. The underlying alphabets are denoted by $\mathcal{X}_i$, and specific values by x_i. We use the usual short hand notation for (conditional) probability distributions where the lower case arguments also denote the random variables whose (conditional) distribution is being considered. For example, $p(y_i|x_i)$ denotes $p_{\mathrm{Y}_i|\mathrm{X}_i}(y_i|x_i)$.

We use $\mathbb{C}$ to denote the set of complex numbers and $Z \sim \mathcal{CN}(0, I_N)$ to denote a N-dimensional random vector Z that obeys the complex circularly symmetric Gaussian distribution with zero mean and covariance matrix I_N (the $N \times N$ identity matrix). The note either $\det(\cdot)$ or $|\cdot|$ is used to represent the determinant of a matrix. The number of antennas at Tx$i.j$ and Rxi are denoted as $M_{i.j}$ and N_i. The Frobenius norm of a matrix H is denoted by $\|H\|_{\mathrm{F}}^2$, i.e., $\|H\|_{\mathrm{F}}^2 = \mathrm{Tr}(HH^\dagger)$, where $\mathrm{Tr}(\cdot)$ returns the trace of a given matrix. We use $\mathbb{U}^{N\times N}$ to represent the set of $N \times N$ unitary matrices. The k-th row and column of the matrix H are denoted as $H^{(k)}$ and $H^{[k]}$ respectively. A sub-matrix obtained by taking the rows k_1 to k_2 of the matrix H is written as $H^{(k_1:k_2)}$. A sub-matrix obtained by taking the columns k_1 to k_2 of the matrix H is written as $H^{[k_1:k_2]}$. The linear span of matrix H is denoted as $\langle H \rangle$. For two matrices A and B, if $(A - B)$ is positive definite (p.d.) or positive semi-definite (p.s.d), we write the relationship as $A \succ B$ or $A \succeq B$, respectively. We use $o(1)$ to represent a term which approaches zero asymptotically, and $\mathcal{O}(1)$ to represent a term which is bounded above by some constant. The function $(M)^+$ returns the maximum value of M and 0, i.e., $(M)^+ = \max\{M, 0\}$. The minimum and maximum singular value of a matrix H are denoted as $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$, respectively. We refer rectangular diagonal matrix as any matrix whose nonzero entries only appear on one particular diagonal (not necessarily the main diagonal). The diagonal values of a rectangular diagonal matrix are the entries on that diagonal which contains nonzero values. The minimum and maximum nonzero diagonal values of a rectangular diagonal matrix Σ are denoted as $\sigma_{\min}(\Sigma)$

and $\sigma_{\max}(\Sigma)$, respectively.

The rest of the chapter is organized as follows. Section 4.2 defines the DM and MIMO IC-ZIC channel models and discuss the channel structure of the MIMO IC-ZIC. Section 4.3 presents the multi-level superposition coding and the resulting inner bound for the DM IC-ZIC. Section 4.4 presents single region inner and outer bounds for the MIMO IC-ZIC. The gap between the bounds is quantified. The GDoF region of the MIMO IC-ZIC will be characterized in Section 4.5. Section 4.6 concludes the chapter. Many proofs are relegated to the Appendices.

4.2 Channel Models

In this section, we first introduce the general discrete memoryless IC-ZIC and the MIMO IC-ZIC. Then we explain the channel structure of the MIMO IC-ZIC using the generalized singular value decomposition of the two cross channel matrices.

4.2.1 The Discrete Memoryless (DM) IC-ZIC

An IC-ZIC channel consists three transmitters and three receivers. The sub-channel between the Tx1/Rx1 and Tx2/Rx2 is a two-user interference channel, the sub-channel between the Tx1/Rx1 and Tx3/Rx3 a Z interference channel, and the sub-channel between Tx1/Rx1 and Tx3/Rx3 a parallel channel. The discrete memoryless (DM) IC-ZIC is defined in Definition 4.1 and depicted in Fig. 4.4.

Definition 4.1. A discrete memoryless IC-ZIC is a three-transmitter and three-receiver network $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3,\ p(y_1, y_2, y_3|x_1, x_2, x_3),\ \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3)$ with transition probability satisfying

$$\begin{aligned} &p(y_1^n, y_2^n, y_3^n|x_1^n, x_2^n, x_3^n) \\ &\quad = \prod_{t=1}^{n} \left(p(y_{1t}|x_{1t}, x_{2t})p(y_{2t}|x_{1t}, x_{2t})p(y_{3t}|x_{1t,}x_{3t})\right) \end{aligned} \tag{4.1}$$

The input and output symbols X_i and Y_i are taken from the discrete alphabets $\mathcal{X}_i$ and $\mathcal{Y}_i$ respectively, for each $i \in \{1, 2, 3\}$. Message $\mathtt{M}_i$ is generated from set $\mathcal{M}_i$ uniformly at random, and encoded at transmitter Txi. Receiver Rxi decodes $\mathtt{M}_i$ as $\hat{\mathtt{M}}_i$.

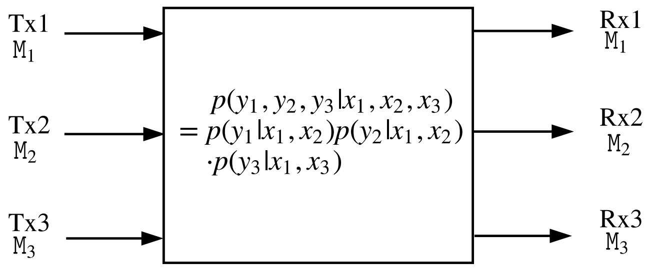

Figure 4.4: DM IC-ZIC

Given the channel as defined in Definition 4.1, a $(n, R_1, R_2, R_3, P_e^{(n)})$ coding scheme for a DM IC-ZIC consists of

- $\mathtt{M}_i$, the message to transmit at Txi, assumed to be uniformly distributed over $\mathcal{M}_i \in \{1, \cdots, 2^{nR_i}\}$, for each $i \in \{1, 2, 3\}$;
- Encoding functions $f_i(\cdot)$ such that

$$\mathtt{f}_i(\cdot): \qquad \mathcal{M}_i \longmapsto \mathcal{X}_i^n,\ m_i \longmapsto x_i^n(m_i).$$

- Decoding functions $g_i(\cdot)$ such that

$$\mathtt{g}_i(\cdot): \ \mathcal{Y}_i^n \longmapsto \mathcal{M}_i,\ y_i^n \longmapsto \hat{m}_i(y_i^n).$$

The probability of error $P_e^{(n)}$ is defined to be

$$P_e^{(n)} = P\left\{\mathtt{M}_1 \neq \hat{\mathtt{M}}_1, \mathtt{M}_2 \neq \hat{\mathtt{M}}_2 \text{ or } \mathtt{M}_3 \neq \hat{\mathtt{M}}_3\right\}.$$

A three rate-tuple (R_1, R_2, R_3) is said to be achievable if there exists a sequence of $(n, R_1, R_2, R_3, P_e^{(n)})$ coding schemes for which $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity region of DM IC-ZIC is the polytope containing all achievable rate tuples, denoted as $\mathcal{C}^{\text{DM}}$.

4.2.2 The MIMO IC-ZIC

A $(M_1, N_1, M_2, N_2, M_3, N_3)$ Gaussian MIMO (multiple-input-multiple-output) IC-ZIC, as shown in Fig. 4.2.2, has M_i antennas at Txi, and N_i antennas at Rxi for each $i \in \{1, 2, 3\}$. Assuming the channel transfer matrices and path attenuations (also referred as channel gains) are time-invariant during each transmission, the input-output relationship of this channel is described by

$$Y_1 = h_{11}H_{11}X_1 + +h_{21}H_{12}X_2 + Z_1 \tag{4.2}$$

$$Y_2 = h_{12}H_{12}X_1 + h_{22}H_{22}X_2 + Z_2 \tag{4.3}$$

$$Y_3 = h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3, \tag{4.4}$$

where $X_i \in \mathbb{C}^{M_i \times 1}$ and $Y_i \in \mathbb{C}^{N_i \times 1}$ are complex input and output vectors, and $H_{ij} \in \mathbb{C}^{N_j \times M_i}$ is the channel transfer matrix from Txi to Rxj whose Frobenius norm satisfies $\|H_{ij}\|_F^2 = 1$. We assume the entries of the transfer matrix H_{ij} are drawn i.i.d. from a continuous and unitarily invariant distribution [45], i.e., $UH_{ij}V$ is identically distributed to H_{ij} for any $U \in \mathbb{U}^{N_i \times N_j}$ and $V \in \mathbb{U}^{M_i \times M_i}$, so H_{ij} has full rank with probability one (w.p.1). The path attenuation h_{ij} from Txi to Rxj is a complex number. The Gaussian noise Z_i are i.i.d. $\mathcal{CN}(\mathbf{0}, I_{N_i})$ across i. Let Cov$[x_{it}]$ be the covariance of the t-th symbol of the transmitted codeword $x_i^n \in \mathcal{X}_i^n$ at Txi. The codeword x_i^n should meet the average per-codeword power constraints

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}(x_{it}x_{it}^{\dagger}) \leq P_i. \tag{4.5}$$

The SNR and INR at receiver Rxi are defined to be

$$\mathsf{SNR}_{ii} = P_i|h_{ii}|^2 \triangleq \rho^{\alpha_{ii}},\ i \in \{1,2,3\} \tag{4.6}$$

$$\mathsf{INR}_{1i} = P_1|h_{1i}|^2 \triangleq \rho^{\alpha_{1i}},\ i \in \{2,3\}, \tag{4.7}$$

where ρ is the nominal SNR based on which the direct channel SNRs and the two cross-channel INRs are defined. The distinct SNR and INR exponents allow us to express the disparities in power levels observed across the direct and cross channels as multiplicative terms associated with the nominal SNR in the dB scale. We denote the capacity region of the MIMO IC-ZIC as $\mathcal{C}$.

4.2.3 The Channel Structure of the MIMO IC-ZIC

We use the generalized singular value decomposition of H_{12} and H_{13} and the singular value decomposition of H_{21} to illustrate the channel structure of the MIMO IC-ZIC. Since channel matrices H_{11}, H_{12}

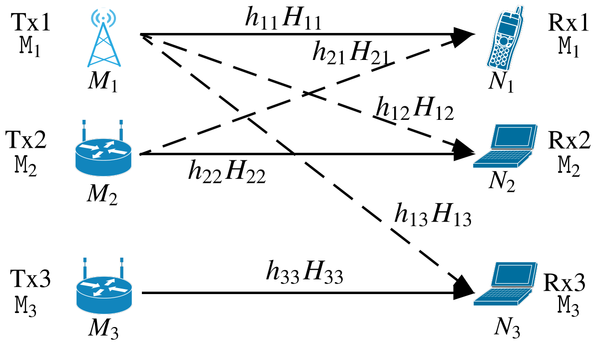

Figure 4.5: The MIMO IC-ZIC

and H_{13} have full rank w.p.1, we have

$$r_{11} \triangleq \text{rank}(H_{11}) = \min\{M_1, N_1\} \tag{4.8}$$

$$r_{12} \triangleq \text{rank}(H_{12}) = \min\{M_1, N_2\} \tag{4.9}$$

$$r_{13} \triangleq \text{rank}(H_{13}) = \min\{M_1, N_3\} \tag{4.10}$$

$$r \triangleq \text{rank}\begin{pmatrix} H_{12} \\ H_{13} \end{pmatrix} = \min\{M_1, N_2 + N_3\} \tag{4.11}$$

The intersection of the two resulting spaces have dimension

$$r_{123} \triangleq r_{12} + r_{13} - r. \tag{4.12}$$

We jointly decompose the two channel matrices H_{12} and H_{13} via the generalized singular value decomposition (GSVD) [35], which is

$$H_{12} = U_{12}\Sigma_{12}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix}^\dagger U^\dagger \triangleq U_{12}\Sigma_{12}V^\dagger \tag{4.13}$$

$$H_{13} = U_{13}\Sigma_{13}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix}^\dagger U^\dagger \triangleq U_{13}\Sigma_{13}V^\dagger. \tag{4.14}$$

$U_{1i} \in \mathbb{U}^{N_i\times N_i}$ and $U \in \mathbb{U}^{M_1\times M_1}$ are unitary matrices. $\Sigma_{1i} \in \mathbb{R}^{N_i\times r}$ is a real and rectangular diagonal matrix. $V_r \in \mathbb{C}^{r\times r}$ is a non-singular upper triangular matrix and $V \triangleq U\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix} \in \mathbb{C}^{M_1\times r}$. Matrices Σ_{12}

and Σ_{13} have the following structure

$$\Sigma_{12} = \begin{array}{c} \\ r-r_{13} \\ r_{123} \\ N_2 - r_{12} \end{array} \begin{array}{c} \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array} \tag{4.15}$$

$$\Sigma_{13} = \begin{array}{c} \\ r_{123} \\ r-r_{12} \\ N_3 - r_{13} \end{array} \begin{array}{c} \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{pmatrix} \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array} \tag{4.16}$$

where C and S are both non-negative real diagonal matrices satisfying $C^2 + S^2 = I$. The acronym N.E. means "never exists". Note $N_2 - r_{12}$ and $r - r_{12}$ cannot be simultaneously positive according to (4.9) and (4.11), and the matrix Σ_{12} is in form of either

$$\Sigma_{12} = \begin{array}{c} \\ r-r_{13} \\ r_{123} \end{array} \begin{array}{c} \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \end{pmatrix} \end{array}$$

or

$$\Sigma_{12} = \begin{array}{c} \\ r-r_{13} \\ r_{123} \\ N_2 - r_{12} \end{array} \begin{array}{c} \begin{array}{cc} r-r_{13} & r_{123} \end{array} \\ \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & C \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}.$$

These two different forms of the matrix Σ_{12} can be unified as the one given by (4.15), where the acronym N.E. comes from the fact that Σ_{12} can only be a 2×3 or 3×2 block matrices. Similarly, $N_3 - r_{13}$ and $r - r_{13}$ cannot be simultaneously positive because of (4.10) and (4.11), and Σ_{13} also has two different forms which can be unified as (4.16).

Define two matrices I_{12} and Λ_{12} to be

$$I_{12} \triangleq \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r-r_{13} \\ r_{123} \\ N_2-r_{12} \end{array} & \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array} \tag{4.17}$$

and

$$\Lambda_{12} \triangleq \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r-r_{13} \\ r_{123} \\ N_2-r_{12} \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \text{N.E.} \end{pmatrix} \end{array} \tag{4.18}$$

respectively, and the matrix Σ_{12} can be written as

$$\Sigma_{12} = I_{12} + \Lambda_{12}.$$

Let the matrices G_{12} and J_{12} be

$$G_{12} \triangleq U_{12}\Lambda_{12}V^{\dagger} \ \text{ and } \ J_{12} \triangleq U_{12}I_{12}V^{\dagger}.$$

Hence, $H_{12} = G_{12} + J_{12}$. Similarly we define two matrices I_{13} and Λ_{13} to be

$$I_{13} \triangleq \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r_{123} \\ r-r_{12} \\ N_3-r_{13} \end{array} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array} \tag{4.19}$$

and

$$\Lambda_{13} \triangleq \begin{array}{c c} & \begin{array}{ccc} r-r_{13} & r_{123} & r-r_{12} \end{array} \\ \begin{array}{c} r_{123} \\ r-r_{12} \\ N_3-r_{13} \end{array} & \begin{pmatrix} \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \text{N.E.} & \mathbf{0} & \mathbf{0} \end{pmatrix} \end{array}, \tag{4.20}$$

respectively, and the matrix Σ_{13} can be written as

$$\Sigma_{13} = I_{13} + \Lambda_{13}.$$

Let the matrices G_{13} and J_{13} be

$$G_{13} \triangleq U_{13}\Lambda_{13}V^{\dagger} \;\text{ and }\; J_{13} \triangleq U_{13}I_{13}V^{\dagger}.$$

Hence, $H_{13} = G_{13} + J_{13}$.

The channel matrix H_{21} from Tx2 to Rx1 can be decomposed with singular value decomposition (SVD) as follows,

$$H_{21} = U_{21}\Sigma_{21}V_{21}^{\dagger}, \tag{4.21}$$

where U_{21} and V_{21} are $N_1 \times N_1$ and $M_2 \times M_2$ unitary matrices respectively and Σ_{21} is a diagonal matrix whose diagonal values consists of the singular values of H_{21}.

Remark 4.1. Through the lens of GSVD, the transmit directions of the interference from Tx1 to Rx2 and Rx3 are established by the column vectors of the matrix V in (4.13) and (4.14). The rectangular diagonal matrix Σ_{1i} determines which components in $V^{\dagger}X_1$ are transferred to Rxi for $i \in \{2,3\}$. The first $r - r_{13}$ components of $V^{\dagger}X_1$ will be heard by Rx2 but not by Rx3, because Σ_{12} has identity block matrix in the upper left, whereas Σ_{13} has all-zero block matrix in the upper left. Similarly, the last $r - r_{12}$ components of $V^{\dagger}X_1$ will be received at Rx3 but not Rx2. The middle r_{123} components of $V^{\dagger}X_1$ will be received by both Rx2 and Rx3, but with different interference strengths (INR_{12} and INR_{13}). Similarly, through the SVD of H_{21}, the transmit directions of the interference from Tx2 to Rx1 are determined by the column vectors of the matrix $V_{21}^{\dagger}$ in (4.21). The diagonal matrix Σ_{21} determines which components of $V_{21}^{\dagger}X_2$ are transferred to Rx1. As pointed out in Remark 3.1, there are a few differences between the SVD and GSVD. In particular, the matrix $V_{21}^{\dagger}$ is unitary therefore its column vectors not only define the transmit signal direction but also form an orthonormal basis of the transmit signal space. However, because the matrix $V^{\dagger}$ is generally not unitary or orthogonal, it can be rank deficient.

Remark 4.2. Note we have assumed INR_{12} and INR_{13} are sufficiently large so that we can disregard the gains contributed by C and S which is justifiable for analysis up to GDoF accuracy.

Next, we express the channel outputs at Rx1-Rx3 in terms of channel side informations. Define the channel side information

$$S_{21} \triangleq h_{21}H_{21}X_2 + Z_1. \tag{4.22}$$

The received signal Y_1 at Rx1 given by 4.2 can be written as the sum of the received intended signal $h_{11}H_{11}X_1$ with the channel side information S_{21}.

To describe the received signal Y_2 and Y_3, we use three channel side informations S_{123}, S_{12} and S_{13} which represent the parts of the interference (and its associated noise) hearable at Rx2-Rx3, Rx2 and Rx3. They are given by (4.23)-(4.25).

$$S_{123} \triangleq \begin{cases} h_{13}G_{13}X_1 + U_{13}\begin{pmatrix} U_{13}^{-1(1:r_{123})}Z_3 \\ \mathbf{0}_{(N_3-r_{123})\times 1} \end{pmatrix} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h_{12}G_{12}X_1 + U_{12}\begin{pmatrix} \mathbf{0}_{(r-r_{13})\times 1} \\ U_{12}^{-1(r-r_{13}+1:r_{12})}Z_2 \\ \mathbf{0}_{(N_2-r_{12})\times 1} \end{pmatrix} & \mathsf{INR}_{12} < \mathsf{iNR}_{13} \end{cases} \tag{4.23}$$

$$S_{12} \triangleq \begin{cases} h_{12}H_{12}X_1 + Z_2 & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h_{12}J_{12}X_1 + U_{12}\begin{pmatrix} U_{12}^{-1(1:r-r_{13})}Z_2 \\ \mathbf{0}_{r_{123}\times 1} \\ U_{12}^{-1(r_{12}+1:N_2)}Z_2 \end{pmatrix} & \mathsf{INR}_{12} < \mathsf{iNR}_{13} \end{cases} \tag{4.24}$$

$$S_{13} \triangleq \begin{cases} h_{13}J_{13}X_1 + U_{13}\begin{pmatrix} \mathbf{0}_{r_{123}\times 1} \\ U_{13}^{-1(r_{123}+1:N_3)}Z_3 \end{pmatrix} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h_{13}H_{13}X_1 + Z_3. & \mathsf{INR}_{12} < \mathsf{iNR}_{13} \end{cases} \tag{4.25}$$

The reason that channel side informations S_{123}, S_{12} and S_{13} depends on the relationship ofINR_{12} and INR_{13} is because the channel gain associated with common part of the interference is determined by the channel gain of the weaker interference receiver. When $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, the channel side informations are the same as these informations in the MIMO one-to-three IC, i.e., (3.25)-(3.27). The side information S_{123} contains the common part of the interference signal $h_{13}G_{13}X_1$ that can be heard by both Rx2 and Rx3 with the associated r_{123} noise elements. There is no S_{123} explicitly in (4.24) because $S_{12} = h_{12}U_{12}I_{12}V^{\dagger}X_1 + h_{12}U_{12}\Lambda_{12}V^{\dagger}X_1 + Z_2$ already contains a scaled and linearly transformed version of the interference signal in S_{123} in the term

$h_{12}U_{12}\Lambda_{12}V^{\dagger}X_1$. The side information S_{13} contains the partial interference signal exclusively for Rx3. S_{12} contains the interference sent along the exclusive signal directions to Rx2 ($h_{12}U_{12}I_{12}V^{\dagger}X_1$) and the amplified version of S_{123} ($h_{12}U_{12}J_{12}V^{\dagger}X_1$) to Rx2. As will be seen in what follows, the GDoF optimal coding scheme incorporates signal direction alignment to utilize the exclusive transmit directions from Tx1 to Rx2 and Rx3 respectively, as well as signal level alignment to adapt the disparity of the interference strengths along the common transmit directions from Tx1 to Rx2-Rx3.

With the channel side informations defined above, the channel input-output relationship (4.2)-(4.4) can be written as

$$Y_1 = h_{11}H_{11}X_1 + S_{21} \tag{4.26}$$

$$Y_2 = \begin{cases} S_{12} + h_{22}H_{22}X_2 & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ S_{123}^{'} + S_{12}^{'} + h_{22}H_{22}X_2 & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.27}$$

$$Y_3 = \begin{cases} S_{123} + S_{13} + h_{33}H_{33}X_3 & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ S_{13}^{'} + h_{33}H_{33}X_3 & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases}. \tag{4.28}$$

The relationship between $V_rV_r^{\dagger}$ and the two scaled identity matrices $\lambda_{\min}^2(V_r)I_r$ and $\lambda_{\max}^2(V_r)I_r$ will be frequently used in the rest of the chapter. We present it in Fact 4.1.

Fact 4.1. *Let V_r be a $r \times r$ full rank square matrix. The following relationship holds between the matrices $V_rV_r^{\dagger}$, $\lambda_{\min}^2(V_r)I_r$ and $\lambda_{\max}^2(V_r)I_r$*

$$\lambda_{\min}^2(V_r)I_r \preceq V_rV_r^{\dagger} \preceq \lambda_{\max}^2(V_r)I_r \tag{4.29}$$

which is equivalent to

$$\frac{V_rV_r^{\dagger}}{\lambda_{\max}^2(V_r)} \preceq I_r \preceq \frac{V_rV_r^{\dagger}}{\lambda_{\min}^2(V_r)}, \tag{4.30}$$

and

$$\lambda_{\max}^{-2}(V_r)I_r \preceq V_r^{-1}V_r^{\dagger-1} \preceq \lambda_{\min}^{-2}(V_r)I_r. \tag{4.31}$$

4.3 Multi-level Superposition Coding Scheme and the Inner Bound for the DM IC-ZIC

The structure of the channel 4.1 suggests a natural coding scheme for the DM IC-ZIC: three level superposition coding at Tx1 as in the DM one-to-three IC (c.f. Section 3.3), two level superposition coding (either HK [23] or CMG type coding [10, 11]) at Tx2 as has been done for the two-user DM IC and single user random coding at Tx3. More specifically, we split the message $\mathtt{M}_1$ at Tx1 into four parts, namely $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$, and perform three level superposition coding to encode them to intermediate codewords W_{123}^n,W_{12}^n, W_{13}^n and W_{1p}^n so that those four messages can be decoded by Rx1-Rx3, Rx1 and Rx2, Rx1 and Rx3 and Rx1 only, respectively. We split the message $\mathtt{M}_2$ at Tx2 into two parts ($\mathtt{M}_{12}$ and $\mathtt{M}_{1p}$) and perform CMG type superposition coding to encode them to intermediate codewords W_{21}^n and W_{2p}^n so that $\mathtt{M}_{21}$ and $\mathtt{M}_{1p}$ can be decoded by Rx1-Rx2 and Rx1 only. The Tx3 simply encodes its entire message using a single user random codebook.

According to the coding scheme introduced above, we define the set of the coding distributions accordingly in Definition 4.2.

Definition 4.2. Let $\mathcal{P}_{\text{in}}$ be the set of distributions P_{in} of joint random variables $(Q, W_{123}, W_{12}, W_{13}, W_{12}, X_1,$ $X_2, X_3)$ that can be factored as

$$p(q, w_{123}, w_{12}, w_{13}, w_{21}, x_1, x_2, x_3) = p(q)p(w_{123}|q)p(w_{12}|w_{123})p(w_{13}|w_{123})p(w_{21}|q) \\ \cdot p(x_1|w_{123}, w_{12}, w_{13})p(x_2|w_{21})p(x_3|q). \tag{4.32}$$

An inner bound is obtained in Theorem 4.1 for any fixed coding distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$ through a detailed joint typicality analysis. For the sake of convenience, we define the relevant set functions in Definition 4.3, in terms of which the inner bound can be written succinctly. Define Θ_i to be the index set of the sub-messages which needs to be decoded at Rxi (intended or non-intended) for each $i \in \{1, 2, 3\}$, i.e.,

$$\Theta_1 \triangleq \{123, 12, 13, 1p, 21\} = \{1, 21\}$$

$$\Theta_2 \triangleq \{123, 12, 21, 2p\} = \{123, 12, 2\}$$

$$\Theta_3 \triangleq \{123, 13, 3\}.$$

Note by the message split scheme introduced in the beginning of this section, we have $\{1\} = \{123, 12, 13, 1p\}$ and $\{2\} = \{21, 2p\}$. Also, define Φ_1, Φ_2 and Φ_3 to be the three sets defined in the following.

$$\Phi_1 \triangleq \{\{1p\}, \{12, 1p\}, \{13, 1p\}, \{12, 13, 1p\}, \{1\}, \{1p, 21\}, \{12, 1p, 21\}, \{13, 1p, 21\}, \{12, 13, 1p, 21\}, \{1, 21\}\} \tag{4.33}$$

$$\Phi_2 \triangleq \{\{2p\}, \{2\}, \{12, 2p\}, \{12, 2\}, \{123, 12, 2p\}, \{123, 12, 2\}\} \tag{4.34}$$

$$\Phi_3 \triangleq \{\{3\}, \{13, 3\}, \{123, 13, 3\}\} \tag{4.35}$$

Let $\mathtt{M}_{\phi_i} = \{\cup_{k\in\phi_i} \mathtt{M}_k\}$ and $W_{\phi_i} = \{\cup_{k\in\phi_i} W_k\}$. For a given P_{in}, the set function $\mathsf{F}_i(\mathtt{M}_{\phi_i})$ takes the sub-message set $\mathtt{M}_{\phi_i} \in \Phi_i$ (for $i \in \{1, 2, 3\}$) as input and returns the mutual information between the set of auxiliary random variables W_{ϕ_i} (which are used to encode the sub-messages in the set $\mathtt{M}_{\phi_i}$) and the received signal Y_i conditioned on $W_{\Theta_i \setminus \phi_i}$ and time sharing variable Q. For example, $\mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{12})$ is the mutual information between W_{1p}, W_{12} (which are used to encode $\mathtt{M}_{1p}$ and $\mathtt{M}_{12}$) and Y_1 conditioned on W_{123}, W_{13}, W_{21} and Q.

Definition 4.3. For a fix distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$, define the following set functions for the DM IC-ZIC inner bound.

$$\mathsf{F}_1(\mathtt{M}_{1p}) \triangleq I(X_1; Y_1 | W_{1c}, W_{21}, Q) \tag{4.36}$$

$$\mathsf{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) \triangleq I(X_1; Y_1 | W_{123}, W_{12}, W_{21}, Q) \tag{4.37}$$

$$\mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) \triangleq I(X_1; Y_1 | W_{123}, W_{13}, W_{21}, Q) \tag{4.38}$$

$$\mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) \triangleq I(X_1; Y_1 | W_{123}, W_{21}, Q) \tag{4.39}$$

$$\mathsf{F}_1(\mathtt{M}_1) \triangleq I(X_1; Y_1 | W_{21}, Q) \tag{4.40}$$

$$\mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq I(X_1, W_{21}; Y_1 | W_{1c}, Q) \tag{4.41}$$

$$\mathsf{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq I(X_1, W_{21}; Y_1 | W_{123}, W_{12}, Q) \tag{4.42}$$

$$\mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq I(X_1, W_{21}; Y_1 | W_{123}, W_{13}, Q) \tag{4.43}$$

$$\mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq I(X_1, W_{21}; Y_1 | W_{123}, Q) \tag{4.44}$$

$$\mathsf{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) \triangleq I(X_1, W_{21}; Y_1 | Q) \tag{4.45}$$

$$\mathsf{F}_2(\mathtt{M}_{2p}) \triangleq I(X_2; Y_2 | W_{123}, W_{12}, W_{21}, Q) \tag{4.46}$$

$$\mathsf{F}_2(\mathtt{M}_2) \triangleq I(X_2;Y_2|W_{123},W_{12},Q) \tag{4.47}$$

$$\mathsf{F}_2(\mathtt{M}_{12},\mathtt{M}_{2p}) \triangleq I(X_2,W_{12};Y_2|W_{123},W_{21},Q) \tag{4.48}$$

$$\mathsf{F}_2(\mathtt{M}_{12},\mathtt{M}_2) \triangleq I(X_2,W_{12};Y_2|W_{123},Q) \tag{4.49}$$

$$\mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) \triangleq I(X_2,W_{123},W_{12};Y_2|W_{21},Q) \tag{4.50}$$

$$\mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) \triangleq I(X_2,W_{123},W_{12};Y_2|Q) \tag{4.51}$$

$$\mathsf{F}_3(\mathtt{M}_3) \triangleq I(X_3;Y_3|W_{123},W_{13},Q) \tag{4.52}$$

$$\mathsf{F}_3(\mathtt{M}_{13},\mathtt{M}_3) \triangleq I(X_3,W_{13};Y_3|W_{123},Q) \tag{4.53}$$

$$\mathsf{F}_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \triangleq I(X_3,W_{123},W_{13};Y_3|Q) \tag{4.54}$$

where for the sake of simplicity we define $W_{1c} \triangleq \{W_{123},W_{12},W_{13}\}$.

The inner bound for the DM IC-ZIC is stated in the theorem below.

Theorem 4.1. *For the DM IC-ZIC and some fixed distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$, the region $\mathcal{R}_{\text{in}}^{\text{DM}}(P_{\text{in}})$ defined by (4.55)-(4.85) is achievable, i.e., $\mathcal{R}_{\text{in}}^{\text{DM}} \subseteq \mathcal{C}^{\text{DM}}$.*

$$\mathcal{R}_{\text{in}}^{\text{DM}} \triangleq \big\{(R_1,R_2,R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \le \mathsf{F}(\mathtt{M}_1) \tag{4.55}$$

$$R_1 \le \mathsf{F}_1(\mathtt{M}_{13},\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) \tag{4.56}$$

$$R_2 \le \mathsf{F}_2(\mathtt{M}_2) \tag{4.57}$$

$$R_2 \le \mathsf{F}_1(\mathtt{M}_{1p},\mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) \tag{4.58}$$

$$R_3 \le \mathsf{F}_3(\mathtt{M}_3) \tag{4.59}$$

$$R_1+R_2 \le \mathsf{F}_1(\mathtt{M}_1,\mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) \tag{4.60}$$

$$R_1+R_2 \le \mathsf{F}_1(\mathtt{M}_{13},\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) \tag{4.61}$$

$$R_1+R_2 \le \mathsf{F}_1(\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p})$$

$$R_1+R_3 \le \mathsf{F}_1(\mathtt{M}_{12},\mathtt{M}_{1p}) + \mathsf{F}_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.62}$$

$$R_1+R_3 \le \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{12},\mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3)$$

$$R_1+R_3 \le \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{13},\mathtt{M}_3) \tag{4.63}$$

$$R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.64}$$

$$R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.65}$$

$$R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.66}$$

$$R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.67}$$

$$R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.68}$$

$$R_1 + 2R_2 \leq \mathsf{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) \tag{4.69}$$

$$2R_1 + R_2 \leq \mathsf{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \tag{4.70}$$

$$2R_1 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.71}$$

$$2R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.72}$$

$$2R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.73}$$

$$2R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.74}$$

$$2R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.75}$$

$$R_1 + 2R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.76}$$

$$R_1 + 2R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.77}$$

$$2R_1 + 2R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.78}$$

$$2R_1 + 2R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.79}$$

$$2R_1 + 2R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.80}$$

$$2R_1 + R_2 + 2R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + 2\mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.81}$$

$$2R_1 + R_2 + 2R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.82}$$

$$3R_1 + R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.83}$$

$$3R_1 + 2R_2 + 2R_3 \leq 2\mathsf{F}_1(\mathtt{M}_{1p}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + 2\mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.84}$$

$$2R_1 + 3R_2 + R_3 \leq \mathsf{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \mathsf{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + 2\mathsf{F}_2(\mathtt{M}_{2p}) + \mathsf{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \mathsf{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)\} \tag{4.85}$$

Proof Outline. We outline the proof here and relegate the full proof to Appendix B.1. As previously stated, Tx1 performs three level superposition coding; Tx2 performs two level superposition coding; and Tx3 performs single user random coding. More specifically, Tx1 splits a message m_1 into four parts m_{123}, m_{12}, m_{13} and m_{1p}. The sub-message m_{123}, which needs to be decoded by Rx1-Rx3, is first encoded to the first level codeword $w_{123}^n(m_{123})$. Then the multicast sub-message m_{1i} is encoded to $w_{1i}^n(m_{1i}, w_{123}^n(m_{123}))$, which needs to be decoded by Rx1 and Rxi for some $i \in \{2,3\}$. This is the second level superposition coding. Finally, based on m_{1p}, which is the private message to be decoded by Rx1, the entire message is encoded to the codeword $x_1(m_{1p}, w_{12}^n(m_{12}, w_{123}^n(m_{123})), w_{13}^n(m_{13}, w_{123}^n(m_{123})))$. Tx2 splits a message m_2 into two parts m_{21} and m_{2p}. The sub-message m_{21} needs to be decoded by Rx2 and Rx1, and it is encoded to the codeword $w_{21}^n(m_{21})$ first. Then based on m_{2p}, which is the private message to be decoded by Rx2, the entire message $\mathtt{M}_2$ is encoded to the codeword $x_2^n(m_{2p}, w_{21}^n(m_{21}))$. Tx3 sends information m_3 via some codeword $x_3^n(m_3)$ using a single-user random codebook, and Rx3 decodes the intended message m_3. Fourier-Motzkin elimination is performed to eliminate the four rate variables associated with the auxiliary random variables W_{123}, W_{12}, W_{13} and W_{21} to obtain the achievable region. □

4.4 Bounds on the Capacity Region for MIMO IC-ZIC

We present single region inner and outer bounds for the MIMO IC-ZIC which are within quantifiable gap (independent of channel SNR/INR) in this section. In Section 4.4.1, we present an **explicit** additive superposition coding scheme for the general MIMIO IC-ZIC with Gaussian codebooks and specified covariance matrices. We then obtain a **single** region inner bound which has the form of a single polytope. In Section 4.4.2, we derive a single region outer bound through the genie aided argument. In Section 4.4, the gap between the inner and outer bounds is then quantified and shown to depend only on the entries of the cross channel matrices H_{12} and H_{13}.

4.4.1 The Inner Bound

A per-distribution achievable region $\mathcal{R}_{\text{in}}(P_{\text{in}})$ for the DM IC-ZIC has been derived in Theorem 4.1. In this subsection, we apply that result to the MIMO IC-ZIC to obtain a single region inner bound in Theorem 4.2 to follow. In particular, to prove Theorem 4.2, we have to specify an explicit coding distribution P_{in} and then compute the set functions in Definition 4.3 for the MIMO setting and for this particular coding distribution. The set functions for the MIMO IC-ZIC inner bound are given in Definition 4.4 and the inner bound itself is stated in Theorem 4.2.

Definition 4.4. Define the set functions listed as (4.86)-(4.104), where the constant β_1, β_2 and β_3 are

$$F_1(\mathtt{M}_{1p}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger\right| - \beta_1 \tag{4.86}$$

$$F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{13} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger\right| - \beta_1 \tag{4.87}$$

$$F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{12} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger\right| - \beta_1 \tag{4.88}$$

$$F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{12} + Q_{13} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger\right| - \beta_1 \tag{4.89}$$

$$F_1(\mathtt{M}_{1}) = \log|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger| - \beta_1 \tag{4.90}$$

$$F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) = \log|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2} H_{21}^\dagger| - \beta_1 \tag{4.91}$$

$$F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{13} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2} H_{21}^\dagger\right| - \beta_1 \tag{4.92}$$

$$F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{12} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2} H_{21}^\dagger\right| - \beta_1 \tag{4.93}$$

$$F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) = \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (Q_{12} + Q_{13} + Q_{1p}) H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2} H_{21}^\dagger\right| - \beta_1 \tag{4.94}$$

$$F_1(\mathtt{M}_{1}, \mathtt{M}_{21}) = \log|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2} H_{21}^\dagger| - \beta_1 \tag{4.95}$$

$$F_2(\mathtt{M}_{2p}) = \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} (Q_{13} + Q_{1p}) H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2p} H_{22}^\dagger\right| - \beta_2 \tag{4.96}$$

$$F_2(\mathtt{M}_{2}) = \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} (Q_{13} + Q_{1p}) H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2} H_{22}^\dagger\right| - \beta_2 \tag{4.97}$$

$$F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) = \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} (Q_{12} + Q_{13} + Q_{1p}) H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2p} H_{22}^\dagger\right| - \beta_2 \tag{4.98}$$

$$F_2(\mathtt{M}_{12}, \mathtt{M}_{2}) = \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} (Q_{12} + Q_{13} + Q_{1p}) H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2} H_{22}^\dagger\right| - \beta_2 \tag{4.99}$$

$$F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) = \log|I_{N_2} + \rho^{\alpha_{12}} H_{12} Q_{1} H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2p} H_{22}^\dagger| - \beta_2 \tag{4.100}$$

$$F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2}) = \log|I_{N_2} + \rho^{\alpha_{12}} H_{12} Q_{1} H_{12}^\dagger + \rho^{\alpha_{22}} H_{22} Q_{2} H_{22}^\dagger| - \beta_2 \tag{4.101}$$

$$F_3(\mathtt{M}_3) = \log \left| I_{N_3} + \rho^{\alpha_{13}} H_{13}(Q_{12} + Q_{1p})H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33}H_{33}^\dagger \right| - \beta_3 \tag{4.102}$$

$$F_3(\mathtt{M}_{13}, \mathtt{M}_3) = \log \left| I_{N_3} + \rho^{\alpha_{13}} H_{13}(Q_{12} + Q_{13} + Q_{1p})H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33}H_{33}^\dagger \right| - \beta_3 \tag{4.103}$$

$$F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) = \log |I_{N_3} + \rho^{\alpha_{13}} H_{13}Q_1H_{13}^\dagger + \frac{\rho^{\alpha_{33}}}{M_3} H_{33}H_{33}^\dagger| - \beta_3 \tag{4.104}$$

$$\beta_1 \triangleq \min\{N_1, M_2\} \log \frac{1 + M_2}{M_2} \tag{4.105}$$

$$\beta_2 \triangleq \log \left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| + r_{123} \log \left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{12})}\right) + (r - r_{13}) \tag{4.106}$$

$$\beta_3 \triangleq \log \left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| + r_{123} \log \left(1 + \frac{\sigma_{\max}^2(\Lambda_{13})}{\sigma_{\min}^2(\Lambda_{13})}\right) + (r - r_{12}), \tag{4.107}$$

$$\zeta_{\min} \triangleq \frac{r}{\lambda_{\max}^2(V_r)} + (M_1 - r)^+$$

$$\zeta_{\max} \triangleq \frac{r}{\lambda_{\min}^2(V_r)} + (M_1 - r)^+$$

and the covariance matrices Q_{1p}, Q_{12}, Q_{13} are given by the restrictions (4.141)-(4.145), and Q_{2p} by (4.149).

Theorem 4.2. *For the MIMO one-to-three IC, the region $\mathcal{R}_{\text{in}}$ defined by* (4.108)-(4.139) *is achievable, i.e., $\mathcal{R}_{\text{in}} \subseteq \mathcal{C}$.*

$$\mathcal{R}_{\text{in}}(F_1, F_2, F_3) \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq F_1(\mathtt{M}_1) \tag{4.108}$$

$$R_1 \leq F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \tag{4.109}$$

$$R_2 \leq F_2(\mathtt{M}_2) \tag{4.110}$$

$$R_2 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) \tag{4.111}$$

$$R_3 \leq F_3(\mathtt{M}_3) \tag{4.112}$$

$$R_1 + R_2 \leq F_1(\mathtt{M}_1, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) \tag{4.113}$$

$$R_1 + R_2 \leq F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) \tag{4.114}$$

$$R_1 + R_2 \leq F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})$$

$$R_1 + R_3 \leq F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.115}$$

$$R_1 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.116}$$

$$R_1 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.117}$$

$$R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.118}$$

$$R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.119}$$

$$R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.120}$$

$$R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.121}$$

$$R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.122}$$

$$R_1 + 2R_2 \leq F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) \tag{4.123}$$

$$2R_1 + R_2 \leq F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + F_1(\mathtt{M}_1, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \tag{4.124}$$

$$2R_1 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.125}$$

$$2R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.126}$$

$$2R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.127}$$

$$2R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_1, \mathtt{M}_{21}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.128}$$

$$2R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_1, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.129}$$

$$R_1 + 2R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.130}$$

$$R_1 + 2R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.131}$$

$$2R_1 + 2R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.132}$$

$$2R_1 + 2R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.133}$$

$$2R_1 + 2R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.134}$$

$$2R_1 + R_2 + 2R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + 2F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.135}$$

$$2R_1 + R_2 + 2R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.136}$$

$$3R_1 + R_2 + R_3 \leq F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_1(\mathtt{M}_1, \mathtt{M}_{21}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.137}$$

$$3R_1 + 2R_2 + 2R_3 \leq 2F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + 2F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.138}$$

$$2R_1 + 3R_2 + R_3 \leq F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + 2F_2(\mathtt{M}_{2p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)\Big\} \tag{4.139}$$

Proof. As stated in the beginning of this subsection, we apply the inner bound in Theorem 4.1 for the DM IC-ZIC to derive the single region inner bound for the MIMIO IC-ZIC. We explicitly pick one coding distribution from the set of distributions defined in Definition 4.2 which adapts the MIMO IC-ZIC channel parameters, allowing for auxiliary and input random vectors over continuous alphabets for the MIMO IC-ZIC. First, time sharing is disabled and all the transmitters use full power. The coding scheme is motivated from the three level superposition coding scheme for the MIMO one-to-three IC (c.f. Section 3.4.2) and Karmakar-Varanasi coding scheme for the MIMO two-user IC. More specifically, since Tx1 produces one-sided interferences to both Rx2-Rx3, we employ the same three-level superposition coding at Tx1 as in the MIMO one-to-three IC (c.f. Section 3.4.2). Let $W_{123} \sim \mathcal{CN}(\mathbf{0}, Q_{123})$, $W_{12} \sim \mathcal{CN}(\mathbf{0}, Q_{12})$, $W_{13} \sim \mathcal{CN}(\mathbf{0}, Q_{13})$ and $W_{1p} \sim \mathcal{CN}(\mathbf{0}, Q_{1p})$ be four independent Gaussian random vectors to encode $\mathtt{M}_{123}$, $\mathtt{M}_{12}$, $\mathtt{M}_{13}$ and $\mathtt{M}_{1p}$. The transmitted signal X_1 is the direct sum of W_{123}, W_{12}, W_{13} and W_{1p} scaled by the transmit power, i.e.,

$$X_1 = \sqrt{P_1}(W_{123} + W_{12} + W_{13} + W_{1p}).$$

Define

$$V_p \triangleq U^{\dagger-1} \begin{pmatrix} V_r^{\dagger-1} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+} \end{pmatrix} \tag{4.140}$$

as a linear precoding matrix, and let the covariance matrices Q_{123}, Q_{12}, Q_{13} and Q_{1p} satisfy the restrictions given by (4.141)-(4.153).

$$Q_{1p}$$

$$= \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} + \rho^{\alpha_{13}}\Sigma_{13}^\dagger\Sigma_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \tag{4.141}$$

$$Q_{12} + Q_{1p}$$

$$
= \begin{cases} \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} \geq \mathsf{INR}_{13} \\ \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}} \Lambda_{12}^\dagger \Lambda_{12} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} < \mathsf{INR}_{13} \end{cases} \tag{4.142}
$$

$$
Q_{13} + Q_{1p}
$$

$$
= \begin{cases} \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13} + \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} \geq \mathsf{INR}_{13} \\ \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} < \mathsf{INR}_{13} \end{cases} \tag{4.143}
$$

$$
Q_{12} + Q_{13} + Q_{1p}
$$

$$
= \begin{cases} \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} \geq \mathsf{INR}_{13} \\ \frac{V_p}{\mathrm{Tr}(V_p V_p^\dagger)} \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}} \Lambda_{12}^\dagger \Lambda_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger & \mathsf{INR}_{21} < \mathsf{INR}_{13} \end{cases} \tag{4.144}
$$

$$
Q_1
$$

$$
\triangleq Q_{12} + Q_{13} + Q_{1p} + Q_{123} = \frac{V_p V_p^\dagger}{\mathrm{Tr}(V_p V_p^\dagger)}. \tag{4.145}
$$

This coding scheme was shown to be GDoF optimal for the MIMO one-to-three IC in Section 3.4.4, a special case of the MIMO IC-ZIC when the cross link channel gain $h_{21} = 0$. This three level superposition coding ensures that the contributions of W_{1p} at Rx2 and Rx3 have covariances that satisfy

$$
\rho^{\alpha_{12}} H_{12} Q_{1p} H_{12}^\dagger \preceq I_{N_2} \text{ and } \rho^{\alpha_{13}} H_{13} Q_{1p} H_{13}^\dagger \preceq I_{N_3}.
$$

Therefore, W_{1p} arrives at the Rx2 and Rx3 under the noise floor. Hence, W_{1p} is used to encode $\mathtt{M}_{1p}$, the private message to Rx1. The contribution of $W_{1p} + W_{13}$ at Rx2 has covariance which satisfies

$$\rho^{\alpha_{12}} H_{12}(Q_{12} + Q_{1p})H_{12}^{\dagger} \preceq I_{N_2},$$

and the contribution of $W_{1p} + W_{12}$ at Rx3 has covariance which satisfies

$$\rho^{\alpha_{13}} H_{13}(Q_{12} + Q_{1p})H_{13}^{\dagger} \preceq I_{N_3}.$$

Hence, W_{12} and W_{13} are the auxiliary random vectors to encode $\mathtt{M}_{12}$ and $\mathtt{M}_{13}$, respectively. The random vector W_{123} is received above the noise floor at all three receivers; therefore, it is used to encode the sub-message $\mathtt{M}_{123}$ to be decoded by Rx1-Rx3. Besides, it is not difficult to see that (4.144) results from adding the left and right hand sides of (4.143) and (4.142) and subtracting from that result the left and right hand sides of (4.141).

Using the inequality (4.29), the trace $\mathrm{Tr}(V_p V_p^{\dagger})$ can be upper and lower bounded as

$$\mathrm{Tr}(V_p V_p^{\dagger}) = \mathrm{Tr}(V_p^{\dagger} V_p) \geq \frac{r}{\lambda_{\max}^2(V_r)} + (M_1 - r)^+ = \zeta_{\min} \tag{4.146}$$

and

$$\mathrm{Tr}(V_p V_p^{\dagger}) = \mathrm{Tr}(V_p^{\dagger} V_p) \leq \frac{r}{\lambda_{\min}^2(V_r)} + (M_1 - r)^+ = \zeta_{\max}, \tag{4.147}$$

respectively.

Since Tx2 only interferes Rx1, we let Tx2 perform Karmakar-Varanasi type coding scheme [27], where the transmitted signal X_2 is the sum of the auxiliary random vectors $W_{21} \sim \mathcal{CN}(\mathbf{0}, Q_{21})$ and $W_{2p} \sim \mathcal{CN}(\mathbf{0}, Q_{2p})$ scaled by the transmit power, i.e.

$$X_2 = \sqrt{P_2}\,(W_{21} + W_{2p}).$$

The random vectors W_{21} and W_{2p} are used to encode $\mathtt{M}_{21}$ and $\mathtt{M}_{2p}$, respectively. The entire signal is transmitted with full power and with covariance

$$Q_2 \triangleq \mathrm{Cov}[X_2] = \frac{1}{M_2} I_{M_2}. \tag{4.148}$$

The covariance matrices Q_{21} and Q_{2p} are chosen as follows.

$$Q_{2p} = \frac{1}{M_2}(I_{M_2} + \rho^{\alpha_{21}} H_{21}^{\dagger} H_{21})^{-1} \tag{4.149}$$

$$Q_{21} = \frac{1}{M_2} I_{M_2} - Q_{2p} \tag{4.150}$$

This coding scheme was first proposed in [26] for the two-user MIMO interference channel. It was shown therein that the resulting rate region is within constant gap to the capacity region of the MIMO two-user interference channel [27, Theorem 2].

Lastly, we let Tx3 perform single user coding with Gaussian random codebook to encode the intended message $\mathtt{M}_3$ directly to the transmitter signal X_3^n at full power, i.e.

$$\mathrm{Cov}[X_3] = \frac{P_3}{M_3} I_{M_3}.$$

Note there is no water-filling at Tx3, and the total transmit power is uniformly and independently allocated among all transmit antennas. The earlier works in [39] and Section 5.3.1 already pointed out that the scaled identity matrix is sufficient to achieve constant-gap-to-capacity region for the MIMO MAC (hence also for MIMO P2P channel). For the purpose of deriving GDoF region for the MIMO IC-ZIC, water-filling for Tx3 turns out to be unnecessary, which will be shown in the gap result in Section 4.4.3.

With the distributions for the inputs specified this way, we are now ready to obtain the inner bound of Theorem 4.2 from Theorem 4.1. Please refer to Appendix (B.2) for the evaluation of the set functions in $\mathcal{R}_{\text{in}}$ for the MIMO IC-ZIC. $\square$

Let us take a deeper look at the coding scheme. We temporarily assume $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$ before the end of this subsection. From the restrictions (4.141)-(4.145), the individual covariance matrices Q_{12}, Q_{13} and Q_{123} can be obtained as (4.151)-(4.153).

$$\begin{aligned} Q_{12} = & \frac{V_p}{\mathrm{Tr}(V_p V_p^{\dagger})} \\ & \cdot \begin{pmatrix} \frac{\rho^{\alpha_{12}}}{1+\rho^{\alpha_{12}}} I_{12}^{\dagger} I_{12} + (I_r + \rho^{\alpha_{13}} \Lambda_{13}^{\dagger} \Lambda_{13})^{-1} - (I_r + \rho^{\alpha_{12}} \Lambda_{12}^{\dagger} \Lambda_{12} + \rho^{\alpha_{13}} \Lambda_{13}^{\dagger} \Lambda_{13})^{-1} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} \\ & \cdot V_p^{\dagger} \end{aligned} \tag{4.151}$$

$$Q_{13} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \begin{pmatrix} \frac{\rho^{\alpha_{13}}}{1+\rho^{\alpha_{13}}} I_{13}^\dagger I_{13} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} V_p^\dagger \tag{4.152}$$

$$Q_{123} = \frac{V_pV_p^\dagger}{\mathrm{Tr}(V_pV_p^\dagger)} - Q_{1p} - Q_{12} - Q_{13} = \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)} \begin{pmatrix} I_r - (I_r + \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13})^{-1} & \\ & \mathbf{0}_{(M_1-r)\times(M_1-r)} \end{pmatrix} V_p^\dagger \tag{4.153}$$

Let $\mathtt{X}_{123}$, $\mathtt{X}_{12}$, $\mathtt{X}_{13}$ and $\mathtt{X}_{1p}$ be zero mean Gaussian vectors with identity covariance matrices, of length r_{123}, r_{12}, r_{13} and M_1, respectively. With their chosen covariance matrices, the auxiliary random vectors W_{123}, W_{12}, W_{13} and W_{1p} can be alternatively written as follows.

$$W_{123} = \sum_{k=1}^{r_{123}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{13}}\lambda_{13,r-r_{13}+k}^2}} V_p^{[r-r_{13}+k]} \mathtt{X}_{123}^{(k)} \tag{4.154}$$

$$\begin{aligned} W_{12} = &\sum_{k=1}^{r-r_{13}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} V_p^{[k]} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{12}}}} \mathtt{X}_{12}^{(k)} \\ &+ \sum_{k=r-r_{13}+1}^{r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{13}}\lambda_{13,k}^2} - \frac{1}{1+\rho^{\alpha_{12}}\lambda_{12,k}^2 + \rho^{\alpha_{13}}\lambda_{13,k}^2}} V_p^{[k]} \mathtt{X}_{12}^{(k)} \end{aligned} \tag{4.155}$$

$$W_{13} = \sum_{k=1}^{r-r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} V_p^{[r_{12}+k]} \sqrt{1 - \frac{1}{1+\rho^{\alpha_{13}}}} \mathtt{X}_{13}^{(k)} \tag{4.156}$$

$$\begin{aligned} W_{1p} = &\sum_{k=1}^{r-r_{13}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{12}}}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} + \sum_{k=r-r_{13}+1}^{r_{12}} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{12}}\lambda_{12,k}^2 + \rho^{\alpha_{13}}\lambda_{13,k}^2}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} \\ &+ \sum_{k=r_{12}+1}^{r} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} \sqrt{\frac{1}{1+\rho^{\alpha_{13}}}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} + \sum_{k=r+1}^{M} \sqrt{\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}} V_p^{[k]} \mathtt{X}_{1p}^{(k)} \end{aligned} \tag{4.157}$$

In the aforementioned equation, $\mathtt{X}_{123}^{(k)}$ denotes the k-th data stream to Rx1-Rx3 that carries the public sub-message (k-th public data stream) along the transmit direction $V_p^{[r-r_{13}+k]}$. It needs to be decoded by all three receivers. Similarly, $\mathtt{X}_{12}^{(k)}$ is the k-th public data stream to Rx1-Rx2 along the transmit direction $V_p^{[k]}$. The data streams $\mathtt{X}_{12}$ can be divided into two groups. The first $r - r_{13}$ data streams are sent at power level ρ^0, and they are received by Rx1 and Rx2, but not Rx3, since they are sent along the null space of $\langle H_{13}\rangle$. The other r_{123} data streams will be sent at power level $\rho^{-\alpha_{13}}$ so they are decodable at Rx1 and Rx2, but not Rx3, since they arrive under the noise floor at Rx3. The data streams $\mathtt{X}_{13}$ are received by Rx1 and Rx3, but not Rx2, as they are sent along the null space of $\langle H_{12}\rangle$. The data streams $\mathtt{X}_{1p}$ are received by Rx1 only, as a result of which we call them private data streams. The first $r - r_{13}$ private data streams are hearable by Rx2, but not Rx3; therefore, they are transmitted at power level $\rho^{-\alpha_{12}}$ so they arrive at Rx2 under the noise floor. The data streams $\mathtt{X}_{1p}^{(r-r_{13}+1)}, \cdots, \mathtt{X}_{1p}^{(r_{12})}$ are hearable by both Rx2 and Rx3, and they are sent at power level $\rho^{-\alpha_{12}}$ so they arrive at Rx2 and Rx3 under the noise floor. The next $r - r_{12}$ private data streams $\mathtt{X}_{1p}^{(r_{12}+1)}, \cdots, \mathtt{X}_{1p}^{(r)}$ are hearable by Rx3, but not Rx2, and they are sent at power level $\rho^{-\alpha_{13}}$ so they arrive below the noise floor at Rx3. Lastly, when there are more transmit antennas at Tx1 than the sum of receiver antennas at Rx2 and Rx3, the precoding matrix V_p lets $M_1 - r$ private data streams (the last part on the right hand of (4.157)) transmit along the null space of $\left\langle \begin{array}{c} H_{12} \\ H_{13} \end{array} \right\rangle$, and these data streams are exclusively hearable by Rx1; thus, these private data streams are sent at power level ρ^0.

Similarly, let $\mathtt{X}_{21}$ and $\mathtt{X}_{2p}$ be zero mean Gaussian vectors with identity covariance matrices, of length $\min\{M_2, N_1\}$ and M_2 respectively. According to their chosen covariance matrices, the auxiliary random vectors W_{21} and W_{2p} can be alternatively written as

$$W_{21} = \sum_{k=1}^{\min\{M_2,N_1\}} V_{21}^{[k]} \sqrt{\frac{P_2}{M_2}} \sqrt{\frac{\rho^{\alpha_{21}}\sigma_{21,k}^2}{1+\rho^{\alpha_{21}}\sigma_{21,k}^2}} \mathtt{X}_{21}^{(k)} \tag{4.158}$$

$$W_{2p} = \sum_{k=1}^{\min\{M_2,N_1\}} V_{21}^{[k]} \sqrt{\frac{P_2}{M_2}} \sqrt{\frac{1}{1+\rho^{\alpha_{21}}\sigma_{21,k}^2}} \mathtt{X}_{2p}^{(k)} + \sum_{k=\min\{M_2,N_1\}+1}^{M_2} V_{21}^{[k]} \sqrt{\frac{P_2}{M_2}} \mathtt{X}_{2p}^{(k)}. \tag{4.159}$$

The public data streams $\mathtt{X}_{21}$ need to be decoded by both Rx1 and Rx2, and the private data streams $\mathtt{X}_{2p}$ are decoded by Rx2 only. The first $\min\{M_2, N_1\}$ private data streams in $\mathtt{X}_{2p}$ are sent at power level $\rho^{-\alpha_{21}}$ and are received under the noise floor at Rx1. These data streams are the first part in the right hand side of (4.159). When there are more transmit antennas at Tx2 than receive antennas at Rx1, there are $M_2 - N_2$ data streams sent along the null space of $\langle H_{21} \rangle$ causing no interference to Rx1; therefore, these data streams are sent at power level ρ^0 which are the second part in the right hand side of (4.159). The non-interfering signal X_3 can also be rewritten in terms of independent data streams

$$X_3 = \sum_{k=1}^{\min\{M_3,N_3\}} \sqrt{\frac{P_3}{M_3}} \mathtt{X}_3^{(k)}, \tag{4.160}$$

and it is only to be received and decoded by Rx3.

4.4.2 The Outer Bound

We derive a single region outer bound for the MIMO IC-ZIC in this subsection. We provide various genie informations to Rxi to produce upper bounds on R_i in several different forms, and then linearly combine those upper bounds across $i \in \{1,2,3\}$ to obtain sum rate upper bounds. To present the outer bound, we define relevant set functions in Definition 4.5. The outer bound itself is stated in Theorem 4.3.

Definition 4.5. Define the following matrices given by (4.161)-(4.164)

$$K_{1p} \triangleq \left(I_{M_1} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13}\right)^{-1} \tag{4.161}$$

$$K_{12,1p} \triangleq \begin{cases} \left(I_{M_1} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13}\right)^{-1} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ \left(I_{M_1} + \rho^{\alpha_{12}} G_{12}^\dagger G_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13}\right)^{-1} & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.162}$$

$$K_{13,1p} \triangleq \begin{cases} \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^\dagger G_{13} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12}\right)^{-1} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ \left(I_{M_1} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12}\right)^{-1} & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.163}$$

$$K_{12,13,1p} \triangleq \begin{cases} \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^\dagger G_{13}\right)^{-1} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ \left(I_{M_1} + \rho^{\alpha_{12}} G_{12}^\dagger G_{12}\right)^{-1} & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.164}$$

$$K_{2p} \triangleq \left(I_{M_1} + \rho^{\alpha_{21}} H_{21}^\dagger H_{21}\right)^{-1} \tag{4.165}$$

and the set functions listed in (4.166)-(4.184).

$$\overline{F}_1(\mathtt{M}_{1p}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^\dagger) \tag{4.166}$$

$$\overline{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^\dagger) \tag{4.167}$$

$$\overline{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,1p} H_{11}^\dagger) \tag{4.168}$$

$$\overline{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,13,1p} H_{11}^\dagger) \tag{4.169}$$

$$\overline{F}_1(\mathtt{M}_1) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger) \tag{4.170}$$

$$\overline{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger) \tag{4.171}$$

$$\overline{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger) \tag{4.172}$$

$$\overline{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger) \tag{4.173}$$

$$\overline{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,13,1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger) \tag{4.174}$$

$$\overline{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) \triangleq \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger) \tag{4.175}$$

$$\overline{F}_2(\mathtt{M}_{2p}) \triangleq \log(I_{N_2} + \rho^{\alpha_{22}} H_{22} K_{2p} H_{22}) \tag{4.176}$$

$$\overline{F}_2(\mathtt{M}_2) \triangleq \log(I_{N_2} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}) \tag{4.177}$$

$$\overline{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) \triangleq \log(I_{N_2} + \rho^{\alpha_{12}} H_{12} K_{12,13,1p} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} K_{2p} H_{22}^{\dagger}) \tag{4.178}$$

$$\overline{F}_2(\mathtt{M}_{12}, \mathtt{M}_2) \triangleq \log(I_{N_2} + \rho^{\alpha_{12}} H_{12} K_{12,13,1p} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}) \tag{4.179}$$

$$\overline{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \triangleq \log(I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} K_{2p} H_{22}^{\dagger}) \tag{4.180}$$

$$\overline{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) \triangleq \log(I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}) \tag{4.181}$$

$$\overline{F}_3(\mathtt{M}_3) \triangleq \log(I_{N_3} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}) \tag{4.182}$$

$$\overline{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \triangleq \log(I_{N_3} + \rho^{\alpha_{13}} H_{13} K_{12,13,1p} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}) \tag{4.183}$$

$$\overline{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \triangleq \log(I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}) \tag{4.184}$$

Theorem 4.3. *For the MIMO IC-ZIC, let*

$$\eta \triangleq \begin{cases} \log\left|\max\left\{\lambda_{\max}^2(V_r), 1\right\}\right| & \\ \qquad + r_{123}\log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{13})}{\sigma_{\min}^2(\Lambda_{12})}\right) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ \log\left|\max\left\{\lambda_{\max}^2(V_r), 1\right\}\right| & \\ \qquad + r_{123}\log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{13})}\right) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases}, \tag{4.185}$$

and the capacity region $\mathcal{C}$ is contained in $\mathcal{R}_{\mathrm{o}}$ which is defined by inequalities (4.186) (4.213)*, i.e., $\mathcal{C} \subseteq \mathcal{R}_{\mathrm{o}}$.*

$$\mathcal{R}_{\mathrm{o}}(\bar{F}_1, \bar{F}_2, \bar{F}_3) \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq \bar{F}_1(\mathtt{M}_1) \tag{4.186}$$

$$R_2 \leq \bar{F}_2(\mathtt{M}_2) \tag{4.187}$$

$$R_3 \leq \bar{F}_3(\mathtt{M}_3) \tag{4.188}$$

$$R_1 + R_2 \leq \bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) \tag{4.189}$$

$$R_1 + R_2 \leq \bar{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \eta \tag{4.190}$$

$$R_1 + R_2 \leq \bar{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \eta \tag{4.191}$$

$$R_1 + R_3 \leq \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.192}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.193}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.194}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.195}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.196}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.197}$$

$$R_1 + 2R_2 \leq \bar{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \eta \tag{4.198}$$

$$2R_1 + R_2 \leq \bar{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \eta \tag{4.199}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.200}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.201}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.202}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.203}$$

$$R_1 + 2R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.204}$$

$$R_1 + 2R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.205}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.206}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.207}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.208}$$

$$2R_1 + R_2 + 2R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + 2\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.209}$$

$$2R_1 + R_2 + 2R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.210}$$

$$3R_1 + R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{4.211}$$

$$3R_1 + 2R_2 + 2R_3 \leq 2\bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + 2\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$$

$$+\eta \tag{4.212}$$

$$2R_1 + 3R_2 + R_3 \leq \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + 2\bar{F}_2(\mathtt{M}_{2p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$$

$$+\eta\} \tag{4.213}$$

Proof Outline. The fundamental principle in the proof of the outer bound is to construct virtual channels whose outputs are then regarded as genie informations to each receiver to decode its intended signal (and therefore making the receiver more interference-resilient). We construct three genie informations (T_{123}, T_{12} and T_{13}) which are identically distributed as the channel side informations S_{123}, S_{12} and S_{13}, respectively, but each pair of corresponding "T" and "S" random variables (with the same subscripts) are independent conditioned on X_1. The upper bound is proved in three steps. First, by providing one or more of those genie informations to Rxi, $i \in \{1, 2, 3\}$, we derive a series of individual upper bounds on R_i. Some of the bounds may contain entropy terms which cannot be single-letterized. There is one set of individual upper bounds for R_1, but there are two sets of individual upper bounds for R_2 (and also R_3) for the two cases $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$ and $\mathsf{INR}_{12} < \mathsf{INR}_{13}$. Secondly, we linearly combine those individual upper bounds across $i \in \{1, 2, 3\}$ to obtain sum rate upper bounds with unsingle-letterized entropy terms vanished. We then get two intermediate outer bounds in terms of channel side and genie information symbols for the cases $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$ and $\mathsf{INR}_{12} < \mathsf{INR}_{13}$. We unify these two outer bounds into one bound. This bound is a union of polytopes over all admissible input distributions. Finally, we optimize the input distributions in the context of MIMO setting and plug in the optimized distribution to obtain a single region output bound in terms of the channel parameters. Detailed proof is relegated to Appendix B.3. □

4.4.3 Quantifiable Gap

An achievable rate region of a MIMO IC-ZIC is within gap (n_1, n_2, n_3) to its capacity if for any given rate tuple $(R_1, R_2, R_3) \in \mathcal{C}$, the rate tuple $(R_1 - n_1, R_2 - n_2, R_3 - n_3)$ is within that achievable region. We call the tuple n_i the individual gap on R_i. Since we do not know the capacity region $\mathcal{C}$, we quantify the gap between the inner bound $\mathcal{R}_{\text{in}}$ and the outer bound $\mathcal{R}_{\text{o}}$, and the resulting gap will be an upper bound of the gap between $\mathcal{R}_{\text{in}}$ and $\mathcal{C}$. The main result in this subsection is stated in Theorem (4.4).

Theorem 4.4. *Define the following constants.*

$$\gamma_{11} \triangleq \min\{M_1, N_1\} \log\left(\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}\right)^+ \tag{4.214}$$

$$\gamma_{12} \triangleq \min\{M_1, N_2\} \log\left(\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}\right)^+ \tag{4.215}$$

$$\gamma_{21} \triangleq \min\{M_2, N_1\} \log M_2 \tag{4.216}$$

$$\gamma_{22} \triangleq \min\{M_2, N_2\} \log M_2 \tag{4.217}$$

$$\delta_1 \triangleq \min\{M_1 + M_2, N_1\} \log \max\left\{\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}, M_2\right\} \tag{4.218}$$

$$\delta_2 \triangleq \min\{M_1 + M_2, N_2\} t \log \max\left\{\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}, M_2\right\} \tag{4.219}$$

$$\delta_3 \triangleq \min\{M_1 + M_3, N_3\} \log \max\left\{\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}, M_3\right\}. \tag{4.220}$$

$$n^{(1)} \triangleq (n_1^{(1)}, n_2^{(1)}, n_3^{(1)}) \triangleq (\beta_1 + \delta_1 + \eta, \beta_2 + \delta_2, \beta_3 + \delta_3)$$

$$n^{(2)} \triangleq (n_1^{(2)}, n_2^{(2)}, n_3^{(2)}) \triangleq (\beta_1 + \beta_2 + \gamma_{11} + \gamma_{12} + \eta, \beta_1 + \beta_2 + \gamma_{21} + \gamma_{22}, \beta_3 + \delta_3)$$

For any $(R_1, R_2, R_3) \in \mathcal{R}_{\mathrm{o}}$, *let*

$$(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3) = \left(\left(R_1 - \max\{n_1^{(1)}, n_1^{(2)}\}\right)^+, \left(R_2 - \max\{n_2^{(1)}, n_2^{(2)}\}\right)^+, (R_3 - \beta_3 - \delta_3))^+\right),$$

then we have

$$(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3) \in \mathcal{R}_{\mathrm{in}}.$$

Proof Outline. There are 33 inequalities in $\mathcal{R}_{\mathrm{in}}$. Taking out (4.109), (4.111), (4.116), (4.117) and (4.125) from $\mathcal{R}_{\mathrm{in}}$, there is a one-to-one correspondence between the rest of the 28 inequalities in $\mathcal{R}_{\mathrm{in}}$ and the 28 inequalities in $\mathcal{R}_{\mathrm{o}}$. More specifically, the k-th inequality in the rest of the 28 inequalities of $\mathcal{R}_{\mathrm{in}}$ and the k-th inequality in $\mathcal{R}_{\mathrm{o}}$ differ by the set function name ($F_i(\cdot)$ and $\bar{F}_i(\cdot)$) and a constant η (excepting the first three inequalities, which do not have η). To demonstrate the gap, we first quantify the gap between the 28 inequalities in $\mathcal{R}_{\mathrm{in}}$ and $\mathcal{R}_{\mathrm{o}}$ as $(n_1^{(1)}, n_2^{(1)}, n_3^{(2)})$. Then we quantify the five gaps from (4.109) to (4.186), (4.111) to (4.187), (4.116) to (4.192), (4.117) to (4.192) and (4.125) to (4.186)+(4.192). We choose another gap tuple $(n_1^{(2)}, n_2^{(2)}, n_3^{(2)})$ to settle these five gaps. The overall individual gap n_i is determined as $\max(n_i^{(1)}, n_i^{(2)})$. Please refer to Appendix B.4 for the detailed proof. $\square$

4.5 The GDoF Region of the MIMO IC-ZIC

The generalized degrees of freedom (GDoF) is an information-theoretic performance metric that characterizes the number of independent data streams a network could support simultaneously among all users at high SNR regime. In this section, we first compute the GDoF region of the MIMO IC-ZIC, and then focus on the achievability of the key corner points in the GDoF region and the sum GDoF curve in various numerical examples. In what follows, we define $\bar{\alpha} = \{\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}, \alpha_{21}\}$.

4.5.1 GDoF Region

The definition of GDoF region of the MIMO IC-ZIC is given in Definition 4.6.

Definition 4.6. The *generalized degrees of freedom* region of a MIMO IC-ZIC $\mathcal{D}(\bar{\alpha}) \in \mathbb{R}_+^3$ with a capacity region $\mathcal{C}(\bar{\alpha})$ is defined as

$$\left\{(d_1, d_2, d_3) : d_i = \lim_{\rho\to\infty} \frac{R_i}{\log\rho}, i \in \{1,2,3\} \text{ and } (R_1, R_2, R_3) \in \mathcal{C}(\bar{\alpha})\right\}. \tag{4.221}$$

In the rest of the chapter, we call (d_1, d_2, d_3) a GDoF tuple. To compute the GDoF region in this section, we need a slightly different version of Lemma 5.1 which is stated in Fact 4.2. They differ in that the matrices $H_1,\ H_2,\ \cdots,\ H_n$ only need to be full rank w.p.1 here, whereas the entries of the matrices in Lemma 5.1 are drawn i.i.d. from a continuous unitarily invariant distribution. Fact 4.2 can be proved with similar mathematical induction as in the proof of Lemma 5.1.

Fact 4.2. *Let $H_1 \in \mathbb{C}^{u\times u_i}$, $H_2 \in \mathbb{C}^{u\times u_2}$, $\cdots$, $H_n \in \mathbb{C}^{u\times u_n}$ be n full rank matrices (w.p.1) such that $H = [H_1, H_2, \cdots, H_n]$ is also full rank w.p.1. Then, for asymptotic ρ*

$$\log\det\left(I_u + \sum_{i=1}^{n} \rho^{a_i} H_i H_i^\dagger\right) = g(u, (a_1, u_1), \cdots, (a_n, u_n)) \log(\rho) + \mathcal{O}(1) \tag{4.222}$$

where for any $(u, u_1, \cdots, u_n) \in \mathbb{Z}^{+(n+1)}$ and $(a_1, \cdots, a_n) \in \mathbb{R}^n$, the function $g(u, (a_1, u_1), \cdots, (a_n, u_n))$ is defined as

$$\begin{aligned} &g(u, (a_1, u_1), (a_2, u_2), \cdots, (a_n, u_n)) \\ &= \sum_{i=i_1}^{i_n} \left\{ \min\{u, u_{i_1}\} a_{i_1}^+ + \min\left\{(u - u_{i_1})^+, u_{i_2}\right\} a_{i_2}^+ + \cdots + \min\left\{\left(u - \sum_{j=1}^{i_{n-1}} u_j\right)^+, u_{i_n}\right\} a_{i_n}^+ \right\} \end{aligned}$$

for $i_1 \neq i_2 \neq \cdots \neq i_n \in \{1, \cdots, n\}$ *such that* $a_{i_1} \geq a_{i_2} \geq \cdots \geq a_{i_n}$.

To present the GDoF region, we need to define the relevant set functions in Definition 4.7. The GDoF region is stated in Theorem 4.5.

Definition 4.7. Define the set functions given by (4.223)-(4.241).

$$f_1(\mathtt{M}_{1p})$$

$$\triangleq \begin{cases} g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r_{12}), ((\alpha_{11}-\alpha_{13})^+, r-r_{12}), (\alpha_{11}, M_1-r)\right) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r-r_{13}), ((\alpha_{11}-\alpha_{13})^+, r_{13}), (\alpha_{11}, M_1-r)\right) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.223}$$

$$f_1(\mathtt{M}_{13}, \mathtt{M}_{1p})$$

$$\triangleq g\left(N_1, \left((\alpha_{11}-\alpha_{12})^+, r_{12}\right), (\alpha_{11}, M_1 - r_{12})\right) \tag{4.224}$$

$$f_1(\mathtt{M}_{12}, \mathtt{M}_{1p})$$

$$\triangleq g\left(N_1, \left((\alpha_{11}-\alpha_{13})^+, r_{13}\right), (\alpha_{11}, M_1 - r_{13})\right) \tag{4.225}$$

$$f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p})$$

$$\triangleq \begin{cases} g\left(N_1, ((\alpha_{11}-\alpha_{13})^+, r_{123}), (\alpha_{11}, M_1-r_{123})\right) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r_{123}), (\alpha_{11}, M_1-r_{123})\right) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.226}$$

$$f_1(\mathtt{M}_1)$$

$$\triangleq \min\{M_1, N_1\}\alpha_{11} \tag{4.227}$$

$$f_1(\mathtt{M}_{1p}, \mathtt{M}_{21})$$

$$\triangleq \begin{cases} g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r_{12}), ((\alpha_{11}-\alpha_{13})^+, r-r_{12}), (\alpha_{11}, M_1-r), (\alpha_{21}, M_2)\right) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r-r_{13}), ((\alpha_{11}-\alpha_{13})^+, r_{13}), (\alpha_{11}, M_1-r), (\alpha_{21}, M_2)\right) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.228}$$

$$f_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21})$$

$$\triangleq g\left(N_1, \left((\alpha_{11}-\alpha_{12})^+, r_{12}\right), (\alpha_{11}, M_1 - r_{12}), (\alpha_{21}, M_2)\right) \tag{4.229}$$

$$f_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21})$$

$$\triangleq g\left(N_1, \left((\alpha_{11}-\alpha_{13})^+, r_{13}\right), (\alpha_{11}, M_1 - r_{13}), (\alpha_{21}, M_2)\right) \tag{4.230}$$

$$f_1(\mathtt{M}_{12},\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21})$$

$$\triangleq \begin{cases} g\left(N_1,((\alpha_{11}-\alpha_{13})^+,r_{123}),(\alpha_{11},M_1-r_{123}),(\alpha_{21},M_2)\right) & \mathsf{INR}_{12}\geq\mathsf{INR}_{13} \\ g\left(N_1,((\alpha_{11}-\alpha_{12})^+,r_{123}),(\alpha_{11},M_1-r_{123}),(\alpha_{21},M_2)\right) & \mathsf{INR}_{12}<\mathsf{INR}_{13} \end{cases} \tag{4.231}$$

$$f_1(\mathtt{M}_1,\mathtt{M}_{21})$$

$$\triangleq g\left(N_1,(\alpha_{11},M_1),(\alpha_{21},M_2)\right) \tag{4.232}$$

$$f_2(\mathtt{M}_{2p})$$

$$\triangleq g\left(N_2,\left((\alpha_{22}-\alpha_{21})^+,\min\{M_2,N_1\}\right),\left(\alpha_{22},(M_2-N_1)^+\right)\right) \tag{4.233}$$

$$f_2(\mathtt{M}_2)$$

$$\triangleq \min\{M_2,N_2\}\alpha_{22} \tag{4.234}$$

$$f_2(\mathtt{M}_{12},\mathtt{M}_{2p})$$

$$\triangleq \begin{cases} g\left(N_2,((\alpha_{12}-\alpha_{13}),r_{123}),(\alpha_{12},r_{12}-r_{123}),((\alpha_{22}-\alpha_{21})^+,\min\{M_2,N_1\}),\right. & \mathsf{INR}_{12}\geq\mathsf{INR}_{13} \\ \qquad \left.(\alpha_{22},(M_2-N_1)^+)\right) & \\ g\left(N_2,(\alpha_{12},r_{12}-r_{123}),((\alpha_{22}-\alpha_{21})^+,\min\{M_2,N_1\}),(\alpha_{22},(M_2-N_1)^+)\right) & \mathsf{INR}_{12}<\mathsf{INR}_{13} \end{cases} \tag{4.235}$$

$$f_2(\mathtt{M}_{12},\mathtt{M}_2)$$

$$\triangleq \begin{cases} g\left(N_2,((\alpha_{12}-\alpha_{13}),r_{123}),(\alpha_{12},r_{12}-r_{123}),(\alpha_{22},M_2)\right) & \mathsf{INR}_{12}\geq\mathsf{INR}_{13} \\ g\left(N_2,(\alpha_{12},r_{12}-r_{123}),(\alpha_{22},M_2)\right) & \mathsf{INR}_{12}<\mathsf{INR}_{13} \end{cases} \tag{4.236}$$

$$f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p})$$

$$\triangleq g\left(N_2,(\alpha_{12},M_1),\left((\alpha_{22}-\alpha_{21})^+,\min\{M_2,N_1\}\right),\left(\alpha_{22},(M_2-N_1)^+\right)\right) \tag{4.237}$$

$$f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2)$$

$$\triangleq g\left(N_2,(\alpha_{12},M_1),(\alpha_{22},M_2)\right) \tag{4.238}$$

$$f_3(\mathtt{M}_3)$$

$$\triangleq \min\{M_3,N_3\}\alpha_{33} \tag{4.239}$$

$$f_3(\mathtt{M}_{13},\mathtt{M}_3)$$

$$\triangleq \begin{cases} g\left(N_3,(\alpha_{13},r_{13}-r_{123}),(\alpha_{33},M_3)\right) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ g\left(N_3,(\alpha_{13}-\alpha_{12},r_{123}),(\alpha_{13},r_{13}-r_{123}),(\alpha_{33},M_3)\right) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{4.240}$$

$$f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3)$$

$$\triangleq g\left(N_3,(\alpha_{13},M_1),(\alpha_{33},M_3)\right) \tag{4.241}$$

Theorem 4.5. *The GDoF region $\mathcal{D}(\bar{\alpha})$ of the MIMO IC-ZIC is given by* (4.242)-(4.269)*.*

$$\mathcal{D}(\bar{\alpha}) \triangleq \{(d_1,d_2,d_3) \in \mathbb{R}_+^3 :$$

$$d_1 \leq f_1(\mathtt{M}_1) \tag{4.242}$$

$$d_2 \leq f_2(\mathtt{M}_2) \tag{4.243}$$

$$d_3 \leq f_3(\mathtt{M}_3) \tag{4.244}$$

$$d_1+d_2 \leq f_1(\mathtt{M}_1,\mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) \tag{4.245}$$

$$d_1+d_2 \leq f_1(\mathtt{M}_{13},\mathtt{M}_{1p}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) \tag{4.246}$$

$$d_1+d_2 \leq f_1(\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) \tag{4.247}$$

$$d_1+d_3 \leq f_1(\mathtt{M}_{12},\mathtt{M}_{1p}) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.248}$$

$$d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_2(\mathtt{M}_{12},\mathtt{M}_2) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.249}$$

$$d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) + f_3(\mathtt{M}_{13},\mathtt{M}_3) \tag{4.250}$$

$$d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{12},\mathtt{M}_{2p}) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.251}$$

$$d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) + f_3(\mathtt{M}_{13},\mathtt{M}_3) \tag{4.252}$$

$$d_1+d_2+d_3 \leq f_1(\mathtt{M}_{12},\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.253}$$

$$d_1+2d_2 \leq f_1(\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) \tag{4.254}$$

$$2d_1+d_2 \leq f_1(\mathtt{M}_{13},\mathtt{M}_{1p}) + f_1(\mathtt{M}_1,\mathtt{M}_{21}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) \tag{4.255}$$

$$2d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12},\mathtt{M}_{13},\mathtt{M}_{1p}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.256}$$

$$2d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12},\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.257}$$

$$2d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_1,\mathtt{M}_{21}) + f_2(\mathtt{M}_{12},\mathtt{M}_{2p}) + f_3(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_3) \tag{4.258}$$

$$2d_1+d_2+d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_1,\mathtt{M}_{21}) + f_2(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) + f_3(\mathtt{M}_{13},\mathtt{M}_3) \tag{4.259}$$

$$d_1 + 2d_2 + d_3 \leq f_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{12}, \mathtt{M}_2) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.260}$$

$$d_1 + 2d_2 + d_3 \leq f_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + f_3(\mathtt{M}_{13}, \mathtt{M}_3) \tag{4.261}$$

$$2d_1 + 2d_2 + d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.262}$$

$$2d_1 + 2d_2 + d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.263}$$

$$2d_1 + 2d_2 + d_3 \leq f_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.264}$$

$$2d_1 + d_2 + 2d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + 2f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.265}$$

$$2d_1 + d_2 + 2d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + f_3(\mathtt{M}_{13}, \mathtt{M}_3) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.266}$$

$$3d_1 + d_2 + d_3 \leq f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + f_1(\mathtt{M}_1, \mathtt{M}_{21}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.267}$$

$$3d_1 + 2d_2 + 2d_3 \leq 2f_1(\mathtt{M}_{1p}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + f_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + 2f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \tag{4.268}$$

$$2d_1 + 3d_2 + d_3 \leq f_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + f_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + 2f_2(\mathtt{M}_{2p}) + f_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + f_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)\big\} \tag{4.269}$$

Proof. In Definition 4.6, the GDoF region is defined by the capacity region $\mathcal{C}$. We do not have the exact capacity region $\mathcal{C}$ for the MIMO IC-ZIC, but Theorem 4.4 suggests that both $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$ are within a SNR and INR independent gap to the capacity. Because a finite number of bits are insignificant in the GDoF computation, the GDoF region can be obtained from either $\mathcal{R}_{\text{in}}$ or $\mathcal{R}_{\text{o}}$. To characterize the GDoF region, we just compute the limit of each set function in Definition 4.5 when $\rho \to \infty$.

By Fact 4.2, the limit of set functions like (4.170) or (4.175) can immediately be obtained because these set functions are already expressed in the form of $\log(I + \sum_{i=1}^{n} H_i H_i^\dagger)$. However, the limit of set functions like (4.167) cannot be inferred immediately. We alternatively rewrite terms like $H_{11} K_{13,1p} H_{11}^\dagger$ as a sum of terms in the form of $HH^\dagger$ and then apply Fact 4.2. Please refer to (A.46) in Appendix A.5 for the detailed computation of limit of (4.167), which leads to (4.224) of the GDoF region. $\square$

Example 4.1. Consider the MIMO IC-ZIC with the following parameters: $\alpha_{11} = \alpha_{22} = \alpha_{33} = 1$, $\alpha_{12} = \alpha_{21} = 0.6$, $\alpha_{13} = 0.3$, $M_1 = N_1 = 3$ and $M_2 = M_3 = N_2 = N_3 = 2$. Given this setting, we have $r = 3$,

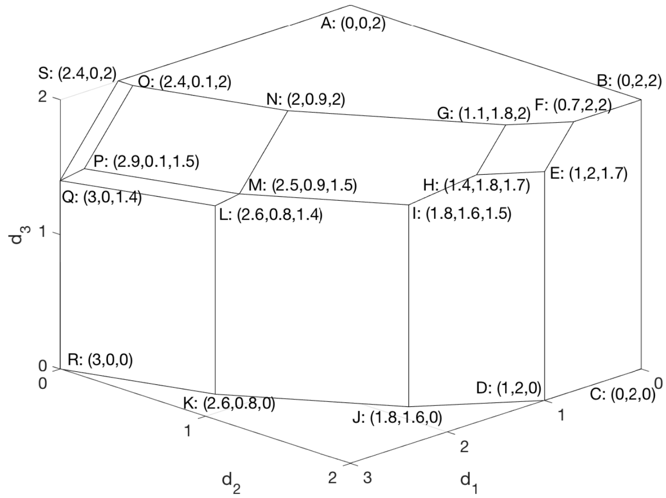

Figure 4.6: GDoF region of a (3,3,2,2,2,2) MIMO IC-ZIC with $\alpha_{12} = \alpha_{21} = 0.6$ and $\alpha_{13} = 0.3$

$r_{123} = 1$ and $r_{12} = r_{13} = 2$. The GDoF region is plotted in Fig. 4.6.

We provide an overview of the GDoF region in Example 4.1. The MIMO IC-ZIC consists of two ZICs as its sub-channels. The tuples on the $(d_1, d_2, 0)$ form the GDoF region of the two-user IC with INR $\rho^{\alpha_{12}}$ and $\rho^{\alpha_{21}}$ which is consistent with the plot in [26, Fig. 2]. The tuples on the $(d_1, 0, d_3)$ plane form the GDoF region of a two-user ZIC with INR $\rho^{\alpha_{13}}$. The rate tuples on d_3 vs d_2 plane when $d_1 = 0$ reflect the GDoF region of a parallel channel between Tx2/Rx2 and Tx3/Rx3 while Tx1 is off. The sum GDoF plane is G-H-I-M-N, and any GDoF tuple on this plane achieves the max sum GDoF 4.9.

Example 4.2. Continue with the MIMO one-to-three IC in Example 4.1. We describe the structure of the transmitted signals from the three transmitters in terms of independent data streams according to (4.154)-(4.157) in Section 4.4.1. The coding scheme suggests we send the following data streams at Tx1.

$$
\begin{aligned}
W_{123} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^{\dagger})}} V_p^{(2)} \sqrt{1 - \frac{1}{1 + \rho^{0.3}\lambda_{13,r-r_{13}+k}^2}} \mathtt{X}_{123}^{(1)} \\
W_{12} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^{\dagger})}} \left(V_p^{(1)} \sqrt{1 - \frac{1}{1 + \rho^{0.6}}} \mathtt{X}_{12}^{(1)} + V_p^{(2)} \sqrt{\frac{1}{1 + \rho^{0.3}\lambda_{13,2}^2} - \frac{1}{1 + \rho^{0.6}\lambda_{12,2}^2 + \rho^{0.3}\lambda_{13,2}^2}} \mathtt{X}_{12}^{(2)} \right) \\
W_{13} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^{\dagger})}} V_p^{(3)} \sqrt{1 - \frac{1}{1 + \rho^{0.3}}} \mathtt{X}_{13}^{(1)} \\
W_{1p} &= \sqrt{\frac{1}{\mathrm{Tr}(V_p V_p^{\dagger})}} \left(V_p^{(1)} \sqrt{\frac{1}{1 + \rho^{0.6}}} \mathtt{X}_{1p}^{(1)} + V_p^{(2)} \sqrt{\frac{1}{1 + \rho^{0.6}\lambda_{12,k}^2 + \rho^{0.3}\lambda_{13,k}^2}} \mathtt{X}_{1p}^{(2)} + V_p^{(3)} \sqrt{\frac{1}{1 + \rho^{0.3}}} \mathtt{X}_{1p}^{(3)} \right).
\end{aligned}
$$

More specifically, there is one public data stream $\mathtt{X}_{123}^{(1)}$ for all receivers, two public data streams $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{12}^{(2)}$ for Rx1 and Rx2, one public data stream $\mathtt{X}_{13}^{(1)}$ for Rx1 and Rx3, and three private data streams for Rx1 only. The data streams $\mathtt{X}_{123}^{(1)}$, $\mathtt{X}_{12}^{(1)}$ and $\mathtt{X}_{13}^{(1)}$ are sent at power level ρ^0. The data stream $\mathtt{X}_{12}^{(2)}$ is sent at power level $\rho^{-0.3}$ as this is the part to be decoded by Rx2, but treated as noise by Rx3. The first private data stream $\mathtt{X}_{1p}^{(1)}$ is sent at power level $\rho^{-0.6}$ so that Rx2 could treat it as noise. The second private data stream $\mathtt{X}_{1p}^{(2)}$ is sent at power level $\rho^{-0.6}$ so that both Rx2 and Rx3 could treat it as noise. The third private data stream $\mathtt{X}_{1p}^{(3)}$ is sent at power level $\rho^{-0.3}$ so Rx3 could treat it as noise.

The coding scheme also implies the following data streams sent at Tx2.

$$X_2 = \sum_{k=1}^{2} V_{21}^{[k]} \sqrt{\frac{\rho^{0.6}\sigma_{21,k}^2}{2(1+\rho^{0.6}\sigma_{21,k}^2)}} \mathtt{X}_{21}^{(k)} + \sum_{k=1}^{2} V_{21}^{[k]} \sqrt{\frac{1}{2(1+\rho^{0.6}\sigma_{21,k}^2)}} \mathtt{X}_{2p}^{(k)}.$$

Tx2 has two data streams, $\mathtt{X}_{21}^{(1)}$ and $\mathtt{X}_{21}^{(2)}$, for its common sub-message to be decoded at both Rx1 and Rx2, as well as two private data streams $\mathtt{X}_{2p}^{(1)}$ and $\mathtt{X}_{2p}^{(2)}$ for its private message to be decoded at Rx2 while arriving under the noise floor at Rx1.

The data streams sent at Tx3 are simply

$$X_3 = \sum_{k=1}^{2} \frac{1}{\sqrt{2}} \mathtt{X}_3^{(k)}.$$

Tx3 sends two private streams, $\mathtt{X}_3^{(1)}$ and $\mathtt{X}_3^{(2)}$, to Rx3 only.

In what follows, we analyze the achievability of the five corner points on the max sum GDoF plane in Example 4.1. For each corner point, we provide GDoF distribution among the data streams revealed in Example 4.2. The detailed GDoF allocation on each data stream will be illustrated via multi-dimensional signal partitioning introduced in Chapter 2. Each GDoF allocation will be plotted in a signal diagram with each of the received independent (transmit) signal directions (from the receiver's perspective) plotted as a multi-leveled bar whose top level marks its signal strength, and the vertical height of each partition is proportional to the DoFs carried by it. The underlying coding scheme can be directly read from the GDoF allocation. A transmitter encodes all data streams on a (transmit) signal direction by multi-level superposition coding from bottom to top (refer to the GDoF allocations figures for the position of the data streams on each signal direction), and the receiver decodes the signal by either successive cancellation or joint decoding. No cross signal level is employed, and each data stream is encoded independently. The underlying coding scheme can be different from the coding scheme we used to derive the inner bound in Section 4.4.1. In all the GDoF analysis figures in the rest of the chapter, transmit signal directions $V_p^{[1]}$, $V_p^{[2]}$, and $V_p^{[3]}$ are sorted from left to right in sequence at Rx1, and$V_{21}^{[1]}$ and $V_{21}^{[2]}$ are shown from left to right at Rx2, and $V_p^{[2]}$ and $V_p^{[3]}$ are shown from left to right at Rx3.

4.5.1.1 The achievability of point G (1.1,1.8,2)

We choose the GDoF distribution $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0.2$, $d_{12}^{(2)} = 0$, $d_{13}^{(1)} = 0$, $d_{1p}^{(1)} = 0.2$, $d_{1p}^{(2)} = 0$, $d_{1p}^{(3)} = 0.7$, $d_{21}^{(1)} = 0.4$, $d_{21}^{(2)} = 0.6$, $d_{2p}^{(1)} = d_{2p}^{(2)} = 0.4$ and $d_3^{(1)} = d_3^{(2)} = 1$. The GDoF allocation among the three transmitters are illustrated in Fig. 4.7. This allocation guarantees an interference free channel between Tx3 and Rx3. Due to the precoding (by matrix V_p), the second and third transmit directions $V_p^{[2]}$ and $V_p^{[3]}$ do not appear at Rx2 and Rx3. All the signal levels at both Rx2 and Rx3 are fully utilized.

Rx1 first removes the effect of $\mathtt{X}_{1p}^{(3)}$ from Y_1 by zero forcing, i.e. projecting the received signal onto the 2-D signal space which is perpendicular to $H_{11}V_p^{[3]}$. In the resulting 2-D signal space, Tx1, Tx2 and Rx1 form a (2,2,2) MIMO MAC channel, and their contributions $\mathtt{X}_{12}^{(1)}$, $\mathtt{X}_{1p}^{(1)}$, $\mathtt{X}_{21}^{(1)}$ and $\mathtt{X}_{21}^{(2)}$ are present. Given the power level assignment in Fig. 4.7, $\mathtt{X}_{12}^{(1)}$ can be recovered by treating the W_{21} and W_{1p} as noise, resulting in GDoF $d_{12}^{(1)} = 0.2$. Since $\mathtt{X}_{12}^{(1)}$ also needs to be decoded by Rx2, it remains to verify if the same GDoF can be achieved at Rx2. Subtracting the contribution of $\mathtt{X}_{12}^{(1)}$, we decode $\mathtt{X}_{21}^{(1)}$, $\mathtt{X}_{21}^{(2)}$ and $\mathtt{X}_{1p}^{(1)}$ jointly, which gives GDoF $d_{12}^{(1)} = 0.4$, $d_{21}^{(2)} = 0.6$ and $d_{1p}^{(1)} = 0.2$, respectively, where achievability of $d_{21}^{(1)}$ and $d_{21}^{(2)}$ has to be confirmed at Rx2 later. After the recovery of $\mathtt{X}_{12}^{(1)}$, $\mathtt{X}_{1p}^{(1)}$, $\mathtt{X}_{21}^{(1)}$ and $\mathtt{X}_{21}^{(2)}$, we remove their effects from Y_1, so $\mathtt{X}_{1p}^{(3)}$ can be recovered, resulting in GDoF $d_{1p}^{(3)} = 0.7$.

At Rx2, the power level assignment of $\mathtt{X}_{21}^{(1)}$, $\mathtt{X}_{21}^{(2)}$ and $\mathtt{X}_{12}^{(1)}$ allow us to decode them jointly by treating $\mathtt{X}_{2p}^{(1)}$ and $\mathtt{X}_{2p}^{(2)}$ as noise, resulting in GDoF $d_{21}^{(1)} = 0.4$, $d_{21}^{(2)} = 0.6$ and $d_{12}^{(1)} = 0.2$, which are consistent with the achievable GDoF allowed by these data streams at Rx1. Removing the effect of $\mathtt{X}_{21}^{(1)}$, $\mathtt{X}_{21}^{(2)}$ and $\mathtt{X}_{12}^{(1)}$, we see an interference free 2×2 MIMO P2P channel between Tx2 and Rx2, and $\mathtt{X}_{2p}^{(1)}$ and $\mathtt{X}_{2p}^{(2)}$ can be decoded with $d_{2p}^{(1)} = d_{2p}^{(2)} = 0.4$.

At Rx3, the interference from Tx1 arrives under the noise floor. So Rx3 simply decodes its intended

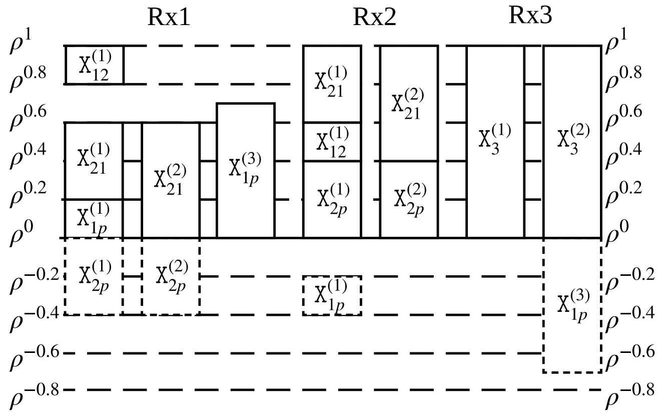

Figure 4.7: GDoF allocation at point G

signal, which leads to $d_3^{(1)} = d_3^{(2)} = 1$.

4.5.1.2 The achievability of point H (1.4,1.8,1.7)

We choose the GDoF distribution $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0.2$, $d_{12}^{(2)} = 0$, $d_{13}^{(1)} = 0.3$, $d_{1p}^{(1)} = 0.2$, $d_{1p}^{(2)} = 0$, $d_{1p}^{(3)} = 0.7$, $d_{21}^{(1)} = 0.4$, $d_{21}^{(2)} = 0.6$, $d_{2p}^{(1)} = d_{2p}^{(2)} = 0.4$ $d_3^{(1)} = 1$ and $d_3^{(2)} = 0.7$. The GDoF allocation among the three transmitters is illustrated in Fig. 4.8. Comparing to the GDoF allocation for point G, the difference is that $\mathtt{X}_{13}^{(1)}$ carries GDoF 0.3. The decoding procedure to recover $\mathtt{X}_{12}^{(1)}$, $\mathtt{X}_{1p}^{(1)}$, $\mathtt{X}_{21}^{(1)}$ and $\mathtt{X}_{21}^{(2)}$ at Rx1 is exactly the same as the procedure at corner point G. But after removing the effects of $\mathtt{X}_{12}^{(1)}$, $\mathtt{X}_{1p}^{(1)}$, $\mathtt{X}_{21}^{(1)}$ and $\mathtt{X}_{21}^{(2)}$ from Y_1, Rx1 decodes $\mathtt{X}_{13}^{(1)}$ and $\mathtt{X}_{1p}^{(3)}$ successively, which gives $d_{13}^{(1)} = 0.3$ and $d_{1p}^{(3)} = 0.7$. The decoding procedure at Rx2 is exactly the same as the procedure at the corner point G. Rx3 decodes $\mathtt{X}_3^{(1)}$, $\mathtt{X}_3^{(2)}$ and $\mathtt{X}_{13}^{(1)}$ jointly, resulting in $d_3^{(1)} = 1$, $d_3^{(2)} = 0.7$ and $d_{13}^{(1)} = 0.3$.

4.5.1.3 The achievability of point I (1.8,1.6,1.5)

We choose the GDoF distribution $d_{123}^{(1)} = 0.2$, $d_{12}^{(1)} = 0.2$, $d_{12}^{(2)} = 0$, $d_{13}^{(1)} = 0.3$, $d_{1p}^{(1)} = 0.2$, $d_{1p}^{(2)} = 0.2$, $d_{1p}^{(3)} = 0.7$, $d_{21}^{(1)} = d_{21}^{(2)} = 0.4$, $d_{2p}^{(1)} = d_{2p}^{(2)} = 0.4$, $d_3^{(1)} = 0.8$ and $d_3^{(2)} = 0.7$. The GDoF allocation among the three transmitters is illustrated in Fig. 4.9. The decoding procedures at Rx1-Rx3 can be learnt from the decoding procedures at the corner points G and H.

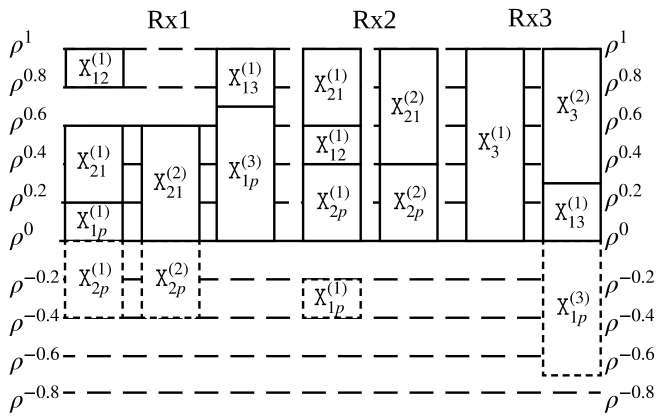

Figure 4.8: GDoF allocation at point H

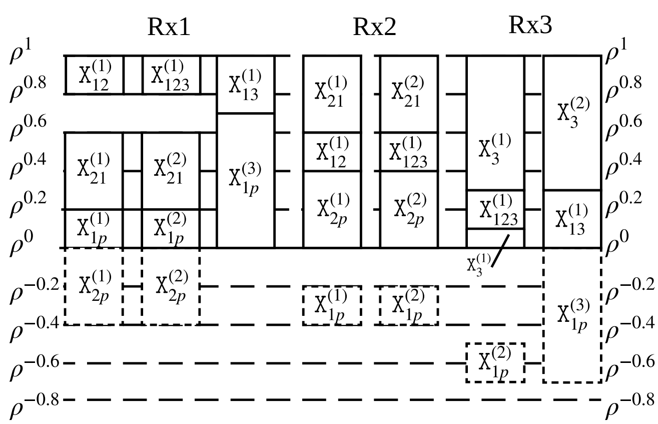

Figure 4.9: GDoF allocation at point I

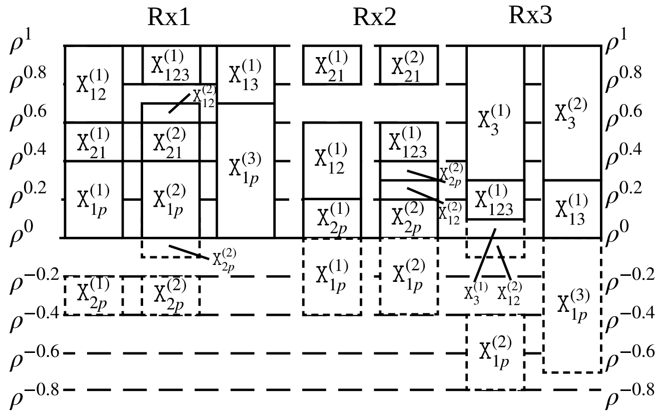

Figure 4.10: GDoF allocation at point M

4.5.1.4 The achievability of point M (2.5,0.9,1.5)

We choose the GDoF distribution $d_{123}^{(1)} = 0.2$, $d_{12}^{(1)} = 0.4$, $d_{12}^{(2)} = 0.1$, $d_{13}^{(1)} = 0.3$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_{21}^{(1)} = d_{21}^{(2)} = 0.2$, $d_{2p}^{(1)} = 0.2$, $d_{2p}^{(2)} = 0.3$, $d_{3}^{(1)} = 0.8$ and $d_{3}^{(2)} = 0.7$. The GDoF allocation among the three transmitters is illustrated in Fig. 4.10. The decoding procedures at Rx1-Rx3 can be learnt from the decoding procedures at the corner points G and H.

4.5.1.5 The achievability of point N (2,0.9,2)

We choose the GDoF distribution $d_{123}^{(1)} = 0$, $d_{12}^{(1)} = 0.4$, $d_{12}^{(2)} = 0.1$, $d_{13}^{(1)} = 0$, $d_{1p}^{(1)} = d_{1p}^{(2)} = 0.4$, $d_{1p}^{(3)} = 0.7$, $d_{21}^{(1)} = d_{21}^{(2)} = 0.2$, $d_{2p}^{(1)} = 0.2$, $d_{2p}^{(2)} = 0.3$ and $d_{3}^{(1)} = d_{3}^{(2)} = 1$. The GDoF allocation among the three transmitters is illustrated in Fig. 4.11. The decoding procedures at Rx1-Rx3 can be learnt from the decoding procedures at the corner points G and H.

4.5.2 The Sum GDoF Curve

Next, we keep the number of transmit and receive antennas unchanged in Example (4.1), and let α run through the internal $[0, 2]$ to see the variation of the sum GDoF. The sum GDoF vs α curve is plotted in Fig. 4.12. There are five corner points in the middle of the curve. At the corner point (0.25,4.75), the interferences between Tx1/Rx1 and Tx2/Rx2 turn moderate, i.e. $\alpha_{12} = \alpha_{21} \in [0.5, 1]$ and the interference between Tx1 to Rx1 stays weak. At the corner point (0.5,4.5), the interferences between Tx1/Rx1 and

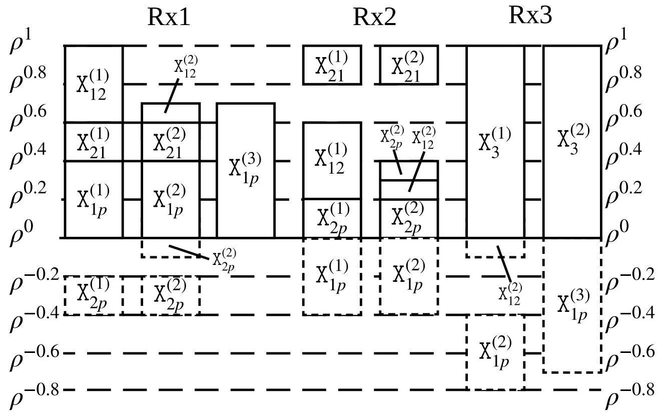

Figure 4.11: GDoF allocation at point N

Tx2/Rx2 turn strong ($\alpha_{12} = \alpha_{21} = [1, 2]$) and the interference from Tx1 to Rx3 turns moderate. At the corner point (1,5), the interferences between Tx1/Rx1 and Tx2/Rx2 turn very strong ($\alpha_{12} = \alpha_{21} \geq 2$), and the interference from Tx1 to Rx3 turns strong. We analyze the achievability of two sum GDoF optimal corner points $(1/3, 5)$ and $(0.8, 4.8)$ on the curve. By Theorem 4.5, the equal GDoF tuples $(5/3, 5/3, 5/3)$ and (1.6,1.6,1.6) are achievable. A coding scheme to achieve $(5/3, 5/3, 5/3)$ when $\alpha = 1/3$ is illustrated in Fig. 4.13. Two coding schemes to achieve (1.6,1.6,1.6) when $\alpha = 0.8$ are illustrated in Fig. 4.14 and Fig. 4.15. Note in the latter two coding schemes, there are signal levels overlapping between two different data streams at Rx1 and Rx2; however, it is clear that with the power level assignments given in the diagram, both Rx1 and Rx2 could recover their received data streams with joint decoding.

Next, let us take an example when $\mathsf{INR}_{12} < \mathsf{INR}_{13}$. Consider a $(3, 3, 2, 2, 2, 2)$ IC-ZIC with $\alpha_{12} = \alpha_{21} = \alpha$ and $\alpha_{13} = 2\alpha$. Its sum GDoF curve is plotted in Fig. 4.16. The achievability of GDoF tuple (14/9,14/9,14/9) when $\alpha = 1/3$ is illustrated in Fig. 4.17.

Lastly, we plot the sum GDoF curve of a SISO one-to-three IC in Fig. 4.18. There is only one antenna at each transmitter and each receiver, and again we choose $\alpha_{13} = \alpha$ and $\alpha_{12} = 2\alpha$.

4.6 Conclusion

We derived a pair of single region inner and outer bounds which are within a SNR/INR independent gap. An explicit coding scheme which incorporates three level superposition coding at Tx1 (as in the MIMO

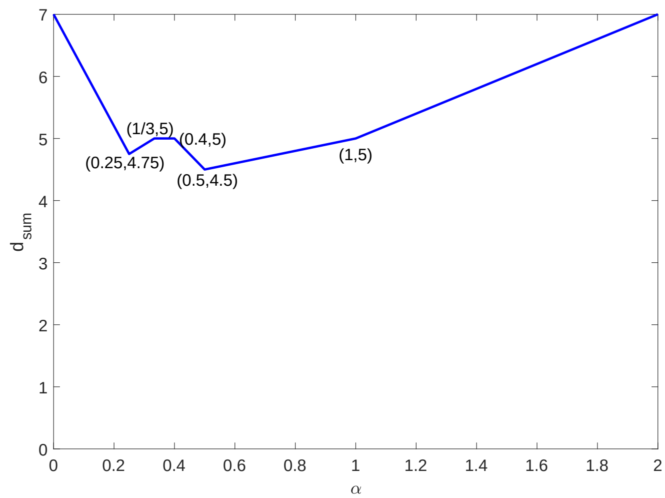

Figure 4.12: Sum GDoG Curve of a (3,3,2,2,2,2) MIMO IC-ZIC with $\alpha_{12} = \alpha_{21} = 2\alpha$ and $\alpha_{13} = \alpha$

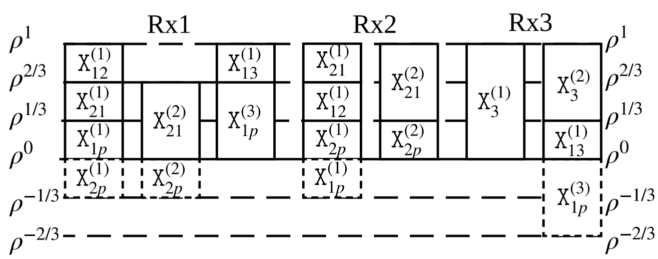

Figure 4.13: A GDoF allocation to achieve (5/3,5/3,5/3) when $\alpha = 1/3$

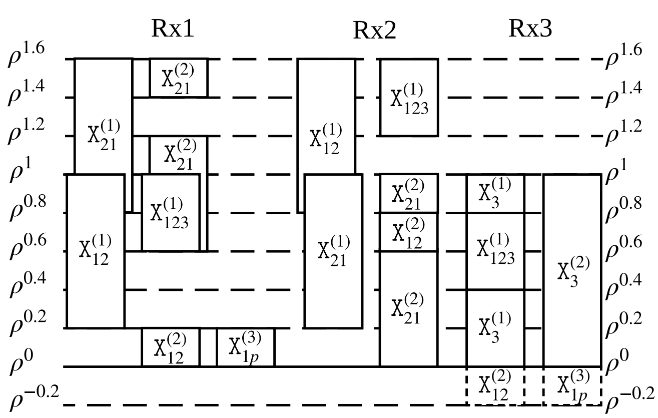

Figure 4.14: A GDoF allocation to achieve (1.6,1.6,1.6) when $\alpha = 0.8$

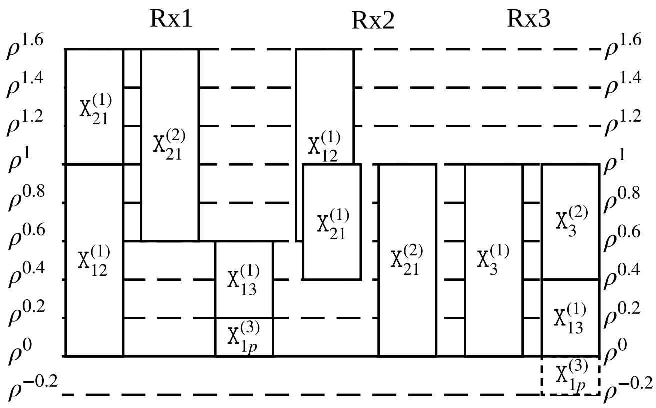

Figure 4.15: Another GDoF allocation to achieve (1.6,1.6,1.6) when $\alpha = 0.8$

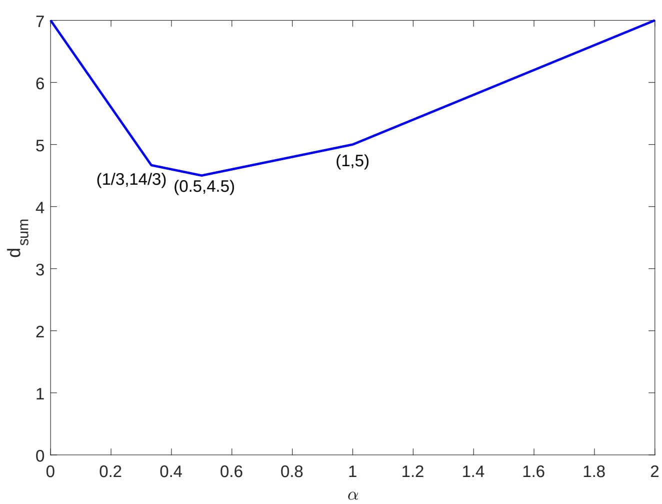

Figure 4.16: Sum GDoG Curve of a (3,3,2,2,2,2) MIMO IC-ZIC with $\alpha_{12} = \alpha_{21} = \alpha$ and $\alpha_{13} = 2\alpha$

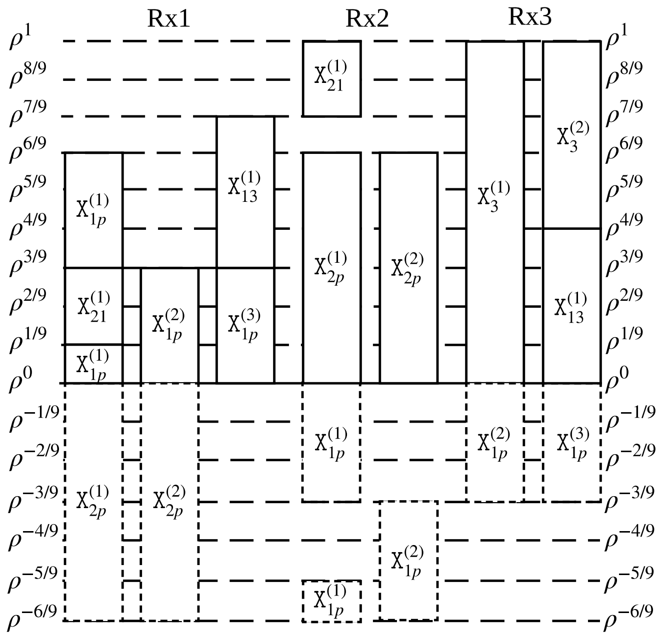

Figure 4.17: A GDoF allocation to achieve (14/9,14/9,14/9) when $\alpha = 1/3$

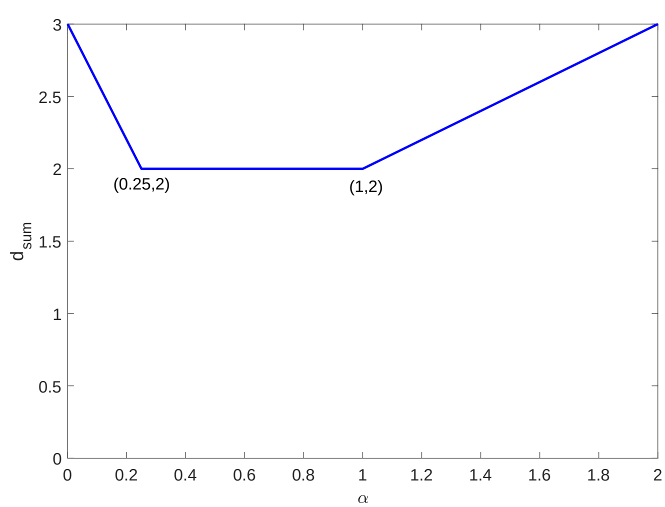

Figure 4.18: Sum GDoG Curve of a SISO IC-ZIC with $\alpha_{12} = \alpha_{21} = 2\alpha$ and $\alpha_{13} = \alpha$

one-to-three IC, c.f. Section 3.4.2), Karmakar-Varanasi type coding at Tx2 and non-water filling single user coding at Tx3 turns out to be GDoF optimal. The GDoF region of the MIMO IC-ZIC is then fully characterized, and we numerically studied achievability of the GDoF region and the sum GDoF curve of several channel examples.

Chapter 5

Constant-Gap-to-Capacity and Generalized Degrees of Freedom Regions of the MIMO MAC-IC-MAC

5.1 Introduction

Spectrum sharing allows the coexistence of heterogeneous wireless networks on the same frequency band. Managing the interference in the same band between such networks is critical to ensure high spectrum efficiency. The MAC-IC-MAC is an abstract channel model inspired by practical co-band network scenarios where two multiple-access channels (MACs) mutually interfere with each other, but in which there is interference only from one of the transmitters of each MAC to the receiver of the other MAC (see [37] for illustrations of practical settings of the MAC-IC-MAC). An approximate capacity region for the scalar Gaussian MAC-IC-MAC within a two bit gap was found in [37].

Multiple-antenna transmission and reception have been widely adopted in many wireless systems in the last decade. Hence, the approximate capacity region of the MIMO MAC-IC-MAC and its achievability analysis can provide the relevant understanding and insight on how coding schemes might be designed for modern co-band network communications in settings that involve terminals with multiple antennas. Fig. 5.1 is an illustration of the MIMO MAC-IC-MAC.

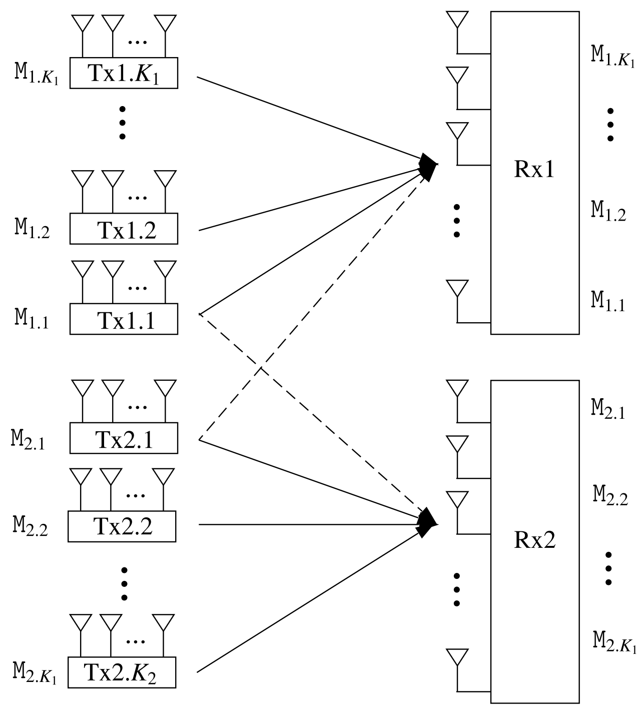

Figure 5.1: MIMO MAC-IC-MAC

5.1.1 Main Contributions

5.1.1.1 A constant-gap-to-capacity region for the MIMO MAC-IC-MAC

The constant-gap-to-capacity region is based on the inner and outer bounds on the semi-deterministic MAC-IC-MAC obtained by the authors in [37]. We explicitly specify one coding scheme and propose its achievable rate region, a single polytope, to be the inner bound on the capacity region of the MIMO Gaussian MAC-IC-MAC. This scheme simply lets interfering transmitters employ the Karmakar-Varanasi (KV) superposition coding scheme of [27] proposed therein for the 2-user MIMO interference channel, and the non-interfering transmitters employ single-user Gaussian codebooks with scaled identity covariance matrices (i.e., with no beamforming or water-filling). These two coding schemes for interfering and non-interfering transmitters by themselves are known to achieve constant-gap-to-capacity regions in the MIMO interference channel [27] and the MIMO MAC [39], respectively. We hence unify and generalize those two results in this chapter. The outer bound on the capacity region in the form of a single polytope is characterized by specifying extremal input and genie signal distributions in the union-of-polytopes outer bound for the semi-deterministic MAC-IC-MAC proposed by the author in the previous work [37], in the context of the MIMO MAC-IC-MAC. The gap between the inner and outer bounds, while dependent on the numbers of transmit/receive antennas and the numbers of users in each cell, is shown to be independent of all channel matrices and signal- and interference-to-noise ratios. Hence, the explicit inner (or outer) bound is an approximation of the capacity region that is guaranteed to be within a constant gap.

5.1.1.2 Analysis of the GDoF region of the MIMO MAC-IC-MAC

The generalized degrees of freedom (GDoF) region is characterized and the achievability of its key corner points is analyzed using the multidimensional signal-level partitioning technique introduced in Chapter 2. We also study the symmetric GDoF curve under various antenna configurations and analyze the role of the non-interfering transmitters in affecting the symmetric GDoF curve. When the interference strength is weak or strong enough, the non-interfering transmitters can fully occupy the receiver's signal partitions which cannot be utilized by the interfering transmitter, which improves the cell spectrum efficiency. This phenomenon has been discovered in our previous work on the scalar Gaussian MAC-IC-MAC [37]. Moreover, when a receiver has more antennas than the interfering transmitter does, the non-interfering transmitters saturate receiver's signal dimensions which are not used by the signal from the interfering transmitter, which also improves the cell spectrum efficiency.

5.1.2 Related Previous Work

The capacity of the time-invariant Gaussian MIMO point-to-point (P2P) channel was characterized in [42], where the optimal Gaussian random coding scheme can be specified via beamforming and water-filling power allocation via the singular value decomposition of the channel matrix. The capacity region of the discrete-memoryless multiple access channel (MAC) was characterized by Ahlswede [1] and Liao [31]. In the MIMO Gaussian MAC, Gaussian inputs are optimal and the determination of the boundary of the capacity region via a maximization of the weighted sum rate over input covariances at each transmitter is a convex optimization problem. Multiple access channels are the best understood multi-terminal networks with the capacity region determined by Liao [31], Ahlswede [1] and Wyner [47]. For the fading MIMO MAC with finite discrete fading state, Mohseni **et al.** [33] characterized its capacity and power regions under various power and rate constraints. Romero and Varanasi [39] studied the fading MIMO MAC with general (private and common) message sets and with discrete fading states and showed that employing scaled identity covariance matrices at every transmitter is sufficient to achieve a rate region that is within a constant gap to the capacity region. Their result evidently holds when specialized to the time-invariant MIMO Gaussian MAC with only private messages. Hence, the result in [39] motivates the use of that simple coding scheme at the non-interfering transmitters in the MIMO MAC-IC-MAC. The constant-gap-to-capacity result of [39] specialized to the MIMO MAC with private messages is discussed in Section 5.3.1.

Some of the key papers on two-user interference channels are $[7, 10, 11, 15, 23, 26, 27, 30, 43]$. For the discrete memoryless two-user interference channel, the Han-Kobayashi achievable scheme (HK scheme) in [23], as well as its alternative, the CMG scheme of [11], give the (same) best inner bound to the capacity region known to date. Telatar and Tse [43] found an outer bound for the class of semi-deterministic interference channels and quantified the gap to the CMG inner bound. Etkin et al [15] characterized approximate capacity regions to within one bit gap. Two constant-gap-to-capacity regions for the MIMO interference channel were obtained in [27]. The first region was obtained using the so-called simple coding scheme, referred to herein as the Karmakar-Varanasi (or KV) coding scheme. The second region was obtained by the explicit coding scheme in [27], which represents a choice of one out of three coding schemes (including the simple coding scheme) depending on the rate-pair to be achieved. The latter region was shown to be within a smaller gap to the capacity region [27]. The KV coding scheme involves message splitting (into private and public messages) at each transmitter with Gaussian random coding distributions. Each sub-message's covariance matrix incorporate transmit beamforming and signal-level superposition coding. The GDoF region of MIMO interference channel was established in [26] with an in-depth analysis.

Bounds on the capacity region of the semi-deterministic MAC-IC-MAC were obtained in [37]. Two inner bounds [37, (46) and (53)] and an outer bound [37, Theorem 3] were provided therein, with both inner bounds being within a quantifiable gap of the proposed outer bound as shown in [37, Theorems 4 and 5]. For instance, it was shown in [37] that the Telatar-Tse type coding scheme of [43] at the interfering transmitters and single-user random coding at the non-interfering transmitters is sufficient to achieve an inner bound [37, (53)] which is within a quantifiable gap of the outer bound. However, those bounds on the semi-deterministic MAC-IC-MAC do not directly yield explicit or closed-form constant-gap-to-capacity region for the MIMO MAC-IC-MAC. This is because each of these inner or outer bounds is the union of rate regions over its associated distributions. A constant-gap-to-capacity region for the Gaussian scalar MAC-IC-MAC was given in [37] as well. The authors determined a single region inner bound which is within two-bit gap to the capacity region. To achieve this inner bound, the interfering transmitters perform Etkin-Tse-Wang type coding [15] and the non-interfering transmitters perform single-user random coding with Gaussian codebooks.

5.1.3 Notations

The notation used throughout this chapter will be consistent with that in [37] except messages will be denoted by the symbol $\mathtt{M}$ instead of M, since the latter is used to represent the number of transmit antennas. The j-th user in the i-th cell is indexed as $i.j$, where $i \in \{1, 2\}$, $j \in \{1, \cdots, K_i\}$ and K_i is the number of the user in cell-i. Hence, the j-th transmitter in the i-th cell is denoted as Tx$i.j$, whose message, transmit symbol, rate and degrees of freedom (GDoF) are denoted as $\mathtt{M}_{i.j}$, $X_{i.j}$, $R_{i.j}$ and $d_{i.j}$, respectively.

Let $\mathit{\Theta}_i$ be the set of indices of all users in the i-th cell, i.e. $\mathit{\Theta}_i = \{i.1, \cdots, i.K_i\}$. For the sake of convenience, the K_i-tuples of messages, input symbols, rates and DoFs of users in cell i are denoted as $\mathtt{M}_{\mathit{\Theta}_i}$, $X_{\mathit{\Theta}_i}$, $R_{\mathit{\Theta}_i}$ and $d_{\mathit{\Theta}_i}$. For example, the input symbols of cell-1 $\{X_{1.1}, \cdots, X_{1.K_1}\}$ are denoted simply as $X_{\mathit{\Theta}_1}$. Similarly, $\mathtt{M}_{\mathit{\Theta}_1}$ denotes the K_1-tuple of messages $\{\mathtt{M}_{1.1}, \cdots, \mathtt{M}_{1.K_1}\}$, $R_{\mathit{\Theta}_1}$ denotes the K_1-tuple of their rates $\{R_{1.1}, \cdots, R_{1.K_1}\}$, and $d_{\mathit{\Theta}_1}$ denotes the K_1-tuple of their DoFs $\{d_{1.1}, \cdots, d_{1.K_1}\}$.

Throughout, we let $\mathit{\Omega}_i$ denote any non-empty subset of $\mathit{\Theta}_i$, i.e., $\mathit{\Omega}_i \in 2^{\mathit{\Theta}_i} \backslash \emptyset$, where $2^{\mathit{\Theta}_i}$ is the power set of $\mathit{\Theta}_i$. Moreover, we let $\mathit{\Upsilon}_i$ denote any non-empty subset of $\mathit{\Theta}_i$ that necessarily contains the element $i.1$. The sets $\bar{\mathit{\Upsilon}}_i$ and $\bar{\mathit{\Omega}}_i$ are defined as the complements of $\mathit{\Upsilon}_i$ and $\mathit{\Omega}_i$ relative to $\mathit{\Theta}_i$. Furthermore, the collection of input symbols of users indexed by elements of $\mathit{\Upsilon}_i$ or $\mathit{\Omega}_i$ are written as $X_{\mathit{\Upsilon}_i}$ and $X_{\mathit{\Omega}_i}$.

We use capital letters to denote random vectors, such as $X_{i.j}$, the underlying alphabets are denoted by $\mathcal{X}_{i.j}$, and specific values by $x_{i.j}$. We use the usual short hand notation for (conditional) probability distributions where the lower case arguments also denote the random variables whose (conditional) distribution is being considered. For example, $p(y_i|x_{i.j})$ denotes $p_{\mathrm{Y}_i|\mathrm{X}_{i,j}}(y_i|x_{i,j})$.

In the MIMO MAC-IC-MAC to be defined in the next section, a signal path from the transmitter Tx$i.j$ to the receiver Rxi is represented as $i.j \to i$, so that $h_{i.j\to i}$ and $H_{i.j\to i}$ denote the path attenuation and transfer matrix from Tx$i.j$ to Rxi respectively. Similarly, the signal-to-noise ratio (SNR) and interference-to-noise ratio (INR) from Tx$i.j$ and Tx$i'.j$ to Rxi are written as $\mathsf{SNR}_{i.j\to i}$ and $\mathsf{INR}_{i'.j\to i}$, respectively, where $i, i^{'} \in \{1, 2\}$ and $i \neq i^{'}$.

The achievable schemes of this chapter involve message splitting at the two transmitters that cause interference at their unintended receiver. The common sub-message sent by Tx$i.1$ and decoded at both receivers is denoted as $m_{i.1c}$. The private sub-message of Tx$i.1$ to be decoded only at the intended receiver Rxi is denoted as $m_{i.1p}$. The rates of $m_{i.1c}$ and $m_{i.1p}$ are written as $R_{i.1c}$ and $R_{i.1p}$ and their GDoFs as $d_{i.1c}$ and $d_{i.1p}$, respectively.

We use $\mathbb{C}$ to denote the set of complex numbers and $Z \sim \mathcal{CN}(0, I_N)$ to denote a N-dimensional random vector Z that is distributed according to the complex circularly symmetric Gaussian distribution with zero mean and covariance matrix I_N (the $N \times N$ identity matrix). Both $\det(\cdot)$ or $|\cdot|$ are used to represent the determinant of a matrix. The number of antennas at Tx$i.j$ and Rxi are denoted as $M_{i.j}$ and N_i. The Frobenius norm of a matrix H is denoted by $\|H\|_{\mathrm{F}}^2$, i.e., $\|H\|_{\mathrm{F}}^2 = \mathrm{Tr}(HH^\dagger)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of its matrix argument. We use $\mathbb{U}^{N\times N}$ to represent the set of $N \times N$ unitary matrices. The k-th row and column of the matrix H are denoted as $H^{(k)}$ and $H^{[k]}$ respectively. A sub-matrix obtained by taking the rows k_1 to k_2 of the matrix H is written as $H^{(k_1:k_2)}$. A sub-matrix obtained by taking the columns k_1 to k_2 of the matrix H is written as $H^{[k_1:k_2]}$. For two matrices A and B, if $(A - B)$ is positive definite (p.d.) or positive semi-definite (p.s.d), we write the relationship as $A \succ B$ or $A \succeq B$, respectively. We use $o(1)$ to represent a term which approaches zero asymptotically and $\mathcal{O}(1)$ to represent a term which is bounded above by some constant. The function $(M)^+$ returns the maximum value of M and 0, i.e., $(M)^+ = \max\{M, 0\}$.

The rest of the chapter is organized as follows. Section II describes the channel models and formulates the problem; Section III presents the constant-gap-to-capacity region of the MIMO MAC-IC-MAC; Section IV introduces multidimensional signal-level partitioning; Section V characterizes the GDoF region, and investigates the achievability of the key corner points in the GDoF region as well as the symmetric GDoF curve; Section VI concludes the chapter. Some detailed proofs are relegated to appendices.

5.2 Channel Model and Problem Formulation

A (K_1, K_2) MIMO (multiple input multiple output) Gaussian MAC-IC-MAC (MIMO MAC-IC-MAC for short) consists two uplink cells: (Tx1.1,$\cdots$,Tx1.$K_1$$\to$Rx1) and (Tx2.1,$\cdots$,Tx2.$K_2$$\to$Rx2). Two interference links exist from Tx1.1 to Rx2 and from Tx2.1 to Rx1. There are $M_{i.j}$ transmit antennas at Tx$i.j$ and N_i receive antennas at Rxi, where $j \in \{1, \cdots, K_i\}$ for some $i \in \{1, 2\}$. Let $H_{i.j\to i} \in \mathbb{C}^{N_i \times M_{i.j}}$ and $H_{i.j,i'} \in \mathbb{C}^{N_{i'} \times M_{i.j}}$ be the channel matrices from Txi to Rxi and Rx$i^{'}$ respectively, whose entries are drawn i.i.d. from a continuous and unitarily invariant distribution [45] i.e., $UH_{i.j\to i}V$ is identically distributed to $H_{i.j\to i}$ for any $U_{N_i} \in \mathbb{U}^{N_i\times N_i}$ and $V \in \mathbb{U}^{M_{i.j}\times M_{i.j}}$ (also $UH_{i.j\to i'}V$ is identically distributed to $H_{i.j\to i'}$). Such matrices $H_{i,j\to i}$ and $H_{i.1\to i'}$ are full rank with probability one (w.p.1). At time t, Txi choose a vector

$X_{i.j,t} \in \mathbb{C}^{M_{i.j}\times 1}$ and send it over the channel. The per-codeword transmission power is constrained as

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{Tr}(x_{i.j,t}x_{i.j,t}^{\dagger}) \le P_{i.j}.$$

The input-output relation of this channel can be written as

$$\begin{aligned} Y_1 &= \left(\sum_{j=1}^{K_1} h_{1.j\to 1}H_{1.j\to 1}X_{1.j}\right) + h_{2.1\to 1}H_{2.1\to 1}X_{2.1} + Z_1 \\ &\triangleq \left(\sum_{j=1}^{K_1} h_{1.j\to 1}H_{1.j\to 1}X_{1.j}\right) + S_2 \qquad (5.1)\\ Y_2 &= h_{1.1\to 2}H_{1.1\to 2}X_{1.1} + \left(\sum_{j=1}^{K_2} h_{2.j\to 2}H_{2.j\to 2}X_{2.j}\right) + Z_2 \\ &\triangleq \left(\sum_{j=1}^{K_2} h_{2.j\to 2}H_{2.j\to 2}X_{2.j}\right) + S_1 \qquad (5.2)\end{aligned}$$

The Gaussian noise vector $Z_i \in \mathcal{CN}(0, I_{N_i})$ is independent of the input signals and the channel gains. The channel side information

$$S_i = h_{i.1\to i'}H_{i.1\to i'}X_{i.1} + Z_{i'} \;\; i \ne i^{'}, i, i^{'} \in \{1,2\} \tag{5.3}$$

includes both the interference from X_i and noise to Rx$i^{'}$. The signal to noise ratio (SNR) and interference to noise ratio (INR) at receiver Rxi are defined as

$$\mathsf{SNR}_{i.j\to i} = P_{i.j}|h_{i.j\to i}|^2 \triangleq \rho^{\alpha_{i.j\to i}} \tag{5.4}$$

$$\mathsf{INR}_{i.1\to i'} = P_{i.1}|h_{i.1\to i'}|^2 \triangleq \rho^{\alpha_{i.1\to i'}}, \tag{5.5}$$

where ρ is a nominal value for SNR and INR.

We shall frequently apply the singular value decomposition (SVD) of the cross link matrix $H_{i.1\to i'}$,

$$H_{i.1\to i'} = U_{i.1\to i'}\Sigma_{i.1\to i'}V_{i.1\to i'}^{\dagger}, \tag{5.6}$$

in the rest of the chapter. The matrix $H_{i.1\to i'}$ has rank $\min\{M_{i.1}, N_{i'}\}$ w.p.1. In the SVD of $H_{i.1\to i'}$, the rectangular diagonal matrix $\Sigma_{i.1\to i'}$ has $N_{i'}$ rows and M_i columns, and $\min\{M_i, N_{i'}\}$ nonzero value on its diagonal, denoted as $\sigma_{i.1\to i',1}$, $\sigma_{i.1\to i',2}$, $\cdots$, $\sigma_{i.1\to i',\min\{M_{i.1},N_{i'}\}}$, which are the singular values of $H_{i.1\to i'}$. The product $\Sigma_{i.1\to i'}^{\dagger}\Sigma_{i.1\to i'}$, which will be frequently referred in the rest of the chapter, is a diagonal matrix of size $M_{i.1} \times M_{i.1}$. The first $\min\{M_{i.1}, N_{i'}\}$ of its diagonal entries are zero, and the rest $(M_{i.1} - N_{i'})^+$ zero. Hence, the matrix $\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}$ has the following structure

$$\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'} = \begin{pmatrix} [\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}]^+ & \\ & \mathbf{0}_{(M_{i.1}-N_{i'})^+} \end{pmatrix}, \tag{5.7}$$

where $[\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}]^+$ is defined to be a square diagonal matrix of size $\min\{M_{i.1}, N_{i'}\} \times \min\{M_{i.1}, N_{i'}\}$ with all the nonzero diagonal values of $\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}$ on its diagonal, i.e.,

$$[\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}]^+ \triangleq \begin{pmatrix} \sigma^2_{i.1\to i',1} & & \\ & \ddots & \\ & & \sigma^2_{i.1\to i',\min\{M_{i.1},N_{i'}\}} \end{pmatrix}. \tag{5.8}$$

The mathematical formulation describing the encoders, decoders, error events, rates, and achievable rate region are consistent with the definitions in [37, Section II-A] and we do not repeat them here for brevity. We denote the capacity region of a MIMO MAC-IC-MAC by $\mathcal{C}$, which is the closure of all achievable rate-tuples of this channel.

5.3 A Constant-Gap-to-Capacity Region

A constant-gap-to-capacity region of a network is an achievable region whose rate tuple $(R_1, \cdots, R_n)$ lies within $(n_1, \cdots n_K)$ bits of the capacity region. The value of n_i is independent of the channel matrices and the SNR and INR of all links. The definition of constant-gap-to-capacity region for MIMO MAC-IC-MAC is stated in Definition 5.1.

Definition 5.1. An achievable rate region of MIMO MAC-IC-MAC is within gap $(n_{\Theta_1}, n_{\Theta_2}) = (n_{1.1}, \ldots, n_{1.K}, n_{2.1}, \cdots, n_{2.K})$ to its capacity region if for any given rate tuple $(R_{\Theta_1}, R_{\Theta_2}) \in \mathcal{C}$, the rate tuple $(\tilde{R}_{\Theta_1}, \tilde{R}_{\Theta_2}) = (R_{1.1} - n_{1.1}, \cdots, R_{1.K} - n_{1.K}, R_{2.1} - n_{2.1}, \cdots, R_{2.K} - n_{2.K})$ lies within the achievable region.

In this section, we first introduce the known results on constant-gap-to-capacity regions for the K-user MIMO MAC and the two-user MIMO IC. We then unify and generalize those results to obtain an explicit achievable rate region and an explicit outer bound for the MIMO MAC-IC-MAC, and show that the two are within a constant gap of each other, and hence, to the capacity region.

5.3.1 The MIMO MAC

In the special case where $\mathsf{INR}_{i.1\to i'} = 0$ for $i = 1, 2$ the MIMO MAC-IC-MAC becomes two decoupled MACs. We hence review here the constant-gap-to-capacity region for the K-user MIMO MAC. Consider the model for the receiver output

$$Y = \sum_{i=1}^{K} H_i X_i + Z$$

where H_i is the channel matrix from i-th transmitter to the receiver and input signal X_i satisfies the power constraint $\mathrm{E}(X_i^\dagger X_i) \le P_i$ for $i \in \{1, \cdots, K\}$. The capacity region of MIMO MAC is the convex closure of the union of achievable rate regions over all admissible input distributions. Earlier works by [9, 48] have shown that Gaussian inputs are sufficient to achieve the capacity region of MIMO MAC and the convex hull operation is not necessary. Let $Q_i = \mathrm{Cov}[X_i]$ be the covariance matrix of input signal X_i. Then, the capacity region is the union of rate regions over all admissible covariance matrices $(Q_1, \cdots, Q_K)$. We present this union region capacity in Fact 5.1. Zero-mean input is assumed since non-zero mean input only contributes to power inefficiency.

Fact 5.1. *[21, Sec III.B] The capacity region of K-user MIMO MAC is*

$$\begin{aligned}\mathcal{C}_{\mathrm{MAC}} = \bigcup_{\substack{\mathrm{Tr}(Q_i)\le P_i\\ \forall i\in\{1,\cdots,K\}}} \Big\{&(R_1, \cdots, R_K) \in \mathbb{R}_+^K : \\ &\sum_{i\in S} R_i \le \log |I_N + \sum_{i\in S} H_i Q_i H_i^\dagger| \\ &\forall S \subseteq \{1, \cdots, K\}\Big\}\end{aligned} \tag{5.9}$$

Determining the optimal covariance matrices $(Q_1, \cdots, Q_K)$ on any boundary point on $\mathcal{C}_{\mathrm{MAC}}$ is a convex optimization problem [48]. An efficient algorithm known as iterative water-filling was found in [48] to solve for the sum capacity.

Next, we specify an explicit inner bound for the MIMO MAC and demonstrate that this bound lies within a constant gap to its capacity. The result is a special case of [39, Corollary 1]. We take the inner bound to be the one obtained with $Q_i = \frac{1}{M_i} I_{M_i}$ for $i \in \{1, \cdots, K\}$, and obtain a single region inner bound $\mathcal{R}_{\text{MAC,in}}$ of $\mathcal{C}_{\text{MAC}}$, which is

$$\begin{aligned}\mathcal{R}_{\text{MAC,in}} = \big\{&(R_1, \cdots, R_K) \in \mathbb{R}_+^K : \\ &\sum_{i \in S} R_i \leq \log |I_N + \sum_{i \in S} \frac{1}{M_i} H_i H_i^\dagger| \\ &\forall S \subseteq \{1, \cdots, K\}\big\}\end{aligned} \tag{5.10}$$

On the other hand, if we choose each covariance matrix to be identity matrix, i.e., $Q_i = I_{M_i}$ in (5.9), the sum rate bounds of (5.9) will be relaxed due to the fact that $\log\det(\cdot)$ is a monotonically increasing function over the cone of p.s.d. matrices, which yields an explicit outer bound for $\mathcal{C}_{\text{MAC}}$:

$$\begin{aligned}\mathcal{R}_{\text{MAC,o}} = \big\{&(R_1, \cdots, R_K) \in \mathbb{R}_+^K : \\ &\sum_{i \in S} R_i \leq \log |I_N + \sum_{i \in S} H_i H_i^\dagger| \\ &\forall S \subseteq \{1, \cdots, K\}\big\}\end{aligned} \tag{5.11}$$

The gap from $\mathcal{R}_{\text{MAC,in}}$ to $\mathcal{C}_{\text{MAC}}$ cannot exceed the gap from $\mathcal{R}_{\text{MAC,in}}$ to $\mathcal{R}_{\text{MAC,o}}$. Hence, it is sufficient to bound the gap between $\mathcal{R}_{\text{MAC,in}}$ and $\mathcal{R}_{\text{MAC,o}}$ to show that $\mathcal{R}_{\text{MAC,in}}$ is an approximation of the capacity region that is within a constant gap to it.

Remark 5.1. The sets of inequalities in $\mathcal{R}_{\text{MAC,in}}$ and $\mathcal{R}_{\text{MAC,o}}$ have the same structure in that when we write the two in the forms $A_1(R_1, \cdots, R_K)^{\text{T}} \leq b_1$ and $A_2(R_1, \cdots, R_K)^{\text{T}} \leq b_2$, they have the same coefficient matrices, i.e., $A_1 = A_2$. Note for any user subset S, there is a one-to-one correspondence between the partial sum rate restrictions in the inner and outer bounds. Let the bounds of these two partial sum rate restrictions in $\mathcal{R}_{\text{MAC,in}}$ and $\mathcal{R}_{\text{MAC,o}}$ be denoted as $B_S \triangleq \log\left|I_N + \frac{1}{M_i}\sum_{i\in S} H_i H_i^\dagger\right|$ and $\overline{B}_S \triangleq \log\left|I_N + \sum_{i\in S} H_i H_i^\dagger\right|$. Let an upper bound on $\overline{B}_S - B_S$ be denoted as n_S, which we will refer to as a **partial sum rate gap**. Also for clarity, we call a component n_i in a gap $(n_1, \cdots, n_K)$ an **individual rate gap**. To prove a gap between $\mathcal{R}_{\text{MAC,in}}$ and $\mathcal{R}_{\text{MAC,o}}$, we first derive a universal partial sum rate gap n_S for arbitrary S, and then we construct individual rate gap n_i for any $i \in \{1, \cdots, K\}$, such that $\sum_{i\in S} n_i \geq n_S$. If such a individual rate gap can indeed be found, then for any rate tuple $(R_1, \cdots, R_K) \in \mathcal{R}_{\text{MAC,o}}$, the rate tuple

$(\tilde{R}_1, \cdots, \tilde{R}_K) = (R_1 - n_1, \cdots, R_K - n_K)$ satisfies

$$\begin{aligned}\sum_{i \in S} \tilde{R}_i &= \sum_{i \in S} R_i - \sum_{i \in S} n_i \\ &\leq \overline{B}_S - \sum_{i \in S} n_i \\ &\leq \overline{B}_S - n_S \\ &\leq B_S\end{aligned}$$

$\forall S \subseteq \{1, \cdots, K\}$. Thus $(\tilde{R}_1, \cdots, \tilde{R}_K) \in \mathcal{R}_{\text{MAC,in}}$ and $\mathcal{R}_{\text{MAC,in}}$ is within constant gap $(n_1, \cdots, n_K)$ to the capacity region of the MIMO MAC. This technique will be repeated in the proof of the constant gap result for the MIMO MAC-IC-MAC in Section 5.3.6 and Appendix C.3.

Now we follow the idea presented in Remark 5.1 to show $\mathcal{R}_{\text{MAC,in}}$ is a constant-gap-to-capacity region. We first derive a partial sum rate gap n_S as

$$\begin{aligned}&\log \left| I_N + \sum_{i \in S} H_i H_i^\dagger \right| - \log \left| I_N + \sum_{i \in S} \frac{1}{M_i} H_i H_i^\dagger \right| \\ &\leq \min \left\{ \sum_{i \in S} M_i, N \right\} \log \max_{i \in \{1, \cdots, K\}} M_i \\ &\triangleq n_S \end{aligned} \tag{5.12}$$

The inequality holds true because the rank of the matrix $\sum_{i \in S} \frac{1}{M_i} H_i H_i^\dagger$ cannot exceed $\min\{\sum_{i \in S} M_i, N\}$. This result on intra-cell sum rate gap is a specialization of the more general result in [39, Corollary 1] on the constant-gap-to-capacity region of the MIMO MAC with discrete time fading state and general message sets consisting of private and common messages.

Next, we pick the individual rate gap n_i as

$$n_i = \min\{M_i, N\} \log \max_{i \in \{1, \cdots, K\}} M_i. \tag{5.13}$$

It is easy to verify that this individual rate gap indeed guarantees

$$\begin{aligned}\sum_{i \in S} n_i &= \left(\sum_{i \in S} \min\{M_i, N\} \right) \log \max_{i \in \{1, \cdots, K\}} M_i \\ &\geq n_S, \forall S \subseteq \{1, \cdots, K\}.\end{aligned}$$

Note the choice of n_i lets the partial sum rate gap n_S to be distributed onto every associated individual rate gap n_i for $i \in S$. Hence, $\mathcal{R}_{\text{MAC,in}}$ is within $(n_1, \cdots, n_K)$ gap to its outer bound $\mathcal{R}_{\text{MAC,o}}$ and hence its capacity region $\mathcal{C}_{\text{MAC}}$.

5.3.2 The Two-user MIMO IC

The two-user MIMO IC is a special case of the MIMO MAC-IC-MAC where each $K_1 = K_2 = 1$. In the work [27] on the two-user MIMO IC, Karmakar and Varanasi proposed a simple coding scheme (which we will henceforth refer to simply as the KV coding scheme) to get an explicit (i.e., a single region) inner bound [27, Lemma 3] that is within a constant gap to the capacity region [27, Theorem 2]. In this chapter, we obtain a constant-gap-to-capacity region for the MIMO MAC-IC-MAC in Theorem 5.3. In particular, the gap of Theorem 5.3, when specialized to the 2-user MIMO IC (i.e., the (1,1) MIMO MAC-IC-MAC), becomes the gap of [27, Theorem 2] as stated in Remark 5.3.

It must be noted here that the inner bound for the two-user MIMO IC in [27, Lemma 3] and the (1,1) MIMO MAC-IC-MAC specialized from the (K_1, K_2) MIMO MAC-IC-MAC inner bound of Theorem 5.1 in this chapter are obtained using **different** approaches. For the encoding, the authors of [27] derived their inner bound by specifying the coding distribution in the Han-Kobayashi (HK) rate region (an achievable rate region for general two-user IC [23]), whereas we derive our inner bound by specifying coding distribution in our achievable region for the MAC-IC-MAC given in [37, Theorem 1], which in turn is based on the Chong-Motani-Garg (CMG) rate region (another achievable rate region for general two-user IC [10]). For decoding, the work in [27] requires the non-intended common sub-message (see Section 5.3.4) to be decoded uniquely at each receiver, whereas in our work we employ a decoding scheme in which the non-intended common sub-message is decoded non-uniquely.

5.3.3 The MIMO MAC-IC-MAC

The MIMO MAC-IC-MAC is a special case of the semi-deterministic MAC-IC-MAC defined in Section II-B of [37]. Hence, bounds and gap results on the capacity region of semi-deterministic MAC-IC-MAC in [37, Section III-C] can be used. If we apply the inner bound of [37, (46)] (which has the form of a a union of rate regions) for the semi-deterministic MAC-IC-MAC to the MIMO MAC-IC-MAC then Theorem 4 of [37] implies that that inner bound would be within the constant gap of

$$\left(\min\{M_{1.1}, N_2\}\mathbf{1}_{K_1}^T, \min\{M_{2.1}, N_1\}\mathbf{1}_{K_2}^T\right) \tag{5.14}$$

to the capacity region, where $\mathbf{1}_n$ is a column vector of length n with each element being 1.

If we apply another inner bound (which has the form of another union of rate regions) for the semi-deterministic MAC-IC-MAC, namely given by [37, (53)] to the MIMO MAC-IC-MAC then Theorem 5 of [37] implies that that inner bound would the constant gap of

$$\left(\sum_{i\in\{1,2\}} \min\{M_{i.1}, N_{i'}\}\mathbf{1}_{K_1}^T, \sum_{i\in\{1,2\}} \min\{M_{i.1}, N_{i'}\}\mathbf{1}_{K_1}^T\right) \tag{5.15}$$

to the capacity region. The two gaps of (5.14) or (5.15) can thus be achieved by considering one of infinitely many distributions.

It is however important to specify an explicit coding scheme for the MIMO MAC-IC-MAC resulting in an explicit inner bound (in the form of a single polytope) which is within a constant gap to the capacity as was done in [27] for the 2-user MIMO IC and in Section 5.3.1 for the MIMO MAC. For instance, such explicit bounds allow for the evaluation of the GDoF region as was done for the 2-user MIMO IC in [26] or the generalized diversity-multiplexing trade-off as was done in [28] for the 2-user MIMO slow-fading Z interference channel.

In Section 5.3.4, we therefore give an explicit specification of a coding scheme which is based on the KV coding scheme of [27] and the simple single-user coding scheme for the K-user MIMO MAC discussed in Section 5.3.1. This simple scheme yields an explicit inner bound for the MIMO MAC-IC-MAC. An explicit outer bound is then derived in Section 5.3.5 by specifying the input distribution and genie information in the known outer bound for the semi-deterministic MAC-IC-MAC [37, Theorem 3], in the MIMO Gaussian setting. The gap between these explicit inner and outer bounds is shown to be constant, albeit different from either (5.14) or (5.15).

5.3.4 An Explicit Inner Bound

A per-distribution or explicit inner bound for the capacity region of the discrete memoryless (DM) MAC-IC-MAC was given in [37, Theorem 1]. In this section, we apply that result to the MIMO MAC-IC-MAC to obtain the explicit inner bound of Theorem 5.1 to follow. In particular, to prove Theorem 5.1, we must explicitly specify a coding distribution and then compute the set functions in [37, Definition 7] for the MIMO MAC-IC-MAC for that coding distribution. Those set functions are given in the following Definitions, and following that, the inner bound is presented in Theorem 5.1.

Definition 5.2. For any sets $\Omega_i \in 2^{\Theta_i}\backslash\emptyset = \{i.1, \cdots, i.K_i\}$, $\Upsilon_i^{'} \subseteq \Theta_i\backslash\{i.1\}$ and $\Upsilon_i = \Upsilon_i^{'} \cup \{i.1\}$, where $i \in \{1,2\}$, let A_{Υ_i} and E_{Υ_i} be non-negative real-valued functions of set Υ_i, and B_{Ω_i} and G_{Ω_i} be non-negative real-valued functions of set Ω_i. The mappings of set functions A_{Υ_i}, B_{Ω_i}, E_{Υ_i} and G_{Ω_i} are given by (5.16)-(5.19).

$$
\begin{aligned}
A_{\Upsilon_i} \triangleq \log \Bigg| & I_{N_i} + \sum_{i.j \in \Upsilon_i\backslash\{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \frac{1}{M_{i.1}} \rho^{\alpha_{i.1\to i}} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger} \\
& + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1\to i}} H_{i'.1\to i} K_{i'.1} H_{i'.1\to i}^{\dagger} \Bigg| - \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \qquad (5.16) \\
B_{\Omega_i} \triangleq \log \Bigg| & I_{N_i} + \sum_{i.j \in \Omega_i} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1\to i}} H_{i'.1\to i} K_{i'.1} H_{i'.1\to i}^{\dagger} \Bigg| \\
& - \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \qquad (5.17) \\
E_{\Upsilon_i} \triangleq \log \Bigg| & I_{N_i} + \sum_{i.j \in \Upsilon_i\backslash\{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \frac{1}{M_{i.1}} \rho^{\alpha_{i.1\to i}} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger} \\
& + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1\to i}} H_{i'.1\to i} H_{i'.1\to i}^{\dagger} \Bigg| - \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \qquad (5.18) \\
G_{\Omega_i} \triangleq \log \Bigg| & I_{N_i} + \sum_{i.j \in \Omega_i} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1\to i}} H_{i'.1\to i} H_{i'.1\to i}^{\dagger} \Bigg| \\
& - \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \qquad (5.19)
\end{aligned}
$$

Definition 5.3. Let the Cartesian product of the domains of $(\Upsilon_1, \Omega_1, \Upsilon_2, \Omega_2)$ be

$$\Xi \triangleq \left\{\Upsilon_1 \in 2^{\Theta_1} : 1.1 \in \Upsilon_1\right\} \times \left\{2^{\Theta_1}\backslash\emptyset\right\}$$

$$\times \left\{ \Upsilon_2 \in 2^{\Theta_2} : 2.1 \in \Upsilon_2 \right\} \times \left\{ 2^{\Theta_2} \backslash \emptyset \right\}. \tag{5.20}$$

Also, for convenience, define the set function

$$B^{'}_{\Omega_i} \triangleq \begin{cases} \min\{B_{\Omega_i}, A_{\Omega_i} + E_{\Upsilon_{i'}}\} & \text{if } i.1 \in \Omega_i \\ \\ B_{\Omega_i} & \text{if } i.1 \notin \Omega_i \end{cases} \tag{5.21}$$

where $i, i^{'} \in \{1, 2\}$, $i \neq i^{'}$ and $(\Upsilon_1, \Omega_1, \Upsilon_2, \Omega_2) \in \Xi$.

Theorem 5.1. *For MIMO MAC-IC-MAC, any rate tuple $(R_{\Theta_1}, R_{\Theta_2})$ in the following region $\mathcal{R}_{\text{in}}$ is achievable, i.e., $\mathcal{R}_{\text{in}} \subseteq \mathcal{C}$.*

$$\mathcal{R}_{\text{in}} = \Bigg\{ (R_{\Theta_1}, R_{\Theta_2}) \in \mathbb{R}_{+}^{K_1+K_2} :$$

$$\forall (\Upsilon_1, \Omega_1, \Upsilon_2, \Omega_2) \in \Xi$$

$$\sum_{1.j \in \Omega_1} R_{1.j} \leq B^{'}_{\Omega_1} \tag{5.22}$$

$$\sum_{2.j \in \Omega_2} R_{2.j} \leq B^{'}_{\Omega_2} \tag{5.23}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Omega_2} R_{2.j} \leq A_{\Upsilon_1} + G_{\Omega_2} \tag{5.24}$$

$$\sum_{1.j \in \Omega_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq G_{\Omega_1} + A_{\Upsilon_2} \tag{5.25}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq E_{\Upsilon_1} + E_{\Upsilon_2} \tag{5.26}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{1.j \in \Omega_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq A_{\Upsilon_1} + G_{\Omega_1} + E_{\Upsilon_2} \tag{5.27}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} + \sum_{2.j \in \Omega_2} R_{2.j} \leq E_{\Upsilon_1} + A_{\Upsilon_2} + G_{\Omega_2} \Bigg\} \tag{5.28}$$

Proof. As stated at the beginning of this section, we apply the inner bound of [37, Theorem 1] for the DM MAC-IC-MAC to derive the single region inner bound for MIMO MAC-IC-MAC. We first note that the MIMO MAC-IC-MAC inner bound of Theorem 5.1 is described by seven classes of inequalities instead of the nine classes of inequalities in [37, Theorem 1]. This is because the nine classes of inequalities that define the inner bound in [37, Theorem 1] can be expressed as a region with seven classes of inequalities as we show in detail in Appendix C.1. In particular, two pairs of classes of inequalities in the nine classes can be coalesced into two classes of inequalities. The inner bound of [37, Theorem 1] involves set functions $\mathsf{A}_{\Upsilon_i}, \mathsf{B}_{\Omega_i}, \mathsf{E}_{\Upsilon_i}$, and G_{Ω_i} defined in [37, equations (31-34)] for any coding distribution $P_{\text{in}} \in \mathcal{P}_{\text{in}}$ defined in [37, Definition 6].

To obtain a inner bound of Theorem 5.1 from the inner bound of [37, Theorem 1] we need to first explicitly specify a single coding distribution in the set of admissible distributions defined in [37, Definition 6] adapted for the MIMO Gaussian case, allowing for auxiliary and input random variables over continuous alphabets for the MIMO MAC-IC-MAC. This would specify the specific coding scheme and compute the resulting bound which then yield Theorem 5.1.

Before we introduce our coding scheme, we first review the coding scheme used in deriving the bounds for the DM MAC-IC-MAC in [37]. That scheme requires the interfering transmitter Txi.1 to split its message into common and private sub-messages $m_{i.1c}$ and $m_{i.1p}$, respectively. The public sub-message is encoded into cloud codeword $U_1^n(m_{i.1c})$, based on which the private message is superimposed on and the transmitted codeword is $X_{i.1}^n(U_{i.1}^n(m_{i.1c}), m_{i.1p})$. A non-interfering transmitter Tx$i.j$, $j \neq 0$, encodes its entire message $m_{i.j}$ into $X_{i.j}^n(m_{i.j})$ using single user random coding. Rxi decodes its intended message from all the transmitters in its cell, as well as the common message $m_{i'.1c}$ from its non-intended transmitter Txi'.1. Time sharing is employed between all transmitters. The resulting inner bound is presented in [37, Theorem 1], which is a per-distribution inner bound.

Next, we explicitly specify one coding distribution for the MIMO MAC-IC-MAC. First, time sharing is disabled. The interfering transmitter Txi.1 performs KV coding [27], i.e., the transmitted signals $X_{i.1}$ is a sum of two independent Gaussian signals $X_{i.1c}$ and $X_{i.1p}$,

$$X_{i.1} = X_{i.1c} + X_{i.1p} \tag{5.29}$$

where $X_{i.1c} \sim \mathcal{CN}(\mathbf{0}, Q_{i.1c})$ and $X_{i.1p} \sim \mathcal{CN}(\mathbf{0}, Q_{i.1p})$ are signals that carry the public and private messages, respectively, at Txi.1. The covariance matrices $Q_{i.1p}$ and $Q_{i.1c}$ of $X_{i.1c}$ and $X_{i.1p}$ are taken to be

$$Q_{i.1p} = \frac{P_{i.1}}{M_{i.1}} \left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}} H_{i.1\to i'}^{\dagger} H_{i.1\to i'} \right)^{-1}$$

$$\triangleq \frac{P_{i.1}}{M_{i.1}} K_{i.1} \tag{5.30}$$

$$Q_{i.1c} = \frac{P_{i.1}}{M_{i.1}} (I_{M_{i.1}} - K_{i.1}) \tag{5.31}$$

This coding scheme has a rate region that is within constant gap to the capacity region in the 2-user MIMO IC as was established in [27, Theorem 2].

The non-interfering transmitters encode their respective messages using a single-user Gaussian codebook with scaled identity covariances, i.e.,

$$X_{i.j} \sim \mathcal{CN}(\mathbf{0}, \frac{1}{M_{i.j}} I_{M_{i.j}})\, j \neq 1. \tag{5.32}$$

As discussed in Section 5.3.1 such a scheme would achieve a rate region which is within a constant gap to the capacity region in a MIMO MAC.

With the distributions for the inputs specified this way (and with the random variable $X_{i.1c}$ playing the role of the auxiliary random variable $U_{i.1}$) we are now ready to obtain the inner bound of Theorem 5.1 from the inner bound of [37, Theorem 1]. In particular, the set functions $\mathsf{A}_{\Upsilon_i}, \mathsf{B}_{\Omega_i}, \mathsf{E}_{\Upsilon_i}$, and G_{Ω_i} of [37, equations (31-34)] must be evaluated (and lower bounded) for the above coding scheme. Please refer to Appendix C.1 for the details of this evaluation. $\square$

Let us take a deeper look at the KV coding scheme. The choice of the covariance ensures the private message signal $X_{i.1p}$ will be received by Rx$i^{'}$ under the noise level on each of its dimensions,

$$\begin{aligned}
&\rho^{\alpha_{i.1\to i'}} H_{i.1\to i'} Q_{i.1p} H^{\dagger}_{i.1\to i'} \\
&= \frac{1}{M_{i.1}} \rho^{\alpha_{i.1\to i'}} H_{i.1\to i'} \left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}} H^{\dagger}_{i.1\to i'} H_{i.1\to i'} \right)^{-1} \\
&\quad \cdot H^{\dagger}_{i.1\to i'} \\
&\preceq U_{i.1\to i'} \begin{pmatrix} \frac{1}{M_{i.1}} I_{\min\{M_{i.1}, N_{i'}\}} & \\ & \mathbf{0}_{(N_{i'} - M_{i.1})^{+}} \end{pmatrix} U^{\dagger}_{i.1\to i'} && (5.33) \\
&\preceq \frac{1}{M_{i.1}} I_{N_{i'}}. && (5.34)
\end{aligned}$$

Note (5.33) is an upper bound of the received private sub-message signal at Rx$i^{'}$, while (5.34) defines the meaning of noise level in the context of vector input and output signals. When Rx$i^{'}$ has more antennas then the Tx$i.1$ does, the difference of (5.34) and (5.33) shows there are extra dimensions at Rx$i^{'}$ that will not be interfered at all. A detailed proof of received signal of non-intended private sub-message can be found in [27, Appendix B].

Recall the channel matrix $H_{i.1\to i'}$ can be decomposed as $H_{i.1\to i'} = U_{i.1\to i'}\Sigma_{i.1\to i'}V^{\dagger}_{i.1\to i'}$, where $U_{i.1\to i'} \in \mathbb{U}^{N_{i'}\times N_{i'}}$ and $V_{i.1\to i'} \in \mathbb{U}^{M_{i.1}\times M_{i.1}}$ are unitary matrices of size $N_{i'} \times N_{i'}$ and $M_{i.1} \times M_{i.1}$ respectively, matrix $\Sigma_{i.1\to i'} \in \mathbb{C}^{N_{i'}\times M_{i.1}}$ is a rectangular diagonal matrix with all the singular values of $H_{i.1\to i'}$ on its diagonal. The covariance matrices (5.31) and (5.30) of $X_{i.1p}$ and $X_{i.1c}$ can be alternatively written as

$$\begin{aligned}
Q_{i.1p} &= \frac{P_{i.1}}{M_{i.1}}\left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}}H^{\dagger}_{i.1\to i'}H_{i.1\to i'}\right)^{-1} \\
&= \frac{P_{i.1}}{M_{i.1}}\Big(V_{i.j\to i'}V^{\dagger}_{i.j\to i'} \\
&\qquad + \rho^{\alpha_{i.1\to i'}}V_{i.1\to i'}\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'}V^{\dagger}_{i.1\to i'}\Big)^{-1} \\
&\triangleq V_{i.1\to i'}D_{i.1\to i'}V^{\dagger}_{i.1\to i'}
\end{aligned} \tag{5.35}$$

$$\begin{aligned}
Q_{i.1c} &= \frac{1}{M_{i.1}}I_{M_{i.1}} - V_{i.1\to i'}D_{i.1\to i'}V^{\dagger}_{i.1\to i'} \\
&\triangleq V_{i.1\to i'}\tilde{D}_{i.1\to i'}V^{\dagger}_{i.1\to i'}
\end{aligned} \tag{5.36}$$

where the structure of matrices $D_{i.1\to i'}$ and $D_{i.1\to i'}$ are expressed in (5.37) and (5.38).

$$D_{i.1\to i'} = \frac{P_{i.1}}{M_{i.1}}\begin{pmatrix} (I_{\min\{M_{i.1},N_{i'}\}} + \rho^{\alpha_{i.1\to i'}}\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'})^{-1} & \mathbf{0} \\ \mathbf{0} & I_{(M_{i.1}-N_{i'})^+} \end{pmatrix} \tag{5.37}$$

$$\tilde{D}_{i.1\to i'} = \frac{P_{i.1}}{M_{i.1}}\begin{pmatrix} I_{\min\{M_{i.1},N_{i'}\}} - (I_{\min\{M_{i.1},N_{i'}\}} + \rho^{\alpha_{i.1\to i'}}\Sigma^{\dagger}_{i.1\to i'}\Sigma_{i.1\to i'})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(M_{i.1}-N_{i'})^+\times(M_{i.1}-N_{i'})^+} \end{pmatrix} \tag{5.38}$$

Define $\mathtt{X}_{i.1c}, \mathtt{X}_{i.1p}$ be two mutually independent Gaussian vectors with zero mean and identity covariance matrices, of length $\min\{M_{i.1}, N_{i'}\}$ and $M_{i.1}$ respectively. Given $Q_{i.1p}$ and $Q_{i.1c}$ in (5.35) and (5.36), the transmitted signal $X_{i.1}$ can be alternatively expressed in terms of $\mathtt{X}_{i.1c}$ and $\mathtt{X}_{i.1p}$ as

$$\begin{aligned}
X_{i.1} &= X_{i.1c} + X_{i.1p} \\
&= \sum_{k=1}^{\min\{M_{i.1},N_{i'}\}} V^{[k]}_{i.1\to i'}\sqrt{[\tilde{D}_{i.1\to i'}]_{kk}}\,\mathtt{X}^{(k)}_{i.1c}
\end{aligned}$$

$$+ \sum_{k=1}^{\min\{M_{i.1},N_{i'}\}} V_{i.1\to i'}^{[k]} \sqrt{[D_{i.1\to i'}]_{kk}} \mathtt{X}_{i.1p}^{(k)}$$

$$+ \sum_{k=\min\{M_{i.1},N_{i'}\}+1}^{M_{i.1}} V_{i.1\to i'}^{[k]} \sqrt{[D_{i.1\to i'}]_{kk}} \mathtt{X}_{i.1p}^{(k)}. \tag{5.39}$$

In (5.39), $\mathtt{X}_{i.1c}^{(k)}$ denotes the k-th data stream that carries the public sub-message (k-th public data stream) along direction $V_{i.j\to i'}^{[k]}$. It needs to be decoded by both Rxi and Rx$i^{'}$. Since the last $(M_{i.1} - N_{i'})^+$ diagonal values of $\tilde{D}_{i.1\to i'}$ is zero (refer to (5.38)), only $\min\{M_{i.1}, N_{i'}\}$ public data streams exist, which are indicated by the first part of the summation in (5.39). According to (5.38), the transmit power of $\mathtt{X}_{i.1c}^{(k)}$, $1 \le k \le \min\{M_{i.1}, N_{i'}\}$ is

$$[\tilde{D}_{i.1\to i'}]_{kk} = \frac{P_{i.1}}{M_{i.1}} - \frac{P_{i.1}}{M_{i.1}(1+\rho^{\alpha_{i.1\to i'}}\sigma^2_{i.1\to i',k})}. \tag{5.40}$$

From (5.40), we see all public data streams are transmitted at power level ρ^0.

The symbol $\mathtt{X}_{i.1p}^{(k)}$ in (5.39) denotes the k-th data stream that carries the private sub-message (k-th private data stream) along directions $V_{i.j\to i'}^{[k]}$ with power

$$[D_{i.1\to i'}]_{kk}$$
$$= \begin{cases} \frac{P_{i.1}}{M_{i.1}} - [\tilde{D}_{i.1\to i'}]_{kk} & 1 \le k \le \min\{M_{i.1}, N_{i'}\} \\ \frac{P_{i.1}}{M_{i.1}} & \min\{M_{i.1}, N_{i'}\} + 1 \le k \le M_{i.1} \end{cases} \tag{5.41}$$

The first $\min\{M_{i.1}, N_{i'}\}$ private data streams are transmitted at power level $\rho^{-\alpha_{i.1\to i'}}$, so they arrive under the noise floor at Rx$i^{'}$. These private data streams are indicated by the second part of the summation in (5.39). When there are more transmit antennas at Txi than receive antennas at Rx$i^{'}$, beamforming (by matrix $V_{i.1\to i'}$) ensures the extra $(M_{i.1} - N_{i'})^+$ private data streams, indicated by the third part of the summation in (5.39), are sent in the direction of the null space of $H_{i,1\to i'}$. Since these private data streams are not "heard" at Rx$i^{'}$, they are transmitted at power level ρ^0.

The non-interfering signal $X_{i.j}$ can also be rewritten in terms of independent data streams

$$X_{i.j} = \frac{\sqrt{P_{i.1}}}{\sqrt{M_{i.j}}} \mathtt{X}_{i.j} = \sum_{k=1}^{\min\{M_{i.j},N_i\}} \frac{\sqrt{P_{i.1}}}{\sqrt{M_{i.j}}} \mathtt{X}_{i.j}^{(k)}. \tag{5.42}$$

They are only to be received and decoded by the intended receiver Rxi.

5.3.5 An Explicit Outer Bound

Since the MIMO MAC-IC-MAC is semi-deterministic, we use the outer bound for the semi-deterministic MAC-IC-MAC in [37, Theorem 3], which is given in the form of a union of polytopes, to derive an explicit single polytope outer bound for the MIMO MAC-IC-MAC in this section. To specify the outer bound we need to define the relevant set functions in Definition 5.4. The outer bound itself is stated in Theorem 5.2.

Definition 5.4. For any sets $\mathit{\Omega}_i \in 2^{\mathit{\Theta}_i} \backslash \emptyset = \{i.1, \cdots, i.K_i\}$, $\mathit{\Upsilon}_i^{'} \subseteq \mathit{\Theta}_i \backslash \{i.1\}$ and $\mathit{\Upsilon}_i = \mathit{\Upsilon}_i^{'} \cup \{i.1\}$, where $i \in \{1, 2\}$, let $\overline{A}_{\mathit{\Upsilon}_i}$ and $\overline{E}_{\mathit{\Upsilon}_i}$ be non-negative real-valued functions of set $\mathit{\Upsilon}_i$, and $\overline{B}_{\mathit{\Omega}_i}$ and $\overline{G}_{\mathit{\Omega}_i}$ be non-negative real-valued functions of set $\mathit{\Omega}_i$. The mappings of set functions $\overline{A}_{\mathit{\Upsilon}_i}$, $\overline{B}_{\mathit{\Omega}_i}$, $\overline{E}_{\mathit{\Upsilon}_i}$ and $\overline{G}_{\mathit{\Omega}_i}$ are given by (5.43)-(5.46).

$$\overline{A}_{\mathit{\Upsilon}_i} \triangleq \log \left| I_{N_i} + \sum_{i.j \in \mathit{\Upsilon}_i \backslash \{i.1\}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} + \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H_{i.1 \to i}^{\dagger} \right| \tag{5.43}$$

$$\overline{B}_{\mathit{\Omega}_i} \triangleq \log \left| I_{N_i} + \sum_{i.j \in \mathit{\Omega}_i} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} \right| \tag{5.44}$$

$$\overline{E}_{\mathit{\Upsilon}_i} \triangleq \log \left| I_{N_i} + \sum_{i.j \in \mathit{\Upsilon}_i \backslash \{i.1\}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} + \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H_{i.1 \to i}^{\dagger} + \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} H_{i'.1 \to i}^{\dagger} \right| \tag{5.45}$$

$$\overline{G}_{\mathit{\Omega}_i} \triangleq \log \left| I_{N_i} + \sum_{i.j \in \mathit{\Omega}_i} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} + \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} H_{i'.1 \to i}^{\dagger} \right| \tag{5.46}$$

Theorem 5.2. *For the MIMO MAC-IC-MAC, any achievable rate tuple is contained in the following region* $\mathcal{R}_{\mathrm{o}}$, *i.e.,* $\mathcal{C} \subseteq \mathcal{R}_{\mathrm{o}}$, *where*

$$\mathcal{R}_{\mathrm{o}} = \Bigg\{ (R_{\mathit{\Theta}_1}, R_{\mathit{\Theta}_2}) \in \mathbb{R}_+^{K_1+K_2} :$$

$$\forall (\mathit{\Upsilon}_1, \mathit{\Omega}_1, \mathit{\Upsilon}_2, \mathit{\Omega}_2) \in \Xi$$

$$\sum_{1.j \in \mathit{\Omega}_1} R_{1.j} \leq \overline{B}_{\mathit{\Omega}_1} \tag{5.47}$$

$$\sum_{2.j \in \Omega_2} R_{2.j} \leq \overline{B}_{\Omega_2} \tag{5.48}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Omega_2} R_{2.j} \leq \overline{A}_{\Upsilon_1} + \overline{G}_{\Omega_2} \tag{5.49}$$

$$\sum_{1.j \in \Omega_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq \overline{G}_{\Omega_1} + \overline{A}_{\Upsilon_2} \tag{5.50}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq \overline{E}_{\Upsilon_1} + \overline{E}_{\Upsilon_2} \tag{5.51}$$

$$\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{1.j \in \Omega_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} \leq \overline{A}_{\Upsilon_1} + \overline{G}_{\Omega_1} + \overline{E}_{\Upsilon_2} \tag{5.52}$$

$$\left.\sum_{1.j \in \Upsilon_1} R_{1.j} + \sum_{2.j \in \Upsilon_2} R_{2.j} + \sum_{2.j \in \Omega_2} R_{2.j} \leq \overline{E}_{\Upsilon_1} + \overline{A}_{\Upsilon_2} + \overline{G}_{\Omega_2} \right\} \tag{5.53}$$

Proof Outline. The key idea in the proof of the outer bound for the semi-deterministic MAC-IC-MAC in [37, Theorem 3] was to construct a virtual channel whose output is then regarded as genie-aided side information to help each receiver to decode its intended signal (by making it more interference-resilient). To get an outer bound in the MIMO setting, we therefore need to specify the distribution of the genie information T_i to Rxi, which, following [37, Definition 10], should be identical to the distribution of channel side information S_i but independent of S_i conditioned on $X_{i.1}$. We hence choose the genie information to Rxi to be

$$T_i = h_{i.1 \to i'} H_{i.1 \to i'} X_{i.1} + Z^{'}_{i'} \quad i \neq i^{'}, i, i^{'} \in \{1, 2\} \tag{5.54}$$

where $Z^{'}_{i'} \sim \mathcal{CN}(0, I_{N_{i'}})$ and $Z^{'}_{i'}$ is independent of $Z_{i'}$. Hence we obtain an outer bound from [37] by specializing it to the MIMO Gaussian setting using (5.54) and it has the form of a union of polytopes. Secondly, we show that a polytope in that union of polytopes when evaluated with a certain Gaussian input distribution (namely, the extremal distribution) results in a region that subsumes that union-of-polytopes outer bound. That explicit outer bound is shown to be $\mathcal{R}_{\mathrm{o}}$ given in the theorem statement. The details of the proof are relegated to Appendix C.2. ∎

5.3.6 Constant Gap

We can now quantify the gap between the inner bound $\mathcal{R}_{\text{in}}$ and the outer bound $\mathcal{R}_{\text{o}}$. That result is stated next.

Theorem 5.3. *Define the following three parameters*

$$\begin{aligned}\beta_{i.j} = \min\{M_{i.j} + M_{i'.1}, N_i\}& \\ \cdot \log \max &\left\{\max_{i.j\in\Theta_i} M_{i.j}, M_{i'.1}\right\} \\ &+ \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}},\end{aligned} \tag{5.55}$$

$$\begin{aligned}\gamma_{i.j} = \min\{M_{i.j}, N_i\} &\log \max_{i.j\in\Theta_i} M_{i.j} \\ &+ \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \\ &+ \min\{M_{i.1}, N_{i'}\} \log(1+M_{i.1}),\end{aligned} \tag{5.56}$$

and $n_{i.j} = \max\{\beta_{i.j}, \gamma_{i.j}\}$ *for any* $i.j \in \Theta_i$ *and* $i \in \{1, 2\}$*. For any rate tuple* $(R_{\Theta_1}, R_{\Theta_2}) \in \mathcal{R}_{\text{o}}$*, let* $\tilde{R}_{\Theta_i}$ *be the rate tuple*

$$\left((R_{i.1} - n_{i.1})^+, \cdots, (R_{i.K} - n_{i.K})^+\right),$$

then we have

$$(\tilde{R}_{\Theta_1}, \tilde{R}_{\Theta_2}) \in \mathcal{R}_{\text{in}}.$$

Proof Outline. The inequality systems of the inner and outer bounds have the same algebraic structure. The proof is based the idea discussed in Remark 5.1 but with a difference. We do not directly quantify the differences of the right hand side values of 3rd to 7th classes of inequalities in $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$. Instead, we make use of the one-to-one correspondence between the involved intra-cell sum rate term on the left hand side and the set function on the right hand side of each class of inequalities in both bounds, and quantify two **intra-cell sum rate gaps** n_{Υ_i} and n_{Ω_i} regarding user subset Υ_i and Ω_i. Since both A_{Υ_i} and E_{Υ_i} (also $\overline{A}_{\Upsilon_i}$ and $\overline{E}_{\Upsilon_i}$) are functions of Υ_i, the intra-cell sum rate gap n_{Υ_i} should upper bound the maximum value of $\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}$ and $\overline{E}_{\Upsilon_i} - E_{\Upsilon_i}$, i.e., $n_{\Upsilon_i} \geq \max\{\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}, \overline{E}_{\Upsilon_i} - E_{\Upsilon_i}\}$ for any Υ_i. Similarly, we should have

$n_{\Omega_i} \geq \max\{\overline{B}_{\Omega_i} - B'_{\Omega_i}, \overline{G}_i - G_i\}$. We show the individual rate gap $n_{i.j}$ satisfies $\sum_{i.j \in \Upsilon_i} n_{i.j} \geq n_{\Upsilon_i}$ and $\sum_{i.j \in \Omega_i} n_{i.j} \geq n_{\Omega_i}$ at the same time. Please refer to Appendix C.3 for details. □

Remark 5.2. When there are only the non-interfering transmitters in each cell, i.e., $\Theta_i = \{i.2, \cdots, i.K_i\}$, $i \in \{1, 2\}$, a MIMO MAC-IC-MAC is specialized to two parallel MIMO MACs. Let $M_{i.1} = M_{i'.1} = 0$ in (5.56) and (5.55), then the resulting individual rate gap becomes

$$n_{i.j} = \min\{M_{i.j}, N_i\} \log \max_{i.j \in \Theta_i} M_{i.j},$$

which reproduces the individual rate gap for the MIMO MAC given in (5.13).

Remark 5.3. As mentioned previously, when each cell contains only the interfering transmitter, a MIMO MAC-IC-MAC is specialized to a two-user MIMO IC. Let $\Theta_1 = \{1.1\}$ and $\Theta_2 = \{2.1\}$ in (5.55) and (5.56) , then we have

$$\begin{aligned}\beta_{i.1} = \min\{M_{i.1} + M_{i'.1}, N_i\}& \\ &\cdot \log \max\{M_{i.1}, M_{i'.1}\} \\ &+ \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}}\end{aligned}$$

and

$$\begin{aligned}\gamma_{i.1} = \min\{M_{i.1}, N_i\} \log M_{i.1}& \\ &+ \min\{M_{i.1}, N_{i'}\} \log(1 + M_{i.1}) \\ &+ \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}},\end{aligned}$$

which reproduces the gap presented in [27, Theorem 2]. As mentioned in Section 5.3.2, despite the same gap result and the same covariance matrices used for the common and private sub-messages, the encoding and decoding procedures in deriving the constant-gap-to-capacity region for (1, 1) MIMO MAC-IC-MAC is different from the ones used for the two-user MIMO IC.

5.4 GDoF Region of the MIMO MAC-IC-MAC

In this section, we first compute the GDoF region of the MIMO MAC-IC-MAC and then describe the achievability of the key corner points of the GDoF region and the symmetric GDoF curve using the multidimensional signal-level partitioning method to provide insight. In what follows, we define $\bar{\alpha}$ as the set of all the SNR and INR exponents, i.e., $\bar{\alpha} = \{\alpha_{1.1\to1}, \cdots \alpha_{1.K_1\to1}, \alpha_{2.1\to2}, \cdots, \alpha_{2.K_2\to2}, \alpha_{1.1\to2}, \alpha_{2.1\to1}\}$.

Definition 5.5. The generalized degrees of freedom (GDoF) region of a (K_1, K_2) MIMO MAC-IC-MAC, the capacity region of which is denoted as $\mathcal{C}$, is defined as

$$\{(d_{\Theta_1}, d_{\Theta_2}) : d_{i.j} = \lim_{\rho\to\infty} \frac{R_{i.j}}{\log\rho}, i.j \in \Theta_i, i \in \{1,2\},$$

$$\text{and}\,(R_{\Theta_1}, R_{\Theta_2}) \in \mathcal{C}\} \tag{5.57}$$

To derive the GDoF region in this section, we must determine $\lim_{\rho\to\infty} \log\det\left(I_u + \sum_{i=1}^{n} \rho^{\alpha_i} H_i H_i^\dagger\right)$ for an arbitrary integer n. In the work on the GDoF region of the two-user MIMO IC [26], this limit was determined for $n = 2$ and $n = 3$. The limit for arbitrary n is stated in Lemma 5.1.

Lemma 5.1. *Let $H_1 \in \mathbb{C}^{u\times u_1}$, $H_2 \in \mathbb{C}^{u\times u_2}$, $\cdots$, $H_n \in \mathbb{C}^{u\times u_n}$ be n channel matrices whose entries are drawn i.i.d. from continuous and unitarily invariant distributions, then for asymptotic ρ*

$$\log\det\left(I_u + \sum_{i=1}^{n} \rho^{a_i} H_i H_i^\dagger\right) \tag{5.58}$$

$$= f(u, (a_1, u_1), \cdots, (a_n, u_n)) \log(\rho) + \mathcal{O}(1) \tag{5.59}$$

where for any $(u, u_1, \cdots, u_n) \in \mathbb{Z}^{+(n+1)}$ and $(a_1, \cdots, a_n) \in \mathbb{R}^n$, the function $f(u, (a_1, u_1), \cdots, (a_n, u_n))$ is defined as

$$\begin{aligned} & f(u, (a_1, u_1), (a_2, u_2), \cdots, (a_n, u_n)) \\ = & \sum_{i=i_1}^{i_n} \left\{ \min\{u, u_{i_1}\} a_{i_1}^+ + \min\left\{(u - u_{i_1})^+, u_{i_2}\right\} a_{i_2}^+ + \cdots \right. \\ & \left. + \min\left\{ \left(u - \sum_{j=1}^{i_{n-1}} u_j\right)^+, u_{i_n} \right\} a_{i_n}^+ \right\} \end{aligned}$$

for indices $\{i_j\}_{j=1}^{n}$ defined such that $a_{i_1} \geq a_{i_2} \geq \cdots \geq a_{i_n}$.

Proof Outline. The proof employs mathematical induction. The details are given in Appendix C.4. □

The result of Lemma 5.1 can be interpreted in an intuitive way. The term $\log\det(I_u + \sum_{i=1}^{n} \rho^{a_i} H_i H_i^\dagger)$ can be approximately viewed as the achievable sum rate of a n-user MIMO MAC. Transmitter Txi_1, that has the strongest SNR, dominates the first $\min\{u, u_{i_1}\}$ dimensions of the receiver's signal space which leads to GDoF $\min\{u, u_{i_1}\}\alpha_{i_1}^{+}$, and then the 2nd strongest transmitter Txi_2 dominates the next $\min\left\{(u - u_{i_1})^+, u_{i_2}\right\}$ dimensions among the remaining ones (if there are any) leading to additional GDoF $\min\left\{(u - u_{i_1})^+, u_{i_2}\right\}\alpha_{i_2}^{+}$, and so on.

5.4.1 The GDoF Region

In this subsection, we present the GDoF region of the MIMO MAC-IC-MAC. We define the relevant set functions for the GDoF region in Definition 5.6, and the GDoF region is characterized in Theorem 5.4.

Definition 5.6. Define the set functions a, b, e, and g as in (5.60)-(5.63).

$$a_{\Upsilon_i} \triangleq f\left(N_i, \bigcup_{i.j\in\Upsilon_i\backslash\{i.1\}} (\alpha_{i.j\to i}, M_{i.j}), \left((\alpha_{i.1\to i} - \alpha_{i.1\to i'})^+, \min\{M_{i.1}, N_{i'}\}\right), \left(\alpha_{i.1\to i}, (M_{i.1} - N_{i'})^+\right)\right) \tag{5.60}$$

$$b_{\Omega_i} \triangleq f\left(N_i, \bigcup_{i.j\in\Omega_i} (\alpha_{i.j\to i}, M_{i.j})\right) \tag{5.61}$$

$$e_{\Upsilon_i} = f\left(N_i, \bigcup_{i.j\in\Upsilon_i\backslash\{i.1\}} (\alpha_{i.j\to i}, M_{i.j}), \left((\alpha_{i.1\to i} - \alpha_{i.1\to i'})^+, \min\{M_{i.1}, N_{i'}\}\right), \left(\alpha_{i.1\to i}, (M_{i.1} - N_{i'})^+\right), (\alpha_{i'.1\to i}, M_{i'.1})\right) \tag{5.62}$$

$$g_{\Omega_i} \triangleq f\left(N_i, \bigcup_{i.j\in\Omega_i} (\alpha_{i.j\to i}, M_{i.j}), (\alpha_{i'.1\to i}, M_{i'.1})\right) \tag{5.63}$$

Theorem 5.4. *The GDoF region of the MIMO MAC-IC-MAC is the following polytope*

$$\mathcal{D}(a,b,e,g) = \Bigg\{(d_{\Theta_1}, d_{\Theta_2}) \in \mathbb{R}_+^{K_1+K_2} :$$

$$\forall(\Upsilon_1, \Omega_1, \Upsilon_2, \Omega_2) \in \Xi$$

$$\sum_{1.j\in\Omega_1} d_{1.j} \le b_{\Omega_1} \tag{5.64}$$

$$\sum_{2.j\in\Omega_2} d_{2.j} \le b_{\Omega_2} \tag{5.65}$$

$$\sum_{1.j\in\Upsilon_1} d_{1.j} + \sum_{2.j\in\Omega_2} d_{2.j} \le a_{\Upsilon_1} + g_{\Omega_2} \tag{5.66}$$

$$\sum_{1.j\in\Omega_1} d_{1.j} + \sum_{2.j\in\Upsilon_2} d_{2.j} \le g_{\Omega_1} + a_{\Upsilon_2} \tag{5.67}$$

$$\sum_{1.j\in\Upsilon_1} d_{1.j} + \sum_{2.j\in\Upsilon_2} d_{2.j} \le e_{\Upsilon_1} + e_{\Upsilon_2} \tag{5.68}$$

$$\sum_{1.j\in\Upsilon_1} d_{1.j} + \sum_{1.j\in\Omega_1} d_{1.j} + \sum_{2.j\in\Upsilon_2} d_{2.j} \le a_{\Upsilon_1} + g_{\Omega_1} + e_{\Upsilon_2} \tag{5.69}$$

$$\left.\sum_{1.j\in\Upsilon_1} d_{1.j} + \sum_{2.j\in\Upsilon_2} d_{2.j} + \sum_{2.j\in\Omega_2} d_{2.j} \le e_{\Upsilon_1} + a_{\Upsilon_2} + g_{\Omega_2}\right\}. \tag{5.70}$$

Proof Outline. In Definition 5.5, the GDoF region is defined via the capacity region $\mathcal{C}$. While the exact capacity region $\mathcal{C}$ is not known for the MIMO MAC-IC-MAC, Theorem 5.3 states that both $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$ are within constant gap to the capacity. Because a constant number of bits are insignificant in the GDoF computation, the GDoF region can be obtained from either $\mathcal{R}_{\text{in}}$ or $\mathcal{R}_{\text{o}}$. We relegate the detailed proof to Appendix C.5, where we shall directly or indirectly apply Lemma 5.1 on the four outer bound set functions (c.f. Definition 5.4) to derive the result. □

Next, we study the GDoF region of a $(2,1)$ MIMO MAC-IC-MAC with an emphasis on the achievability of the key corner points of its GDoF region using the signal partitioning method.

Example 5.1. Consider the $(2,1)$ MAC-IC-MAC with $M_{1.1} = 3$, $M_{1.2} = 2$, N_1=3, $M_{2.1}$=2, N_2=2, $\alpha_{1.1\to1}$=$\alpha_{1.2\to1}$=$\alpha_{2.1\to2}$=1, and $\alpha_{1.1\to2}$=$\alpha_{2.1\to1}$=0.6. The GDoF region of this channel is the three-dimensional polytope plotted in Fig. 5.2.

We provide an overview of the GDoF region of the MIMO MAC-IC-MAC of Example 5.1. Since the MAC-IC-MAC can be viewed as a generalization of the MAC, the IC and the ZIC, we should be able to observe the GDoF regions of these sub-channels by switching off one particular transmitter. The GDoF of a MAC channel formed by Tx1.1 and Tx1.2 is shown on the $(d_{1.1}, d_{1.2}, 0)$ plane, the GDoF of a two-user interference channel constituted by Tx1.1 and Tx2.1 is shown on the $(d_{1.1}, 0, d_{2.1})$ plane, which is identical to Fig. 2 in [26], and finally, the GDoF region of a Z interference channel is shown on the $(0, d_{1.2}, d_{2.1})$ plane

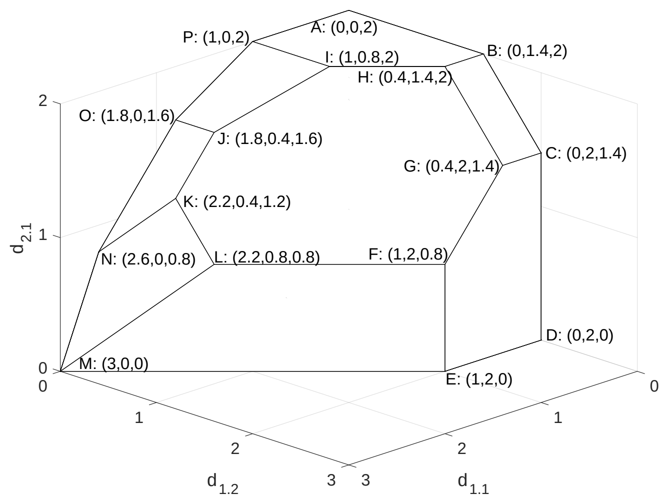

Figure 5.2: The GDoF region of a MIMO MAC-IC-MAC with $K_1 = 2$, $K_2 = 1$, $M_{1.1} = 3$, $M_{1.2} = 2$, $N_1 = 3$, $M_{2.1} = 2$, $N_2 = 2$, $\alpha_{1.1\to1} = \alpha_{1.2\to1} = \alpha_{2.1\to2} = 1$, and $\alpha_{1.1\to2} = \alpha_{2.1\to1} = 0.6$.

in Fig. 5.2. The GDof-tuples on the plane F-G-H-I-J-K-L reach the maximum sum GDoF 3.8, and evidently, we should not switch off any interfering transmitter in order to achieve maximum sum GDoF.

Example 5.2. Continuing with the MIMO MAC-IC-MAC of Example 5.1, we describe here the structures of the transmitted signals from the three transmitters in terms of independent data streams as in (5.39) or (5.42). We assume the transmit power to be unity. In this example, we can accordingly write the signals $X_{1.1}$, $X_{1.2}$ and $X_{2.1}$ as

$$\begin{aligned} X_{1.1} &= \sum_{k=1}^{2} V_{1.1\to2}^{[k]} \sqrt{\frac{\rho^{0.6}\sigma_{1.1\to2,k}^2}{3(1+\rho^{0.6}\sigma_{1.1\to2,k}^2)}}\mathtt{X}_{1.1c}^{(k)} \\ &\quad + \sum_{k=1}^{2} V_{1.1\to2}^{[k]} \sqrt{\frac{1}{3(1+\rho^{0.6}\sigma_{1.1\to2,k}^2)}}\mathtt{X}_{1.1p}^{(k)} \\ &\quad + V_{1.1\to2}^{[3]} \frac{1}{\sqrt{3}}\mathtt{X}_{1.1p}^{(3)} \\ X_{1.2} &= \sum_{k=1}^{2} \frac{1}{\sqrt{2}}\mathtt{X}_{1.2}^{(k)} \end{aligned}$$

and

$$\begin{aligned} X_{2.1} &= \sum_{k=1}^{2} V_{2.1\to1}^{[k]} \sqrt{\frac{\rho^{0.6}\sigma_{2.1\to1,k}^2}{2(1+\rho^{0.6}\sigma_{2.1\to1,k}^2)}}\mathtt{X}_{2.1c}^{(k)} \\ &\quad + \sum_{k=1}^{2} V_{2.1\to1}^{[k]} \sqrt{\frac{1}{2(1+\rho^{0.6}\sigma_{2.1\to1,k}^2)}}\mathtt{X}_{2.1p}^{(k)}. \end{aligned}$$

There are two common and three private data streams from Tx1.1. The common sub-message carried by $\mathtt{X}_{1.1c}^{(1)}$ and $\mathtt{X}_{1.1c}^{(2)}$ will be decoded by both Rx1 and Rx2. The first two private data streams $\mathtt{X}_{1.1p}^{(1)}$ and $\mathtt{X}_{1.1p}^{(2)}$ are decodable by Rx1 but will arrive under the noise floor at Rx2. Since Tx1.1 has one more antenna than Rx2 does, a third private data stream $\mathtt{X}_{1.1p}^{(3)}$ can be sent between Tx1.1 and Rx1 without interfering at Rx2 using transmit zero-forcing beamforming by sending $\mathtt{X}_{1.1p}^{(3)}$ along the null space of $H_{1.1\to 2}$. Hence, we can send $\mathtt{X}_{1.1p}^{(3)}$ at the power level ρ^0. The non-interfering transmitter Tx1.2 has two private streams $\mathtt{X}_{1.2}^{(1)}$ and $\mathtt{X}_{1.2}^{(2)}$ to Rx1 only. Tx2.1 has two data streams $\mathtt{X}_{2.1c}^{(1)}$ and $\mathtt{X}_{2.1c}^{(2)}$ for its common sub-message to be decoded at both Rx1 and Rx2, as well as two private data streams $\mathtt{X}_{2.1p}^{(1)}$ and $\mathtt{X}_{2.1p}^{(2)}$ for its private message to be decoded at Rx2 but under the noise floor at Rx1.

Next, we analyze the achievability of the corner points in the GDoF region of Example 5.1.

The achievability of the 2-user interference channel points A, P, O, N and M have been explored in detail in [26]. The achievability of Points D, E and M can be understood along the lines of achievability of GDoF-tuples in a MIMO MAC, examples of which we have seen previously. The achievability of points A, B, C and D can also be inferred from [26] since the Z interference channel is a special case of the 2-user interference channel. In what follows, we therefore focus on the corner points on the maximum sum GDoF plane. We use the multidimensional signal-level partitioning method in each case to demonstrate the achievability of these GDoF tuples. Note the underlying coding scheme in the multidimensional signal-level partitioning, i.e., multi-level superposition coding as previously mentioned, is different from the coding scheme we used to derive the inner bound.

5.4.1.1 Point I $(d_{1.1}, d_{1.2}, d_{2.1}) = (1, 0.8, 2)$

For this point, we choose the GDoF distribution $d_{1.1c}^{(1)} = d_{1.1c}^{(2)} = 0$, $d_{1.1p}^{(1)} = d_{1.1p}^{(2)} = 0$, $d_{1.1p}^{(3)} = 1$, $d_{1.2}^{(1)} = d_{1.2}^{(2)} = 0.4$, $d_{2.1c}^{(1)} = d_{2.1c}^{(2)} = 0.6$, $d_{2.1p}^{(1)} = d_{2.1p}^{(2)} = 0.4$. The GDoF allocation among the three transmitters can be inferred from the signal partition diagrams at the two receivers in Fig. 5.3. In this allocation, Tx1.1 only uses the private data stream $\mathtt{X}_{1.1p}^{(3)}$ to communicate. This data stream is sent at the power level ρ^0 and along the null space of $H_{1.1\to 2}$ by transmit beamforming. It arrives at Rx1 with power ρ^1, but does not cause interference at Rx2. The coding schemes for Tx1.2 and Tx2.1 can be easily inferred from the GDoF allocation.

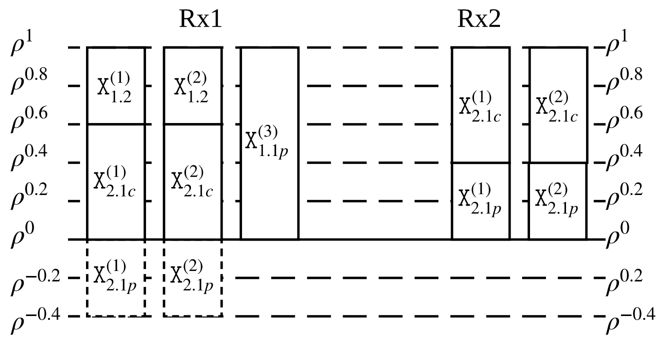

Figure 5.3: GDoF allocation scheme at point I

Rx1 first removes the effect of $\mathtt{X}_{1.1p}^{(3)}$ by zero forcing, i.e., projecting the received signal onto the 2-dimensional signal plane which is perpendicular to $H_{1.1\to 1}V_{1.1\to 2}^{[3]}$. Subsequently, $X_{1.2}$ and $X_{2.1c}$ can be decoded successively. Rx1 recovers $\mathtt{X}_{1.2}^{(1)}$ and $\mathtt{X}_{1.2}^{(2)}$ by treating the $X_{2.1c}$ as noise on its first two dimensions. The noise floor to recover $\mathtt{X}_{1.2}^{(1)}$ and $\mathtt{X}_{1.2}^{(2)}$ will be at $\rho^{0.6}$ and we get $d_{1.2}^{(1)} = d_{1.2}^{(2)} = 0.4$ in this step. Then we recover $\mathtt{X}_{2.1c}^{(1)}$ and $\mathtt{X}_{2.1c}^{(2)}$ resulting $d_{2.1c}^{(1)} = d_{2.1c}^{(2)} = 0.6$. Because $\mathtt{X}_{2.1c}^{(1)}$ and $\mathtt{X}_{2.1c}^{(2)}$ needs to recovered by both Rx1 and Rx2, it remains to check whether we can get $d_{2.1c}^{(1)} = d_{2.1c}^{(2)} = 0.6$ at Rx2 which will be confirmed later. Finally, removing the contributions of $X_{1.2}$ and $X_{2.1c}$ from the received signal, we see an interference-free channel from Tx1.1 to Rx1 so that $\mathtt{X}_{1.1p}^{(3)}$ with GDoF 1 can thus be recovered.

Since there is no interference to Rx2 in this GDoF allocation, Rx2 simply decodes $X_{2.1c}$ and $X_{2.1p}$ successively. The signal $X_{2.1c}$ arrives at Rx2 at power level ρ^1, which can be decoded by treating $X_{2.1p}$, which arrives at power level $\rho^{0.4}$, as noise. We then recover $\mathtt{X}_{2.1c}^{(1)}$ and $\mathtt{X}_{2.1c}^{(2)}$ resulting $d_{2.1c}^{(1)} = d_{2.1c}^{(2)} = 0.6$, the same GDoF obtained from their recovery at Rx1. Lastly, subtracting $X_{2.1c}$ from the received signal, the private message signal $X_{2.1p}$ can be decoded, resulting in $d_{2.1p}^{(1)} = d_{2.1p}^{(2)} = 0.4$.

5.4.1.2 Point J $(d_{1.1}, d_{1.2}, d_{2.1}) = (1.8, 0.4, 1.6)$

Consider the GDoF distribution $d_{1.1c}^{(1)} = d_{1.1c}^{(2)} = 0.2$, $d_{1.1p}^{(1)} = d_{1.1p}^{(2)} = 0.2$, $d_{1.1p}^{(3)} = 1$, $d_{1.2}^{(1)} = d_{1.2}^{(2)} = 0.2$, $d_{2.1c}^{(1)} = d_{2.1c}^{(2)} = 0.4$, $d_{2.1p}^{(1)} = d_{2.1p}^{(2)} = 0.4$, and the GDoF allocation illustrated in Fig. 5.4. Recall that in the simple coding scheme of Section 5.3.4, the non-interfering transmitters transmit at full power. Here however, the GDoF allocation scheme implies the non-interfering transmitter Tx1.2 should transmit at power level $\rho^{0.8}$. In fact, both these two coding schemes achieve Point J. With Tx1.2 transmitting with full power, we need to jointly decode $X_{1.1c}$ and $X_{1.2}$ by treating $X_{2.1c}$ and $X_{1.1p}$ as noise, whereas with Tx1.2 transmitting with power $\rho^{-0.8}$, we can successively decode $X_{1.1c}$, $X_{1.2}$, $X_{2.1c}$ and $X_{1.1p}$ in sequence.

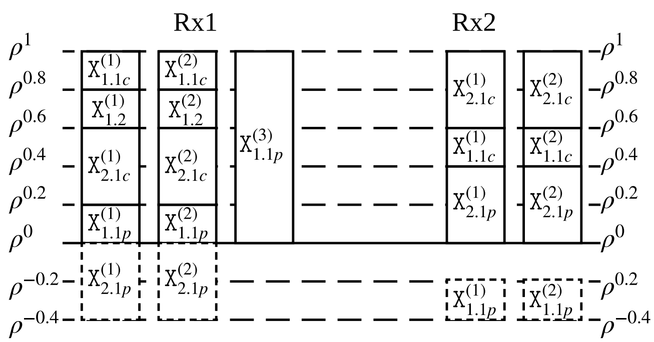

Figure 5.4: GDoF allocation scheme at point J

Rx1 first projects the received signal onto the two-dimensional space perpendicular to $H_{1.1\to 1}V^{[3]}_{1.1\to 2}$ to temporarily get rid of the effect of $\mathtt{X}^{(3)}_{1.1p}$, and then decode $X_{1.1c}$, $X_{1.2}$, $X_{2.1c}$ and $X_{1.1p}$ successively. Lastly, the stream $\mathtt{X}^{(3)}_{1.1p}$ can be recovered after removing the effects of all the other data streams. Rx2 can use successive decoding to decode $X_{2.1c}$, $X_{1.1c}$ and $X_{2.1p}$ sequentially.

5.4.1.3 Point K $(d_{1.1}, d_{1.2}, d_{2.1}) = (2.2, 0.4, 1.2)$

Consider the GDoF distribution $d^{(1)}_{1.1c} = d^{(2)}_{1.1c} = 0.2$, $d^{(1)}_{1.1p} = d^{(2)}_{1.1p} = 0.4$, $d^{(3)}_{1.1p} = 1$, $d^{(1)}_{1.2} = d^{(2)}_{1.2} = 0.2$, $d^{(1)}_{2.1c} = d^{(2)}_{2.1c} = 0.2$, $d^{(1)}_{2.1p} = d^{(2)}_{2.1p} = 0.4$ with the GDoF allocation illustrated in Fig. 5.5. The decoding procedures at Rx1 and Rx2 are similar to what we have done for Point J.

5.4.1.4 Point L $(d_{1.1}, d_{1.2}, d_{2.1}) = (2.2, 0.8, 0.8)$

Consider the GDoF distribution $= d^{(1)}_{1.1c} = d^{(2)}_{1.1c} = 0.2$, $d^{(1)}_{1.1p} = d^{(2)}_{1.1p} = 0.4$, $d^{(3)}_{1.1p} = 1$, $d^{(1)}_{1.2} = d^{(2)}_{1.2} = 0.4$, $d^{(1)}_{2.1c} = d^{(2)}_{2.1c} = 0$, $d^{(1)}_{2.1p} = d^{(2)}_{2.1p} = 0.4$ with the GDoF allocation illustrated in Fig. 5.6.

Rx1 first removes the effect of $\mathtt{X}^{(3)}_{1.1p}$, so $X_{1.1c}$, $X_{1.2}$ and $X_{1.1p}$ can be decoded successively, resulting in $d^{(1)}_{1.1c} = d^{(2)}_{1.1c} = 0.2$, $d^{(1)}_{1.2} = d^{(2)}_{1.2} = 0.4$ and $d^{(1)}_{1.1p} = d^{(2)}_{1.1p} = 0.2$. After that, Rx1 removes the contributions of $X_{1.1c}$, $X_{1.2}$ and $X_{1.1p}$ from the received signal to recover $\mathtt{X}^{(3)}_{1.1p}$.

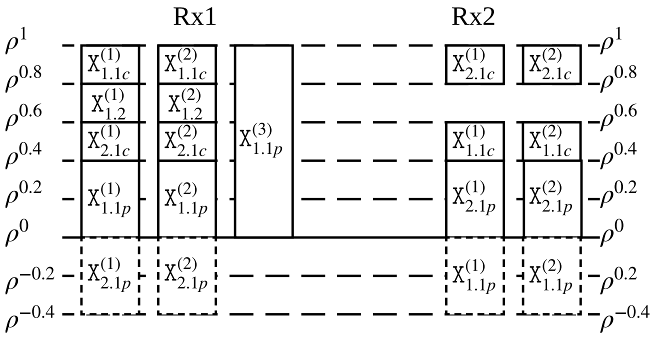

Figure 5.5: GDoF allocation scheme at point K

Rx2 decodes $X_{1.1c}$ and $X_{2.1p}$ using successive cancellation. The equivalent noise floor is $\rho^{0.4}$ to recover $\mathtt{X}_{2.1c}^{(1)}$ and $\mathtt{X}_{2.1c}^{(2)}$, so we have $d_{1.1c}^{(1)} = d_{1.1c}^{(2)} = 0.2$ (the same as we have at Rx1). Lastly, the private message signal $X_{2.1p}$ can be decoded, resulting in $d_{2.1p}^{(1)} = d_{2.1p}^{(2)} = 0.4$.

5.4.1.5 Points F, G and H

We leave the development of the multidimensional signal-partitioning method for these three corner points to the readers.

Remark 5.4. As stated previously, the GDoF region of the $(3, 3, 2, 2)$ two-user MIMO IC is characterized by the curve M-N-O-P-A on the $(d_{1.1}, 0, d_{2.1})$ plane. Comparing points P and I, the DoFs achieved by the two interfering transmitters, i.e., $d_{1.1}$ and $d_{2.1}$ are the same, but at point I the non-interfering transmitter Tx1.2 gets GDoF 0.8 at no reduction of GDoF to Tx1.1 and Tx2.1. The reason is that the signal $X_{2.1c}$ arrives at Rx1 at power level $\rho^{0.6}$, because of which the top two signal partitions in the first two dimensions of Rx1 can be utilized to receive the signals from Tx1.2. Such a phenomenon was first discovered by the authors in the context of the scalar Gaussian MAC-IC-MAC in [37]. It shows that the receivers' power levels are not fully saturated in a two-user MIMO IC under certain channel conditions, and adding non-interfering transmitters (hence, making it a MIMO MAC-IC-MAC) could saturate these power levels (by letting the non-interfering transmitters send signal partitions towards those power levels) and hence improve the overall cell spectrum efficiency. Similar improvement can be observed by comparing points O and J.

Other than more fully occupying a receiver's signal partitions in multiple dimensions, the non-

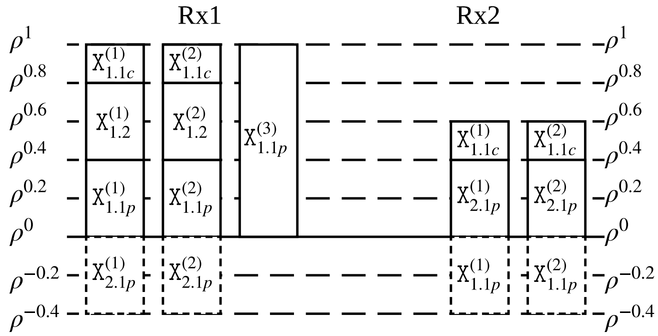

Figure 5.6: GDoF allocation scheme at point L

interfering intended transmitters could also utilize a receiver's signal dimensions that are not used to receive the signals from the intended and non-intended interfering transmitters Tx1.1 and Tx2.1. This is another role the non-interfering transmitters could play in improving the spectrum efficiency. Such an improvement is not seen in Example 5.1, but it is easily understood. Consider a SIMO MAC-IC-MAC which has more than two receive antennas at each receiver. In this case, each receiver has extra signal dimensions after receiving the signals from Tx1.1 and Tx2.1. The non-interfering transmitters (in the same cell), could send their own signals along these extra signal dimensions and henceforth improve the utilization of the available signal partitions and dimensions at each receiver.

5.4.2 The Symmetric GDoF Curve

A MIMO MAC-IC-MAC is said to be symmetric if each cell has the same number of users, all the transmitters (and receivers) have the same number of transmit (receive) antennas M (N) and $\rho_{i.j\to i} = \rho$ and $\rho_{i.1\to i'} = \rho^{\alpha}$ for any $i.j \in \Theta_i$, $i \neq i^{'}$ and $i, i^{'} \in \{1, 2\}$. In regards to the symmetric MIMO MAC-IC-MAC, a more informative performance metric is the symmetric GDoF, which is a function of K, M, N and α, and is defined as follows.

Definition 5.7. For a symmetric K-user, (M, M, N, N) MIMO MAC-IC-MAC with GDoF region $\mathcal{D}_{sym}(K, M, N, \bar{\alpha})$, the *symmetric generalized degree-of-freedom* $d_{sym}(K, M, N, \alpha)$ is defined as the solution to the following maximization problem

$$d_{sym}(K, M, N, \alpha) \triangleq \max_{\substack{d=d_{1.1}=\cdots=d_{1.K}=d_{2.1}=\cdots=d_{2.K} \\ (d_{\Theta_1}, d_{\Theta_2}) \in \mathcal{D}_{sym}(K,M,N,\bar{\alpha})}} d.$$

Given the GDoF region in Theorem 5.4, the symmetric GDoF of MIMO MAC-IC-MAC can be computed by linear programming.

In order to see the GDoF performance at the cell level, we plot the per-cell sum symmetric GDoF (sum symmetric GDoF for short) Kd_{sym} against the interference strength exponent α, for given K, M and N. The plotted curve is called the per-cell sum symmetric GDoF curve (sum symmetric GDoF curve for short). Figs. 5.7-5.9 demonstrate the sum symmetric GDoF curves for the following three antenna configurations: $M = 1$ and $N = 2$, $M = 2$ and $N = 3$, and $M = 3$ and $N = 4$ respectively, each figure has four curves, one for each $K = 1$, 2, 3 and 4.

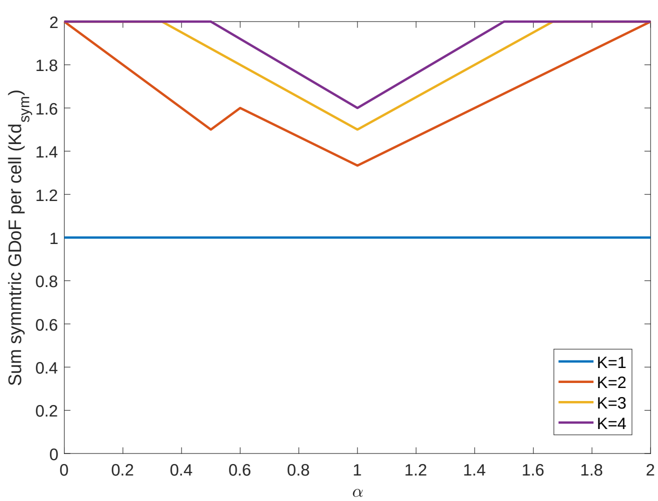

Figure 5.7: Sum symmetric GDoF for $K = 1, 2, 3, 4$, $M = 1$, $N = 2$

Next, we study the sum symmetric GDoFs plotted in Fig. 5.7-5.9 for three distinct symmetric MIMO MAC-IC-MACs described in the figure captions. We focus on the achievability of the corner points on these curves, and the analysis employs the multidimensional signal-level partitioning introduced in Section 2.2.

Example 5.3. $K = 2$, $M = 1$, $N = 2$. This is a two user per cell SIMO (single input multiple output) MAC-IC-MAC. When $\alpha = 0.5$, Fig. 5.7 tells us that $d_{sym} = \frac{3}{4}$ and the sum GDoF per cell of 1.5 is achievable. A GDoF allocation scheme to achieve GDoF 1.5 per cell is illustrated in Fig. 5.10. In this scheme, the private

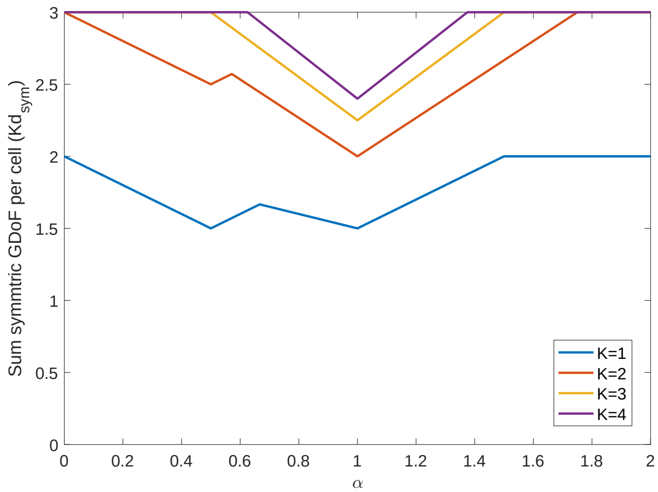

Figure 5.8: Sum symmetric GDoF for $K = 1, 2, 3, 4$, $M = 2$, $N = 3$

data streams $\mathtt{X}_{1.1p}^{(1)}$ and $\mathtt{X}_{2.1p}^{(2)}$ are transmitted with power $\rho^{-0.5}$ so they arrive under the noise floor at their non-intended receivers. The achievability of $d_{sym} = \frac{4}{5}$ when $\alpha = 0.6$ is shown in Fig. 5.11.

Example 5.4. $K = 3$, $M = 1$, $N = 2$. We use this example to show how does sum symmetric GDoF increases with the number of non-interfering transmitters, in comparison to the previous example. The result in Fig. 5.7 suggests that we can achieve full GDoF of 2 per cell at $\alpha = \frac{1}{3}$. A GDoF allocation scheme is illustrated in Fig. 5.12, and the interference free GDoF is achieved in each cell. As can be observed in the figure, the signal partitions of the interfering transmitters are allocated in such a away that the interferences arrive below the noise levels at both the receivers.

Next, we show the achievability of the symmetric GDoF at strong interference with $\alpha = \frac{5}{3}$ which also leads to the achievability of the full GDoF of 2 in each cell. The GDoF allocation of the signal transmission scheme is illustrated in Fig. 5.13.

Example 5.5. $K = 2$, $M = 2$, $N = 3$. In this example, we show how the transmit/receive antenna ratio effects the symmetric GDoF curve, in comparison to Example 5.3 where transmit/receive antenna ratio is 1/2. According to the result in Fig. 5.8, $d_{sym} = \frac{5}{4}$ can be achieved at $\alpha = \frac{1}{2}$. A GDoF allocation that achieves this point is illustrated in Fig. 5.14.

Remark 5.5. As seen from Figs. 5.7-5.9, the per-cell sum symmetric GDoF curve of the MIMO MAC-IC-

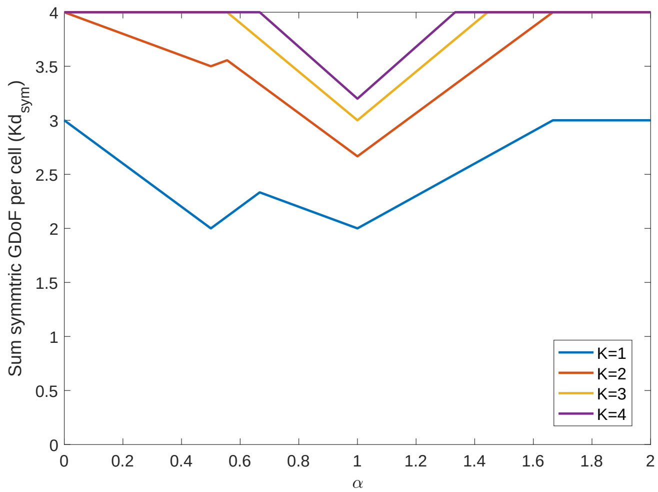

Figure 5.9: Sum symmetric GDoF for $K = 1, 2, 3, 4$, $M = 3$, $N = 4$

MAC moves up as the number of non-interfering transmitters increases. At first glance, this improvement is somehow expected, because with the increasing number of the non-interfering transmitters in a cell, the interfering transmitter has to generally transmit less and therefore emits less interference to the other cell. Hence, the overall cell spectrum efficiency could rise, and an easy way to achieve such improvement is by time-sharing. What is more interesting here is that under certain ranges of α, the interfering transmitter alone cannot fully utilize the spectrum resource as seen in the use of the receivers' signal partitions and dimensions in a cell. However, the interfering and the non-interfering transmitters together, can. The resulting improvement to the sum symmetric GDoF is in general more than what time-sharing can alone achieve. Recall Remark 5.4, the non-interfering transmitter gains positive GDoF at no cost to the two interfering transmitters, and it is such a hidden benefit that we were interested in exploiting in this chapter. Similar GDoF gain can be seen in the sum symmetric GDoF curve too. The most obvious observation is that the full (interference free) GDoF per cell can be achieved ufor certain ranges of α in Fig. 5.7-5.9, whereas time-sharing between the interfering and the non-interfering transmitters cannot achieve full GDoF in those ranges of α. A comparison between the optimal sum symmetric GDoF curve and the sum GDoF curve obtained by several time-sharing schemes has been discussed in the analysis of the per-cell sum symmetric GDoF curve of the Gaussian scalar MAC-IC-MAC in [37, Section III.G].

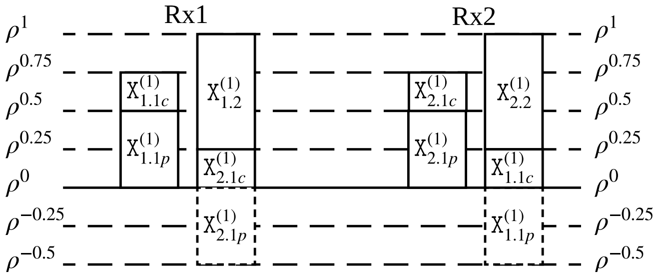

Figure 5.10: GDoF allocation when $K = 2$, $M = 1$, $N = 2$, $\alpha = 0.5$

5.5 Conclusions

The known results on the constant-gap-to-capacity regions of the K-user MIMO MAC [39] and the two-user MIMO IC [27] are generalized and unified in this chapter. In particular, we generalize the coding schemes in [27] and [39] and introduce a simple coding scheme for MIMO MAC-IC-MAC. The resulting achievable region turns out to be within constant gap to the capacity region. The multidimensional signal-level partitioning is formally established and is shown to be a simple and straightforward tool to analyze the achievability of any given GDoF tuple for general MIMO networks. The GDoF region of the MIMO MAC-IC-MAC is characterized. The role of non-interfering transmitters in the MAC-IC-MAC, which has been previously investigated in the Gaussian scalar MAC-IC-MAC [37], is further studied with a variety of antenna configurations. In particular, the per-cell sum symmetric GDoF shows the improvement of spectrum efficiency with the number of the non-interfering transmitters in the MIMO MAC-IC-MAC.

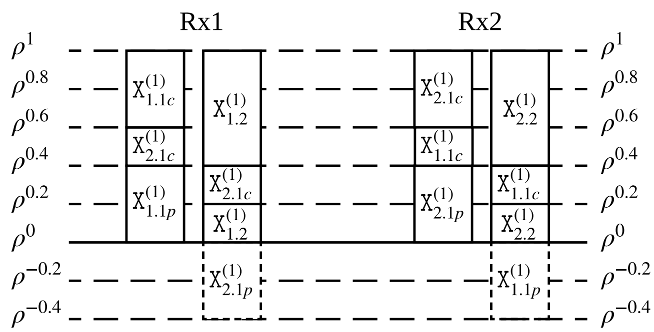

Figure 5.11: GDoF allocation when $K = 2$, $M = 1$, $N = 2$, $\alpha = 0.6$

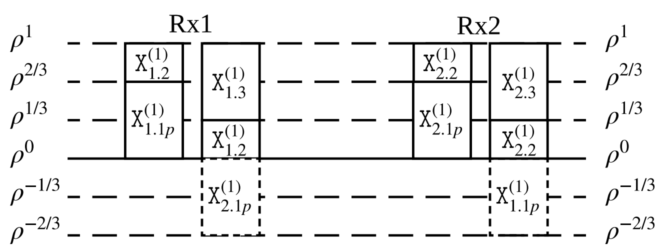

Figure 5.12: GDoF allocation when $K = 3$, $M = 1$, $N = 2$, $\alpha = 1/3$

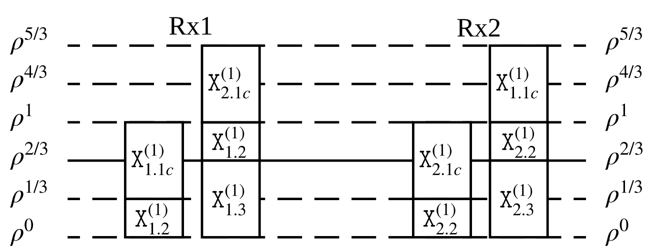

Figure 5.13: GDoF allocation when $K = 3$, $M = 1$, $N = 2$, $\alpha = 5/3$

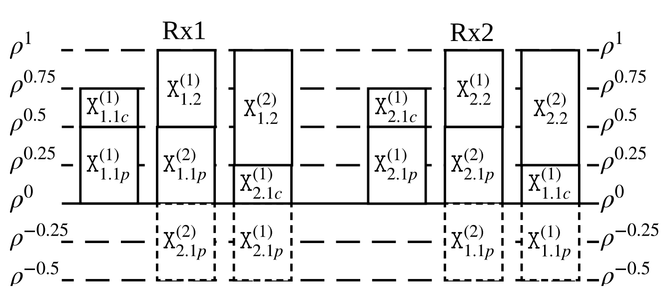

Figure 5.14: GDoF allocation when $K = 2$, $M = 2$, $N = 3$, $\alpha = 0.5$

Chapter 6

Future Research

Now that the GDoF regions of the MIMO one-to-three IC and the MIMO IC-ZIC have been characterized, the next step towards the GDoF region of the fully connected three-user MIMO IC could be obtaining the GDoF region of the MIMO three-to-one IC which is shown in Fig. 6.1. It has the following input-output relations.

$$Y_1 = h_{11}H_{11}X_1 + h_{12}H_{21}X_2 + h_{31}H_{31}X_3 + Z_1$$

$$Y_2 = h_{22}H_{22}X_2 + Z_2$$

$$Y_3 = h_{33}H_{33}X_3 + Z_3$$

The random vectors X_i and Y_i, $i \in \{1,2,3\}$ are the channel inputs and outputs, and h_{ij} and H_{ij} are the channel gain and transfer matrix from Txi to Rxj, where $i, j \in \{1,2,3\}$ and some ij pairs do not exist. Rxi only intends to receive the message from Txi. Note in either MIMO one-to-three IC or IC-ZIC, the GDoF

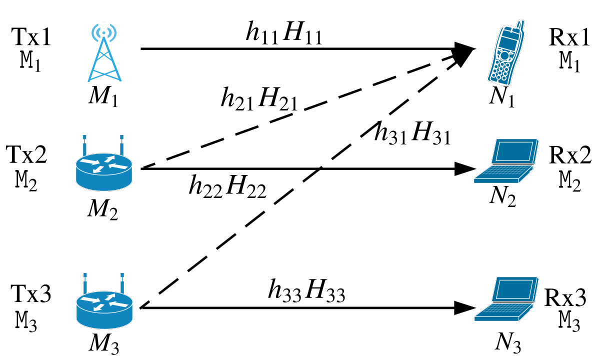

Figure 6.1: The MIMO three-to-one IC

optimal coding scheme does not include interference alignment. This is because in these two channels, no receiver receives more than one interference signal. However, it is not the case in the MIMO three-to-one IC, and interference alignment should be expected. The GDoF region of the MIMO three-to-one IC might be addressed by harnessing insights from the known results on the scalar Gaussian many-to-one IC in [4] and the MIMO one-to-three IC in Chapter 3. We perform GSVD on matrix $H_{21}^{\dagger}$ and $H_{31}^{\dagger}$ such that

$$H_{21}^{\dagger} = U_{21}\Sigma_{21}V^{\dagger} \text{ and } H_{31}^{\dagger} = U_{31}\Sigma_{31}V^{\dagger}$$

where U_{31} and U_{21} are unitary matrices and Σ_{21} and Σ_{31} are rectangular diagonal matrices. Then we write H_{21} and H_{31} as

$$H_{21} = V\Sigma_{21}^{\dagger}U_{21}^{\dagger} \text{ and } H_{31} = V\Sigma_{31}^{\dagger}U_{31}^{\dagger}.$$

This is another form of GSVD which guarantees identical left hand side matrix V in the decomposition. Accordingly, the two interference signals can be expressed as $h_{21}V\Sigma_{21}^{\dagger}U_{21}^{\dagger}X_2$ and $h_{13}V\Sigma_{31}^{\dagger}U_{31}^{\dagger}X_3$. Depending on the diagonal values of $\Sigma_{21}^{\dagger}$ and $\Sigma_{31}^{\dagger}$ as well as the channel gains h_{21} and h_{31}, the interference arriving at Rx1 should contain three parts. The first part consists of the interference signal partitions received along the signal directions and levels (at Rx1) which are only seen by Tx2, the second part consists of the interference signal partitions received along the signal directions and levels which are only seen by Tx3, and the third part consists of the interference signal partitions received along the signal directions and levels which are seen by both Tx2 and Tx3. A possible GDoF optimal coding scheme which adapts this channel structure can be summarized as follows. Tx1 transmits its own message m_1 using single user Gaussian codebook with scaled identity covariance matrix. Tx2 splits its message m_2 into three parts m_{12}, m_{2a} and m_{2p}. Tx3 also splits its message m_3 into three parts m_{13}, m_{3a} and m_{3p}. The sub-messages m_{2p} and m_{3p} are encoded so that they will arrive under the noise floor at Rx1. The sub-message m_{12} is sent to the signal directions and levels (at Rx1) which are exclusively accessible to Tx2, and similarly the sub-message m_{13} is sent at the signal directions and levels which are exclusively accessible to Tx3. Finally, the signal partitions which carry sub-messages m_{2a} and m_{3a} from Tx2 and Tx3 should be aligned with Lattice coding and sent to the signal directions and levels which are accessible to both Tx2 and Tx3. Then Rx1 could take the benefit of interference alignment on those signal directions and levels, and only the sum of aligned interference signal partitions needs to be decoded.

Following the work of the MIMO MAC-IC-MAC, especially the sum symmetric GDoF improvement due to simultaneous transmissions of the interfering and the non-interfering transmitters in each cell, we can already predict the sum symmetric GDoF for a few partially connected symmetric IMACs that have more than one interfering transmitters per cell. Consider a partially connected symmetric IMAC with six transmitters per cell, two of which interfere with the other cell, as shown in Fig. 6.2. Each transmitter is equipped with one antenna and each receiver has two antennas. All the direct links have SNR ρ and all the interference links have INR $\rho^{1/3}$. Recall in Example 5.4, we already know that when $\alpha = 1/3$, a symmetric (3,3) MIMO MAC-IC-MAC with the same antenna configuration could achieve full sum symmetric GDoF 2 (the interference free sum symmetric GDoF) in each cell by the GDoF allocation illustrated in Fig. 5.12. When adding another two non-interfering and one interfering transmitters into the (3,3) MAC-IC-MAC, we let the two added non-interfering transmitters share the signal levels with the existing two non-interfering transmitters in time, and the added interfering transmitter share the signal levels with the existing interfering transmitter in time (see Fig. 6.3). We then get sum symmetric GDoF 2 per-cell which should be the exact per-cell sum symmetric GDoF for this IMAC. Generally speaking, we can scale the number of users in a symmetric MIMO MAC-IC-MAC (while keeping the antenna configuration and the ratio between the numbers of the non-interfering and interfering transmitters the same) to achieve the same per-cell sum symmetric GDoF. Because for any $\alpha \in [0, 1/3]$, a symmetric (3,3) MIMO MAC-IC-MAC could achieve full per-cell sum symmetric GDoF, so could the IMAC in Fig. 6.2. Note the GDoF optimal coding scheme implied by Fig. 6.3 merely treats interference as noise at both receivers. Hence, the MIMO MAC-IC-MAC results could help answer the following question: for a partially connected symmetric IMAC, under what condition is treating interference as noise GDoF optimal? The GDoF optimality condition of treating interference as noise for K-user scalar Gaussian IC with constant channel realization has been determined in [19].

The constant-gap-to-capacity or GDoF region of the MIMO BC-IC-BC, where only one receiver in each cell receives interference from the other cell, may be worth exploring. Unfortunately, the results on the MIMO MAC-IC-MAC do not provide many clues on GDoF optimal coding scheme for the MIMO BC-IC-BC. The known capacity achieving coding scheme for the MIMO BC is the so-called dirty paper coding introduced

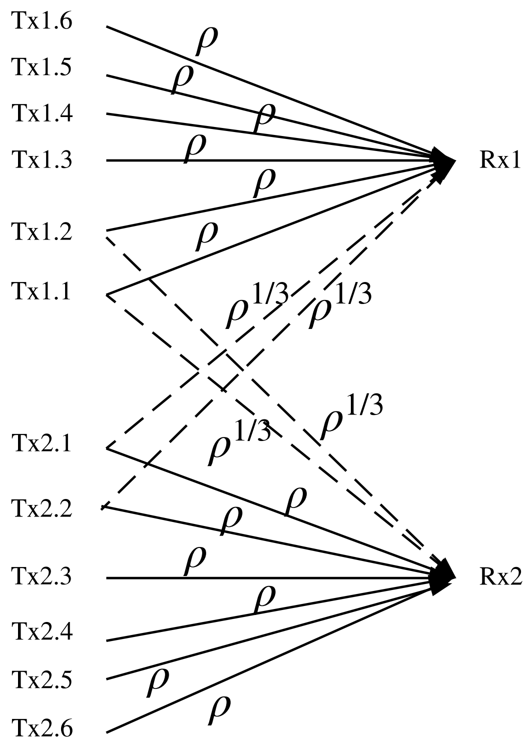

Figure 6.2: A partially connected symmetric IMAC with six transmitters per cell. Two of them interfere with the other cell. Direct links have SNR ρ and interference links have INR $\rho^{1/3}$. Each transmitter has one antenna and each receiver has two antennas.

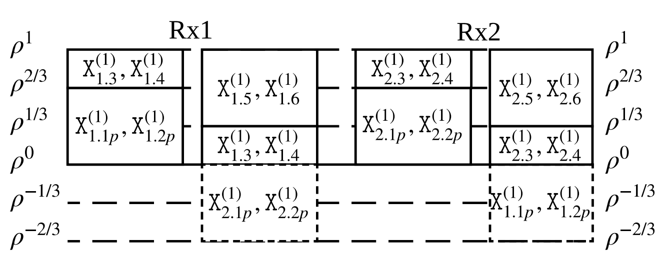

Figure 6.3: A GDoF allocation for the channel given in Fig. 6.2 which achieves full sum symmetric GDoF in each cell. Two signal partitions at the same signal level (separated by comma) share that signal level in time.

by Costa [12], whereas the known GDoF optimal coding scheme for the two-user MIMO IC is the KV coding scheme [27]. It is not clear how these two optimal coding schemes could be combined together to produce a new coding scheme for the MIMO BC-IC-BC. Even for the scalar Gaussian BC-IC-BC, a potential coding scheme is not obvious. One approach to characterize the GDoF region of the MIMO BC-IC-BC might be to establish the duality between the GDoF region of the MIMO MAC-IC-MAC and the MIMO BC-IC-BC.

So far, all the results in Chapters 3-5 are based on constant channel realization. To make these results more practical for wireless applications, channel fading may be incorporated in future research.

Bibliography

[1] Rudolf Ahlswede. Multi-way communication channels. In Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2-8, 1971, 1973.

[2] A Salman Avestimehr, Suhas N Diggavi, and David NC Tse. Wireless network information flow: A deterministic approach. Information Theory, IEEE Transactions on, 57(4):1872–1905, 2011.

[3] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Sukhavasi, Chirag Patel, and Stefan Geirhofer. Network densification: the dominant theme for wireless evolution into 5g. IEEE Communications Magazine, 52(2):82–89, 2014.

[4] Guy Bresler, Abhay Parekh, and NC David. The approximate capacity of the many-to-one and one-to-many gaussian interference channels. IEEE Transactions on Information Theory, 56(9):4566–4592, 2010.

[5] Jörg Bühler and Gerhard Wunder. The multiple access channel interfering with a point to point link: Linear deterministic sum capacity. In Communications (ICC), 2012 IEEE International Conference on, pages 2365–2369. IEEE, 2012.

[6] Viveck R Cadambe and Syed Ali Jafar. Interference alignment and degrees of freedom of the k-user interference channel. Information Theory, IEEE Transactions on, 54(8):3425–3441, 2008.

[7] Aydano Carleial. Interference channels. IEEE Transactions on Information Theory, 24(1):60–70, 1978.

[8] Anas Chaaban and Aydin Sezgin. On the capacity of the 2-user gaussian mac interfering with a p2p link. In Wireless Conference 2011-Sustainable Wireless Technologies (European Wireless), 11th European, pages 1–6. VDE, 2011.

[9] Roger S Cheng and Sergio Verdú. Gaussian multiaccess channels with isi: Capacity region and multiuser water-filling. IEEE Transactions on Information Theory, 39(3):773–785, 1993.

[10] Hon-Fah Chong, Mehul Motani, and Hari Krishna Garg. A comparison of two achievable rate regions for the interference channel. In Proc. Information Theory and Applications Workshop, pages 6–10, 2006.

[11] Hon-Fah Chong, Mehul Motani, Hari Krishna Garg, and H El Gamal. On the han-kobayashi region for the interference channel. IEEE Transactions on Information Theory, 54(7):3188–3194, 2008.

[12] Max Costa. Writing on dirty paper (corresp.). IEEE transactions on information theory, 29(3):439–441, 1983.

[13] Thomas M Cover and Joy A Thomas. Elements of information theory 2nd edition. 2006.

[14] Ersen Ekrem and Sennur Ulukus. An outer bound for the gaussian mimo broadcast channel with common and private messages. IEEE Transactions on Information Theory, 58(11):6766–6772, 2012.

[15] Raul H Etkin, David NC Tse, and Hua Wang. Gaussian interference channel capacity to within one bit. Information Theory, IEEE Transactions on, 54(12):5534–5562, 2008.

[16] Rick Fritschek and Gerhard Wunder. Enabling the multi-user generalized degrees of freedom in the gaussian cellular channel. In Information Theory Workshop (ITW), 2014 IEEE, pages 107–111. IEEE, 2014.

[17] Rick Fritschek and Gerhard Wunder. Upper bounds and duality relations of the linear deterministic sum capacity for cellular systems. In Communications (ICC), 2014 IEEE International Conference on, pages 1884–1889. IEEE, 2014.

[18] Rick Fritschek and Gerhard Wunder. Constant-gap sum-capacity approximation of the deterministic interfering multiple access channel. In Information Theory (ISIT), 2015 IEEE International Symposium on, pages 2643–2647. IEEE, 2015.

[19] Chunhua Geng, Navid Naderializadeh, Amir Salman Avestimehr, and Syed A Jafar. On the optimality of treating interference as noise. Information Theory, IEEE Transactions on, 61(4):1753–1767, 2015.

[20] S. Gherekhloo, A. Chaaban, C. Di, and A. Sezgin. (Sub-)optimality of treating interference as noise in the cellular uplink with weak interference. IEEE Transactions on Information Theory, 62(1):322–356, Jan 2016.

[21] Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity limits of mimo channels. IEEE Journal on selected areas in Communications, 21(5):684–702, 2003.

[22] Tiangao Gou and Syed 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.

[23] Te Sun Han and Kingo Kobayashi. A new achievable rate region for the interference channel. IEEE transactions on information theory, 27(1):49–60, 1981.

[24] Syed A Jafar and Sriram Vishwanath. Generalized degrees of freedom of the symmetric gaussian k user interference channel. IEEE Transactions on Information Theory, 56(7):3297–3303, 2010.

[25] Sang-Woon Jeon and Changho Suh. Degrees of freedom of uplink–downlink multiantenna cellular networks. IEEE Transactions on Information Theory, 62(8):4589–4603, 2016.

[26] Sanjay Karmakar and Mahesh K Varanasi. The generalized degrees of freedom region of the mimo interference channel and its achievability. Information Theory, IEEE Transactions on, 58(12):7188–7203, 2012.

[27] Sanjay Karmakar and Mahesh K Varanasi. The capacity region of the mimo interference channel and its reciprocity to within a constant gap. Information Theory, IEEE Transactions on, 59(8):4781–4797, 2013.

[28] Sanjay Karmakar and Mahesh K Varanasi. The generalized diversity-multiplexing tradeoff of the mimo z interference channel. IEEE Trans. Information Theory, 61(6):3427–3445, 2015.

[29] Taejoon Kim, David J Love, Bruno Clerckx, and Duckdong Hwang. Spatial degrees of freedom of the multicell mimo multiple access channel. In Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE, pages 1–5. IEEE, 2011.

[30] Gerhard Kramer. Review of rate regions for interference channels. In 2006 international Zurich seminar on communications, 2006.

[31] Henry Herng-Jiunn Liao. Multiple Access Channels. PhD thesis, HAWAII UNIV HONOLULU, 1972.

[32] Tingting Liu and Chenyang Yang. Genie tree and degrees of freedom of the symmetric mimo interfering broadcast channel. IEEE Transactions on Signal Processing, 64(22):5914–5929, 2016.

[33] Mehdi Mohseni, Rui Zhang, and John M Cioffi. Optimized transmission for fading multiple-access and broadcast channels with multiple antennas. IEEE Journal on Selected Areas in Communications, 24(8):1627–1639, 2006.

[34] Urs Niesen and Mohammad Ali Maddah-Ali. Interference alignment: From degrees of freedom to constant-gap capacity approximations. Information Theory, IEEE Transactions on, 59(8):4855–4888, 2013.

[35] Christopher C Paige and Michael A Saunders. Towards a generalized singular value decomposition. SIAM Journal on Numerical Analysis, 18(3):398–405, 1981.

[36] Yimin Pang and Mahesh Varanasi. Constant-gap-to-capacity and generalized degrees of freedom regions of the mimo mac-ic-mac. submitted, IEEE Transactions on Information Theory, Dec 18 2018. Also available on Arxiv., 2018.

[37] Yimin Pang and Mahesh Varanasi. A unified theory of multiple-access and interference channels via approximate capacity regions for the mac-ic-mac. IEEE Transactions on Information Theory, 2018.

[38] Peter A Parker, Daniel W Bliss, and Vahid Tarokh. On the degrees-of-freedom of the mimo interference channel. In Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on, pages 62–67. IEEE, 2008.

[39] Henry P Romero and Mahesh K Varanasi. Approximate capacity of the k-user fading mimo mac with common information. In Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on, pages 315–319. IEEE, 2013.

[40] Gokul Sridharan and Wei Yu. Degrees of freedom of mimo cellular networks: Decomposition and linear beamforming design. IEEE Transactions on Information Theory, 61(6):3339–3364, 2015.

[41] Changho Suh and David Tse. Interference alignment for cellular networks. In Communication, Control, and Computing, 2008 46th Annual Allerton Conference on, pages 1037–1044. IEEE, 2008.

[42] Emre Telatar. Capacity of multi-antenna gaussian channels. European transactions on telecommunications, 10(6):585–595, 1999.

[43] Emre Telatar and David Tse. Bounds on the capacity region of a class of interference channels. In Information Theory, 2007. ISIT 2007. IEEE International Symposium on, pages 2871–2874. IEEE, 2007.

[44] J Thomas. Feedback can at most double gaussian multiple access channel capacity (corresp.). IEEE transactions on Information theory, 33(5):711–716, 1987.

[45] Antonia M Tulino, Sergio Verdú, et al. Random matrix theory and wireless communications. Foundations and Trends® in Communications and Information Theory, 3rd edition, (1):1–182, 2004.

[46] Hanan Weingarten, Yossef Steinberg, and Shlomo Shitz Shamai. The capacity region of the gaussian multiple-input multiple-output broadcast channel. IEEE transactions on information theory, 52(9):3936–3964, 2006.

[47] Aaron Wyner. Recent results in the shannon theory. IEEE Transactions on information Theory, 20(1):2–10, 1974.

[48] Wei Yu, Wonjong Rhee, Stephen Boyd, and John M Cioffi. Iterative water-filling for gaussian vector multiple-access channels. IEEE Transactions on Information Theory, 50(1):145–152, 2004.

Appendix A

Proofs for Results on the MIMO One-to-three IC

A.1 Proof of Theorem 3.1

We prove achievability through a random coding argument. We employ three level superposition coding at Tx1, and single use random coding at Tx2 and Tx3. More specifically, the achievable scheme can be described in the following steps.

(1) Generate time sharing sequence q^n according to $p(q^n) = \prod_{t=1}^n p(q_t)$.

(2) Tx1 generates $2^{nR_{123}}$ sequences w_{123}^n according to $p(w_{123}^n|q^n) = \prod_{t=1}^n p(w_{123,t}|q_t)$ and indexes them by $k_{123} \in \{1, \cdots, 2^{nR_{123}}\}$. For each $w_{123}^n(k_{123})$, it generates $2^{nR_{12}}$ sequences w_{12}^n according to $p(w_{12}^n|w_{123}^n(k_{123}), q^n) = \prod_{t=1}^n p(w_{12,t}|w_{123,t}(k_{123}), q_t)$ and indexes then by $(k_{123}, k_{12}) \in \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{12}}\}$ as well as $2^{nR_{13}}$ sequences w_{13}^n according to $p(w_{13}^n|w_{123}^n(k_{123}), q^n) = \prod_{t=1}^n p(w_{13,t}|w_{123,t}(k_{123}), q_t)$ indexed by $(k_{123}, k_{13}) \in \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{13}}\}$. Finally, Tx1 generates $2^{n(R_1 - R_{123} - R_{12} - R_{13})}$ sequences x_1^n according to

$$p(x_1^n|w_{123}^n(k_{123}), w_{12}^n(k_{12}), w_{13}^n(k_{13}), q^n) = \prod_{t=1}^n p(x_{1t}|w_{123,t}(k_{123}), w_{12.t}(k_{12}), w_{13,t}(k_{13}), q_t)$$

and index them as

$$(k_{123}, k_{12}, k_{13}, k_{1p}) \in \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{12}}\} \times \{1, \cdots 2^{nR_{13}}\} \times \{1, \cdots, 2^{n(R_1 - R_{123} - R_{12} - R_{13})}\}.$$

(3) Txi, $i \in \{2, 3\}$, independently generates 2^{nR_i} sequences x_i^n according to $p(x_i^n|q^n) = \prod_{t=1}^n p(x_{it}|q_t)$ and indexes them by $k_i \in \{1, \cdots, 2^{nR_i}\}$.

(4) Once the codebooks are generated, they are fixed for the duration of communication and revealed to receivers Rx1-Rx3.

(5) A 4-tuple message $m_1 = (m_{123}, m_{12}, m_{13}, m_{1p}) = (k_{123}, k_{12}, k_{13}, k_{1p})$ is encoded to $x_1^n(k_{123}, k_{12}, k_{13}, k_{1p})$ at Tx1 and sent over the channel.

(6) A message $m_i = (k_i)$, $i \in \{2, 3\}$ is encoded to $x_i^n(m_i)$ and sent over the channel.

(7) Upon receiving y_1^n, Rx1 declares its decoded messages $(\hat{m}_{123}, \hat{m}_{12}, \hat{m}_{13}, \hat{m}_{1p})$ as the unique index-tuple $(\hat{k}_{123}, \hat{k}_{12}, \hat{k}_{13}, \hat{k}_{1p})$ for which q^n, $w_{123}^n(\hat{k}_{123})$, $w_{12}^n(\hat{k}_{123}, \hat{k}_{12})$, $w_{13}^n(\hat{k}_{123}, \hat{k_{13}})$, $x_1^n(\hat{k}_{123}, \hat{k}_{12}, \hat{k}_{13}, \hat{k}_{1p})$ and y_i^n are jointly typical. If such an index-tuple cannot be found, Rx1 declares an error.

(8) Upon receiving y_i^n, $i \in \{2, 3\}$, Rxi declares its decoded messages $(\hat{m}_{123}, \hat{m}_{1i}, \hat{m}_i)$ as the unique index-tuple $(\hat{k}_{123}, \hat{k}_{1i}, \hat{k}_i)$ for which q^n, $w_{123}^n(\hat{k}_{123})$, $w_{1i}^n(\hat{k}_{123}, \hat{k}_{1i})$, $x_2^n(\hat{k}_i)$ and y_i^n are jointly typical, for some $\hat{k}_{123}$ and $\hat{k}_{1i}$. If such an index-tuple cannot be found, Rxi declares an error.

Suppose $m_1 = (1, 1, 1, 1)$, $m_2 = 1$, and $m_3 = 1$ are sent. The following reliability condition of the coding scheme can be obtained from the typical decoding argument [13, Chapter 7].

$$R_1 - R_{123} - R_{12} - R_{13} \leq I(X_1; Y_1 | W_{123}, W_{12}, W_{13}, Q)$$

$$R_1 - R_{123} - R_{12} \leq I(X_1; Y_1 | W_{123}, W_{12}, Q)$$

$$R_1 - R_{123} - R_{13} \leq I(X_1; Y_1 | W_{123}, W_{13}, Q)$$

$$R_1 - R_{123} \leq I(X_1; Y_1 | W_{123}, Q)$$

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

$$R_2 \leq I(X_2; Y_2 | W_{123}, W_{12}, Q)$$

$$R_2 + R_{12} \leq I(X_2, W_{12}; Y_2 | W_{123}, Q)$$

$$R_2 + R_{123} + R_{12} \leq I(X_2, W_{123}, W_{12}; Y_2 | Q)$$

$$R_3 \leq I(X_3; Y_3 | W_{123}, W_{13}, Q)$$

$$R_3 + R_{13} \leq I(X_3, W_{13}; Y_3 | W_{123}, Q)$$

$$R_3 + R_{123} + R_{13} \leq I(X_3, W_{123}, W_{13}; Y_3 | Q)$$

$$R_{123}, R_{12}, R_{13} \geq 0$$

$$B_{123} + R_{12} + R_{13} \leq R_1$$

$$R_2, R_3 \geq 0$$

Performing Fourier-Motzkin elimination to eliminate R_{123}, R_{12} and R_{13} in the reliability condition, the inner bound can be obtained, which completes the proof.

A.2 Proof of Theorem 3.2

The proof starts from the DM one-to-three inner bound in Theorem 3.1. We evaluate the mutual information terms when specialized to the MIMO setting and for the coding scheme specialized in Section (3.4.2). We prove the fourth inequality (3.63) in $\mathcal{R}_{\text{in}}$ as an example.

According to (3.38), the sum $R_1 + R_2$ is bounded by $I(X_1, Y_1 | W_{123}, W_{12}, Q) + I(X_2, W_{123}, W_{12}; Y_2 | Q)$. The first mutual information term can be evaluated as follows,

$$\begin{aligned}
& I(X_1, Y_1 | W_{123}, W_{12}, Q) \\
&= h(Y_1 | W_{123}, W_{12}) - h(Y_1 | X_1) \\
&= h\left(h_{11} H_{11}(W_{13} + W_{1p}) + Z_1\right) - h(Z_1) \\
&= \log\left(I_{N_1} + \rho^{\alpha_{11}} H_{11}(Q_{13} + Q_{1p}) H_{11}^{\dagger}\right).
\end{aligned}$$

Before we evaluate the second mutual information term, we upper bound the term the term

$$\log\left(I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{12}^{\dagger}\right),$$

which is shown in (A.1), the step (a) is true due to the lower bound on $\mathrm{Tr}(V_p V_p^{\dagger})$ by (3.54). Then the second mutual information term an be computed as following,

$$\begin{aligned}
&I(X_2, W_{123}, W_{12}; Y_2|Q)\\
&= h(Y_2) - h(Y_2|X_2, W_{123}, W_{12})\\
&= h(h_{12}H_{12}X_1 + h_{22}H_{22}X_2 + Z_2)\\
&\quad - h\left(h_{12}H_{12}(W_{13} + W_{1p}) + Z_2\right)\\
&= \log\left(I_{N_2} + \rho^{\alpha_{12}}H_{12}H_{12}^\dagger + \rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right)\\
&\quad - \log\left(I_{N_2} + \rho^{\alpha_{12}}H_{12}(Q_{13} + Q_{1p})H_{12}^\dagger\right)\\
&\geq \log\left(I_{N_2} + \rho^{\alpha_{12}}H_{12}H_{12}^\dagger + \rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right) - \beta_2
\end{aligned}$$

The rest inequalities in $\mathcal{R}_{\text{in}}$ can be proved in a similar fashion. The proof is completed.

$$\begin{aligned}
&\log\left|I_{N_2} + \rho^{\alpha_{12}}H_{12}(Q_{13} + Q_{1p})H_{12}^\dagger\right|\\
&= \log\left|I_{N_2} + \rho^{\alpha_{12}}U_{12}\Sigma_{12}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}^\dagger U^\dagger \frac{V_p}{\mathrm{Tr}(V_pV_p^\dagger)}\right.\\
&\quad\cdot\left.\left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & \mathbf{0}_{(M_1-r)^+\times(M_1-r)^+}\end{pmatrix}\right)^{-1} V_P^\dagger U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}\Sigma_{12}^\dagger U_{12}^\dagger\right|\\
&= \log\left|I_{N_2} + \frac{\rho^{\alpha_{12}}}{\mathrm{Tr}(V_pV_p^\dagger)}\Sigma_{12}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}^\dagger \begin{pmatrix} V_r^{\dagger -1} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+}\end{pmatrix}\right.\\
&\quad\cdot\begin{pmatrix} I_r + \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+}\end{pmatrix}^{-1}\begin{pmatrix} V_r^{-1} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+}\end{pmatrix}\\
&\quad\cdot\left.\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}\Sigma_{12}^\dagger\right|\\
&= \log\left|I_{N_2} + \frac{\rho^{\alpha_{12}}}{\mathrm{Tr}(V_pV_p^\dagger)}\Sigma_{12}\begin{pmatrix} I_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}^\dagger \begin{pmatrix} I_r + \rho^{\alpha_{13}}\Lambda_{13}^\dagger\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12} & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+}\end{pmatrix}^{-1}\right.\\
&\quad\cdot\left.\begin{pmatrix} I_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}\Sigma_{12}^\dagger\right|
\end{aligned}$$

$$
\begin{aligned}
&= \log \left| I_{N_2} + \frac{\rho^{\alpha_{12}}}{\mathrm{Tr}(V_p V_p^\dagger)} \Sigma_{12}(I_r + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13} + \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12})^{-1} \Sigma_{12}^\dagger \right| \\
&= \log \left| I_{N_2} + \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \begin{pmatrix} \frac{\rho^{\alpha_{12}}}{1+\rho^{\alpha_{12}}} I_{r-r_{13}} & & \\ & \rho^{\alpha_{12}} C(I_{r_{123}} + \rho^{\alpha_{13}} S^\dagger S + \rho^{\alpha_{12}} C^\dagger C)^{-1} C^\dagger & \\ & & \mathbf{0}_{(N_2-r_{12})\times(N_2-r_{12})} \end{pmatrix} \right| \\
&\overset{(a)}{\leq} \log \left| \max\left\{ \left(\frac{r}{\lambda_{\max}^2(V_r)} + (M_1 - r)^+ \right)^{-1}, 1 \right\} \right. \\
&\quad \left. \cdot \begin{pmatrix} \left(1 + \frac{\rho^{\alpha_{12}}}{1+\rho^{\alpha_{12}}}\right) I_{r-r_{13}} & & \\ & \left(1 + \frac{\rho^{\alpha_{12}} \sigma_{\max}^2(\Lambda_{12})}{1+\rho^{\alpha_{13}}\sigma_{\min}^2(\Lambda_{13}) + \rho^{\alpha_{12}}\sigma_{\min}^2(\Lambda_{12})}\right) I_{r_{123}} & \\ & & I_{N_2-r_{12}} \end{pmatrix} \right| \\
&\leq \log \left| \max\left\{\zeta_{\min}^{-1}, 1\right\} \begin{pmatrix} 2I_{r-r_{13}} & & \\ & \left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{12})}\right) I_{r_{123}} & \\ & & I_{N_2-r_{12}} \end{pmatrix} \right| \\
&= \log \left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| + (r - r_{13}) + r_{123} \log \left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{12})} \right) \\
&= \beta_2
\end{aligned} \tag{A.1}
$$

A.3 Proof of Theorem 3.3

The outer bound for MIMO one-to-three is characterized in two steps. In the first step, we define the genie information and derive a variety of individual rate upper bounds on R_1, R_2 and R_3. Combing these individual rate restrictions, we characterize an intermediate outer bound with these genie informations and channel side informations S_{12}, S_{13} and S_{123}. It is a union region outer bound over all admissible input distributions. In second step, we optimize the input distribution to be Gaussian and characterize a single region outer bound with this specified distribution.

A.3.1 An Intermediate Outer Bound

We construct three random vectors T_{123}, T_{12} and T_{13} as the genie informations to help receivers to decode their message. They are

$$T_{123} = h_{13}G_{13}X_1 + U_{13}\begin{pmatrix} (U_{13}^{-1}Z_3')^{(1:r_{123})} \\ \mathbf{0}_{(N_3-r_{123})\times 1} \end{pmatrix} \tag{A.2}$$

$$T_{12} = h_{12}H_{12}X_1 + Z_2' \tag{A.3}$$

$$T_{13} = h_{13}J_{13}X_1 + U_{13}\begin{pmatrix} \mathbf{0}_{r_{123}\times 1} \\ (U_{13}^{-1}Z_3')^{(r_{123}+1:N_3)} \end{pmatrix}. \tag{A.4}$$

where $Z_2' \perp Z_2$ and $Z_3' \perp Z_3$ such that T_{123}, T_{12} and T_{13} have identical distribution as the channel side information S_{123}, S_{12} and S_{13} respectively. The basic fact supports the derivation of the individual rate upper bounds to follow is the that providing genie information to the receiver makes the receiver more interference resilient, and therefore should not decrease the capacity of the channel.

If we do not provide any genie information to Rx1, the individual rate R_1 is simply upper bounded by a point-to-point channel capacity

$$nR_1 \leq n\big[h(Y_1|Q) - h(Y_1|X_1,Q)\big] + n\epsilon. \tag{A.5}$$

Providing genie information T_{123}^n to Rx1, we can get another upper bound on the rate R_1 as

$$\begin{aligned}
nR_1 &\overset{(a)}{\leq} I(X_1^n;Y_1^n,T_{123}^n) + n\epsilon \\
&\overset{(b)}{=} I(X_1^n;T_{123}^n) + I(X_1^n;Y_1^n|T_{123}^n) + n\epsilon \\
&\overset{(c)}{=} h(T_{123}^n) - h(T_{123}^n|X_1^n) + h(Y_1^n|T_{123}^n) - h(Y_1^n|X_1^n) \\
&\qquad + n\epsilon \\
&\leq n\big[h(Y_1|T_{123},Q) - h(Y_1|X_1,Q) - h(T_{123}|X_1,Q)\big] \\
&\qquad + h(T_{123}^n) + n\epsilon \\
&\overset{(d)}{=} n\big[h(Y_1|T_{123},Q) - h(Y_1|X_1,Q) - h(S_{123}|X_1,Q)\big]
\end{aligned}$$

$$+ h(S_{123}^n) + n\epsilon. \tag{A.6}$$

The inequalities or equations (a)-(d) hold true because: (a) Providing genie information T_{123}^n to Rx1 will not decrease the channel capacity; (b) chain rule of mutual information; (c) $h(Y_1^n|X_1^n, T_{123}^n) = h(Y_1^n|X_1^n) = h(Z_1^n)$ according to the distribution of X_1, Y_1 and T_{123}; (d) According to the definition of genie information (A.3)-(A.4), we have $h(S_{123}|X_1, Q) = h(T_{123}|X_1, Q)$ and $h(S_{123}^n) = h(T_{123}^n)$.

Similarly, if we provide genie informations (T_{123}^n, T_{12}^n), (T_{123}^n, T_{13}^n) and $(T_{123}^n, T_{12}^n, T_{123}^n)$ we get three more upper bounds on R_1, which are

$$\begin{aligned} nR_1 \le n\big[& h(Y_1|T_{123}, T_{12}, Q) - h(Y_1|X_1, Q) \\ & - h(S_{123}, S_{12}|X_1, Q)\big] + h(S_{123}^n, S_{12}^n) + n\epsilon \end{aligned} \tag{A.7}$$

$$\begin{aligned} nR_1 \le n\big[& h(Y_1|T_{123}, T_{13}, Q) - h(Y_1|X_1, Q) \\ & - h(S_{123}, S_{13}|X_1, Q)\big] + h(S_{123}^n, S_{13}^n) + n\epsilon \end{aligned} \tag{A.8}$$

$$\begin{aligned} nR_1 \le n\big[& h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) \\ & - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] + h(S_{123}^n, S_{12}^n, S_{13}^n) \\ & + n\epsilon. \end{aligned} \tag{A.9}$$

For Rx2 and Rx3, we provide no genie information, genie information T_{123}^n or the entire X_1^n at the decoder, resulting three upper bound on R_2

$$\begin{aligned} nR_2 &\le I(X_2^n; Y_2^n) + n\epsilon \\ &= h(Y_2^n) - h(Y_2^n|X_2^n) + n\epsilon \\ &= h(Y_2^n) - h(S_{12}^n) + n\epsilon \\ &\le nh(Y_2|Q) - h(S_{12}^n) + n\epsilon \end{aligned} \tag{A.10}$$

$$\begin{aligned} nR_2 &\overset{(a)}{\le} I(X_2^n; Y_2^n, T_{123}^n) + n\epsilon \\ &\overset{(b)}{=} I(X_2^n; Y_2^n|T_{123}^n) + n\epsilon \\ &= h(Y_2^n|T_{123}^n) - h(Y_2^n|T_{123}^n, X_2^n) + n\epsilon \\ &= h(Y_2^n|T_{123}^n) - h(S_{12}^n|T_{123}^n) + n\epsilon \end{aligned}$$

$$\leq nh(Y_2|T_{123},Q) - h(S_{12}^n|S_{123}^n) + n\epsilon \tag{A.11}$$

$$nR_2 \overset{(a)}{\leq} I(X_2^n; Y_2^n, X_1^n) + n\epsilon$$

$$\overset{(b)}{=} I(X_2^n; Y_2^n|X_1^n) + n\epsilon$$

$$\leq n\big[h(Y_2|X_1,Q) - h(S_{12}|X_1,Q)\big] + n\epsilon \tag{A.12}$$

and three upper bounds on R_3

$$nR_3 \leq I(X_3^n; Y_3^n) + n\epsilon$$

$$= h(Y_3^n) - h(Y_3^n|X_3^n) + n\epsilon$$

$$\overset{(c)}{=} h(Y_3^n) - h(S_{123}^n, S_{13}^n) + n\epsilon$$

$$\leq nh(Y_3|Q) - h(S_{123}^n, S_{13}^n) + n\epsilon \tag{A.13}$$

$$nR_3 \overset{(a)}{\leq} I(X_3^n; Y_3^n, T_{123}^n) + n\epsilon$$

$$\overset{(b)}{=} I(X_3^n; Y_3^n|T_{123}^n) + n\epsilon$$

$$\overset{(c)}{=} h(Y_3^n|T_{123}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon$$

$$\leq nh(Y_3|T_{123},Q) - h(S_{123}^n, S_{13}^n|S_{123}^n) + n\epsilon \tag{A.14}$$

$$nR_3 \overset{(a)}{\leq} I(X_3^n; Y_3^n, X_1^n) + n\epsilon$$

$$\overset{(b)}{=} I(X_3^n; Y_3^n|X_1^n) + n\epsilon$$

$$\overset{(c)}{=} h(Y_3^n|X_1^n) - h(S_{123}^n, S_{13}^n|X_1^n) + n\epsilon$$

$$\leq n\big[h(Y_3|X_1,Q) - h(S_{123}, S_{13}|X_1,Q)\big] + n\epsilon. \tag{A.15}$$

The steps (a)-(c) hold true because: (a) Providing genie information or X_1^n to Rx2 or Rx3 will not decrease the channel capacity; (b) Chain rule of mutual information and the fact that each genie information is independent of X_i for $i \in \{2,3\}$; (c) The side informations S_{123} and S_{13} are defined to be disjointed (c.f. (3.30)).

Adding (A.7) and (A.10), we obtain an outer bound on sum rate $R_1 + R_2$,

$$n(R_1 + R_2)$$

$$\leq n\big[h(Y_1|T_{123},T_{12},Q) - h(Y_1|X_1,Q) + h(Y_2|Q)$$

$$
\begin{aligned}
&\quad - h(S_{123}, S_{12}|X_1, Q)\big] + h(S^n_{123}, S^n_{12}) - h(S^n_{12}) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, T_{12}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&\quad - h(S_{123}, S_{12}|X_1, Q) + h(S_{123}|S_{12}, Q) + n\epsilon
\end{aligned} \tag{A.16}
$$

Similarly, an outer bound on sum rate $R_1 + R_3$ can be obtained by adding inequalities (A.8) and (A.13),

$$
\begin{aligned}
&n(R_1 + R_3) \\
&\leq n\big[h(Y_1|T_{123}, T_{13}, Q) - h(Y_1|X_1, Q) \\
&\quad + h(Y_3|Q) - h(S_{123}, S_{13}|X_1, Q)\big] + n\epsilon
\end{aligned} \tag{A.17}
$$

Next, we derive two upper bounds on sum rate $R_1 + R_2 + R_3$. Adding (A.9), (A.11) and (A.13), we have

$$
\begin{aligned}
&n(R_1 + R_2 + R_3) \\
&\leq n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|T_{123}, Q) \\
&\quad + h(Y_3|Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&\quad + h(S^n_{123}, S^n_{12}, S^n_{13}) - h(S^n_{12}|S^n_{123}) - h(S^n_{123}, S^n_{13}) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|T_{123}, Q) \\
&\quad + h(Y_3|Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&\quad + h(S^n_{12}|S^n_{123}, S^n_{13}) - h(S^n_{12}|S^n_{123}) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|T_{123}, Q) \\
&\quad + h(Y_3|Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&\quad - I(S^n_{12}; S^n_{13}|S^n_{123}) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|T_{123}, Q) \\
&\quad + h(Y_3|Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] + n\epsilon
\end{aligned} \tag{A.18}
$$

Adding (A.9), (A.10) and (A.14), we have

$$
\begin{aligned}
&n(R_1 + R_2 + R_3) \\
&\leq n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q)
\end{aligned}
$$

$$
\begin{aligned}
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
= n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
\overset{(a)}{=} n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n) \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
\overset{(b)}{\le} n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n, X_1^n) \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
\overset{(c)}{\le} n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - h(T_{123}^n|X_1^n) \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
\overset{(d)}{=} n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q)\big] \\
&- h(T_{123}^n|X_1^n) + h(T_{123}^n|S_{12}^n) \\
&+ h(S_{123}^n, S_{13}^n|S_{12}^n, T_{123}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \\
= n\big[&h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&+ h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q) \\
&- h(T_{123}|X_1, Q)\big] + h(T_{123}^n|S_{12}^n)
\end{aligned}
$$

$$
\begin{aligned}
&- I(S_{123}^n, S_{13}^n; S_{12}^n | T_{123}^n) + n\epsilon \\
&\leq n\big[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q) + h(Y_2|Q) \\
&\quad + h(Y_3|T_{123}, Q) - h(S_{123}, S_{12}, S_{13}|X_1, Q) \\
&\quad - h(S_{123}|X_1, Q) + h(S_{123}|S_{12}, Q)] + n\epsilon
\end{aligned} \tag{A.19}
$$

Inequalities or equations (a)-(d) hold true because: (a) chain rule of conditional differential entropy; (b) conditioning reduces differential entropy; (c) T_{123}^n and $(S_{123}^n, S_{12}^n, S_{13}^n)$ are independent conditioned on X_1^n; (d) chain rule of differential entropy.

Lastly, we generate an upper bound on $2R_1 + R_2 + R_3$, which is obtained by adding (A.6), (A.9), (A.10) and (A.13) in the following.

$$
\begin{aligned}
&n(2R_1 + R_2 + R_3) \\
&\leq n\big[h(Y_1|T_{123}, Q) + h(Y_1|T_{123}, T_{12}, T_{13}, Q) \\
&\quad - 2h(Y_1|X_1, Q) + h(Y_2|Q) + h(Y_3|Q) \\
&\quad - h(S_{123}, S_{12}, S_{13}|X_1, Q) - h(S_{123}|X_1, Q)\big] + h(S_{123}^n) \\
&\quad + h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{12}^n) - h(S_{123}^n, S_{13}^n) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, Q) + h(Y_1|T_{123}, T_{12}, T_{13}, Q) \\
&\quad - 2h(Y_1|X_1, Q) + h(Y_2|Q) + h(Y_3|Q) \\
&\quad - h(S_{123}, S_{12}, S_{13}|X_1, Q) - h(S_{123}|X_1, Q)\big] + h(S_{123}^n) \\
&\quad + h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{12}^n) - h(S_{123}^n, S_{13}^n) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, Q) + h(Y_1|T_{123}, T_{12}, T_{13}, Q) \\
&\quad - 2h(Y_1|X_1, Q) + h(Y_2|Q) + h(Y_3|Q) \\
&\quad - h(S_{123}, S_{12}, S_{13}|X_1, Q) - h(S_{123}|X_1, Q)\big] \\
&\quad + h(S_{123}^n|S_{12}^n) - I(S_{12}^n; S_{13}^n|S_{123}^n) + n\epsilon \\
&= n\big[h(Y_1|T_{123}, Q) + h(Y_1|T_{123}, T_{12}, T_{13}, Q) \\
&\quad - 2h(Y_1|X_1, Q) + h(Y_2|Q) + h(Y_3|Q)
\end{aligned}
$$

$$- h(S_{123}, S_{12}, S_{13}|X_1, Q) - h(S_{123}|X_1, Q)$$

$$+ h(S_{123}|S_{12}, Q)\big] + n\epsilon \tag{A.20}$$

At this point, we can write down an outer bound of MIMO one-to-three IC in terms of side and genie informations by incorporating bounds on individual rate (A.5), (A.12) and (A.15), and on sum rate (A.16)-(A.20). Since the input distribution $p(x_1, x_2, x_3, q)$ is not optimized yet, the resulting outer remains a union of polytopes over all admissible input distributions.

Lemma A.1. *Let $\mathcal{P}_\mathrm{o}$ be the set of distributions P_o of joint random variables (Q, X_1, X_2, X_3) that can be factored as*

$$p(x_1, x_2, x_3) = p(q)p(x_1|q)p(x_2|q)p(x_3|q),$$

and define the following region $\mathcal{R}_\mathrm{o}'(P_\mathrm{o})$.

$$\mathcal{R}_\mathrm{o}'(P_\mathrm{o}) \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq h(Y_1|Q) - h(Y_1|X_1, Q) \tag{A.21}$$

$$R_2 \leq h(Y_2|X_1, Q) - h(S_{12}|X_1, Q) \tag{A.22}$$

$$R_3 \leq h(Y_3|X_1, Q) - h(S_{123}, S_{13}|X_1, Q) \tag{A.23}$$

$$R_1 + R_2 \leq h(Y_1|T_{123}, T_{12}, Q) - h(Y_1|X_1, Q)$$

$$+ h(Y_2|Q) - h(S_{123}, S_{12}|X_1, Q)$$

$$+ h(S_{123}|S_{12}, Q) \tag{A.24}$$

$$R_1 + R_3 \leq h(Y_1|T_{123}, T_{13}, Q) - h(Y_1|X_1, Q)$$

$$+ h(Y_3|Q) - h(S_{123}, S_{13}|X_1, Q) \tag{A.25}$$

$$R_1 + R_2 + R_3 \leq h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q)$$

$$+ h(Y_2|T_{123}, Q) + h(Y_3|Q)$$

$$- h(S_{123}, S_{12}, S_{13}|X_1, Q) \tag{A.26}$$

$$R_1 + R_2 + R_3 \leq h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(Y_1|X_1, Q)$$

$$+ h(Y_2|Q) + h(Y_3|T_{123}, Q)$$

$$
\begin{aligned}
&\quad - h(S_{123}, S_{12}, S_{13}|X_1, Q) \\
&\quad - h(S_{123}|X_1, Q) + h(S_{123}|S_{12}, Q) && \text{(A.27)} \\
2R_1 + R_2 + R_3 &\le h(Y_1|T_{123}, Q) + h(Y_1|T_{123}, T_{12}, T_{13}, Q) \\
&\quad - 2h(Y_1|X_1, Q) + h(Y_2|Q) \\
&\quad - h(S_{123}, S_{12}, S_{13}|X_1, Q) + h(Y_3|Q) \\
&\quad - h(S_{123}|X_1, Q) + h(S_{123}|S_{12}, Q)\big\}. && \text{(A.28)}
\end{aligned}
$$

Then we have

$$\mathcal{C} \subseteq \bigcup_{P_o} \mathcal{R}_o^{'}(P_o).$$

A.3.2 The Single Region Outer Bound

The intermediate upper bound is a union of polytopes over all admissible input distributions P_o. To establish a single region outer bound, we maximize the right hand side values of of inequalities (A.21)-(A.28) by optimizing the input distribution $p(x_1, x_2, x_3, q)$. First of all, the time sharing is disabled. The region $\mathcal{R}_o^{'}(P_o)$ will not shrink because removing random variable Q will not decrease the positive conditional entropy terms and the negative entropy terms are entropies of the Gaussian noises which are independent of Q, for example $h(Y_1|X_1, Q) = h(Z_1|Q) = h(Z_1)$. The positive entropy terms are upper bounded below. Each term reaches its maximum value when X_1, X_2 and X_3 are independent Gaussian random vectors. Because for random vectors X and Y with Zero mean and some fixed joint covariance, the conditional differential entropy of X given Y is maximized when X and Y are joint Gaussian [44, Lemma 1]. We also assumed the inputs have zero mean, i.e., $E(X_i) = \mathbf{0}$ for $i \in \{1, 2, 3\}$, as non-zero means only contribute to power inefficiency.

$$
\begin{aligned}
h(Y_1) &\le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^{\dagger}\right| + N_1 \log 2\pi e \\
h(Y_1|T_{123}) &\le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,13,1p} H_{11}^{\dagger}\right| \\
&\quad + N_1 \log 2\pi e \\
h(Y_1|T_{123}, T_{12}) &\le \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^{\dagger}\right| \\
&\quad + N_1 \log 2\pi e
\end{aligned}
$$

$$h(Y_1|T_{123}, T_{13}) \leq \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,1p} H_{11}^{\dagger}\right| + N_1 \log 2\pi e$$

$$h(Y_1|T_{123}, T_{12}, T_{13}) \leq \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger}\right| + N_1 \log 2\pi e$$

$$h(Y_2) \leq \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}\right| + N_2 \log 2\pi e$$

$$h(Y_2|T_{123}) \leq \log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12} K_{1,12,123} H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}\right| + N_2 \log 2\pi e$$

$$h(Y_2|X_1) \leq \log\left|I_{N_2} + \rho^{\alpha_{22}} H_{22} H_{22}^{\dagger}\right| + N_2 \log 2\pi e$$

$$h(Y_3) \leq \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}\right| + N_3 \log 2\pi e$$

$$h(Y_3|T_{123}) \leq \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} K_{12,13,1p} H_{13}^{\dagger} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}\right| + N_3 \log 2\pi e$$

$$h(Y_3|X_1) \leq \log\left|I_{N_3} + \rho^{\alpha_{33}} H_{33} H_{33}^{\dagger}\right| + N_3 \log 2\pi e$$

$$h(S_{123}|S_{12}) \leq \eta + r_{123} \log(2\pi e)$$

The upper bound of $h(Y_1|T_{123}, T_{12})$ can be obtained in the sequence of steps leading to (A.29). Steps labeled (a)-(c) hold true for the following rationale: (a1)-(a2), the covariance matrix $\mathrm{Cov}[U_{13}(U_{13}^{-1} Z_3)^{(1:r_{123})}]$ satisfies

$$\begin{aligned}
&\mathrm{Cov}[U_{13} U_{13}^{-1(1:r_{123})} Z_3] \\
&= U_{13} U_{13}^{-1(1:r_{123})} I_3 \left(U_{13}^{-1(1:r_{123})}\right)^{\dagger} U_{13}^{\dagger} \\
&= U_{13} \begin{pmatrix} I_{r_{123}} & \mathbf{0}_{r_{123}\times(N_3-r_{123})} \\ \mathbf{0}_{(N_3-r_{123})\times r_{123}} & \mathbf{0}_{(N_3-r_{123})\times N_3-r_{123})} \end{pmatrix} U_{13}^{-1} \\
&\preceq U_{13} I_{N_3} U_{13}^{-1} = I_{N_3}.
\end{aligned}$$

For two p.s.d. matrices A and B, if $A \preceq B$, then $B^{-1} \preceq A^{-1}$ and $-A^{-1} \preceq -B^{-1}$. Therefore, the inverse matrix term with a minus sign will be "greater" if we replace $\mathrm{Cov}[U_{13}(U_{13}^{-1} Z_3)^{(1:r_{123})}]$ with I_{N_3}. Since $\log|\cdot|$

is a monotonically non-decreasing function on the cone of p.s.d. matrices, the value of the entire entropy term will increase after this replacement. (b) follows from the Woodbury's identity. (c) $\mathrm{Tr}(Q_1) \leq P_1$ implies $Q_1 \preceq P_1 I_{M_1}$ and Lemma 6 in [27]. The other positive entropy terms other than $h(S_{123}|S_{12})$ can be upper bounded similarly.

The upper bound of the term $h(S_{123}|S_{12})$ is proved in the sequence of steps leading to (A.29). The rationale for the labeled steps is as follows: step (a) is true because

$$
\begin{aligned}
& h(U_{13}^{-1} S_{123}, U_{12}^{-1} S_{12}) \\
&= h \begin{pmatrix} U_{13}^{-1} S_{123} \\ U_{12}^{-1} S_{12} \end{pmatrix} \\
&= h \begin{pmatrix} U_{13}^{-1} & \\ & U_{12}^{-1} \end{pmatrix} \begin{pmatrix} S_{123} \\ S_{12} \end{pmatrix} \\
&= h \begin{pmatrix} S_{123} \\ S_{12} \end{pmatrix} + \log \begin{vmatrix} U_{13}^{-1} & \\ & U_{12}^{-1} \end{vmatrix} \\
&= h \begin{pmatrix} S_{123} \\ S_{12} \end{pmatrix},
\end{aligned}
$$

step (b) is true due to (3.31).

$$
\begin{aligned}
& h(Y_1|T_{123}, T_{12}) = h(Y_1, T_{123}, T_{12}) - h(T_{123}, T_{12}) \\
&\leq \log \begin{vmatrix} \mathrm{Var}[Y_1] & \mathrm{Cov}[Y_1, T_{123}] & \mathrm{Cov}[Y_1, T_{12}] \\ \mathrm{Cov}[T_{123}, Y_1] & \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, Y_1] & \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{vmatrix} - \log \begin{vmatrix} \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{vmatrix} \\
&\quad + N_1 \log 2\pi e \\
&= \log \left| \mathrm{Var}[Y_1] - \begin{pmatrix} \mathrm{Cov}[Y_1, T_{123}] & \mathrm{Cov}[Y_1, T_{12}] \end{pmatrix} \begin{pmatrix} \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{pmatrix}^{-1} \right.
\end{aligned}
$$

$$
\cdot \begin{pmatrix} \mathrm{Cov}[T_{123}, Y_1] \\ \mathrm{Cov}[T_{12}, Y_1] \end{pmatrix} \Bigg| + N_1 \log 2\pi e
$$

$$
\begin{aligned}
&\overset{(a1)}{\leq} \log \Bigg| I_{N_1} + |h_{11}|^2 H_{11} Q_1 H_{11}^\dagger - \begin{pmatrix} h_{11} h_{13}^* H_{11} Q_1 G_{13}^\dagger & h_{11} h_{12}^* H_{11} Q_1 H_{12}^\dagger \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} U_{13} \begin{pmatrix} I_{r_{123}} & \\ & \mathbf{0}_{(N_3 - r_{123}) \times N_3 - r_{123})} \end{pmatrix} U_{13}^{-1} + |h_{13}|^2 G_{13} Q_1 G_{13}^\dagger & h_{13} h_{12}^* G_{13} Q_1 H_{12}^\dagger \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^\dagger & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^\dagger \end{pmatrix}^{-1} \\
&\quad \cdot \begin{pmatrix} h_{13} h_{11}^* G_{13} Q_1 H_{11}^\dagger \\ h_{12} h_{11}^* H_{12} Q_1 H_{11}^\dagger \end{pmatrix} \Bigg| + N_1 \log 2\pi e \\
&\overset{(a2)}{\leq} \log \Bigg| I_{N_1} + |h_{11}|^2 H_{11} Q_1 H_{11}^\dagger - \begin{pmatrix} h_{11} h_{13}^* H_{11} Q_1 G_{13}^\dagger & h_{11} h_{12}^* H_{11} Q_1 H_{12}^\dagger \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} I_{N_3} + |h_{13}|^2 G_{13} Q_1 G_{13}^\dagger & h_{13} h_{12}^* G_{13} Q_1 H_{12}^\dagger \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^\dagger & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^\dagger \end{pmatrix}^{-1} \begin{pmatrix} h_{13} h_{11}^* G_{13} Q_1 H_{11}^\dagger \\ h_{12} h_{11}^* H_{12} Q_1 H_{11}^\dagger \end{pmatrix} \Bigg| + N_1 \log 2\pi e \\
&= \log \Bigg| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \Bigg[I_{M_1} - \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} I_{N_3} + |h_{13}|^2 G_{13} Q_1 G_{13}^\dagger & h_{13} h_{12}^* G_{13} Q_1 H_{12}^\dagger \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^\dagger & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^\dagger \end{pmatrix}^{-1} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \Bigg] Q_1^{\frac{1}{2}} H_{11}^\dagger \Bigg| + N_1 \log 2\pi e \\
&= \log \Bigg| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \Bigg[I_{M_1} - \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \\
&\quad \cdot \left(I_{N_2+N_3} + \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \right)^{-1} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \Bigg] Q_1^{\frac{1}{2}} H_{11}^\dagger \Bigg| \\
&\quad + N_1 \log 2\pi e \\
&\overset{(b)}{=} \log \left| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \left(I_{M_1} + \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \right)^{-1} Q_1^{\frac{1}{2}} H_{11}^\dagger \right| + N_1 \log 2\pi e \\
&= \log \left| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \left(I_{M_1} + Q_1^{\frac{1}{2}} (|h_{13}|^2 G_{13}^\dagger G_{13} + |h_{12}|^2 H_{12}^\dagger H_{12}) Q_1^{\frac{1}{2}} \right)^{-1} Q_1^{\frac{1}{2}} H_{11}^\dagger \right| + N_1 \log 2\pi e \\
&\overset{(c)}{\leq} \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^\dagger G_{13} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} \right)^{-1} H_{11}^\dagger \right| + N_1 \log 2\pi e \\
&= \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^\dagger \right| + N_1 \log 2\pi e
\end{aligned} \tag{A.29}
$$

$$
\begin{aligned}
&h(S_{123}|S_{12}) = h(S_{123}, S_{12}) - h(S_{12}) \\
&\overset{(a)}{=} h\left(U_{13}^{-1}S_{123}, U_{12}^{-1}S_{12}\right) - h\left(U_{12}^{-1}S_{12}\right) \\
&\leq \log\left|\mathrm{Cov}[U_{13}^{-1}S_{123}] - \mathrm{Cov}[U_{13}^{-1}S_{123}, U_{12}^{-1}S_{12}]\left(\mathrm{Var}[U_{12}^{-1}S_{12}]\right)^{-1}\mathrm{Cov}[U_{12}^{-1}S_{12}, U_{13}^{-1}S_{123}]\right| + r_{123}\log 2\pi e \\
&= \log\Big|I_{r_{123}} + |h_{13}|^2\Lambda_{13}^{(1:r_{123})}V^\dagger Q_1 V\Lambda_{13}^{(1:r_{123})\dagger} - h_{13}h_{12}^\dagger\Lambda_{13}^{(1:r_{123})}V^\dagger Q_1 V\Sigma_{12}^\dagger\left(I_{N_2} + |h_{12}|^2\Sigma_{12}V^\dagger Q_1 V\Sigma_{12}^\dagger\right)^{-1} \\
&\quad \cdot h_{12}h_{13}^\dagger\Sigma_{12}V^\dagger Q_1 V\Lambda_{13}^{(1:r_{123})\dagger}\Big| + r_{123}\log 2\pi e \\
&= \log\Big|I_{r_{123}} + |h_{13}|^2\Lambda_{13}^{(1:r_{123})}V^\dagger Q_1^{\frac{1}{2}}\left[I_{M_1} - |h_{12}|^2 Q_1^{\frac{1}{2}}V\Sigma_{12}^\dagger\left(I_{N_2} + |h_{12}|^2\Sigma_{12}V^\dagger Q_1 V\Sigma_{12}^\dagger\right)^{-1}\Sigma_{12}V^\dagger Q_1^{\frac{1}{2}}\right] \\
&\quad Q_1^{\frac{1}{2}}V\Lambda_{13}^{(1:r_{123})\dagger}\Big| + r_{123}\log 2\pi e \\
&= \log\left|I_{r_{123}} + |h_{13}|^2\Lambda_{13}^{(1:r_{123})}V^\dagger Q_1^{\frac{1}{2}}\left(I_{M_1} + |h_{12}|^2 Q_1^{\frac{1}{2}}V\Sigma_{12}^\dagger\Sigma_{12}V^\dagger Q_1^{\frac{1}{2}}\right)^{-1}Q_1^{\frac{1}{2}}V\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log 2\pi e \\
&\leq \log\left|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}V^\dagger\left(I_{M_1} + \rho^{\alpha_{12}}V\Sigma_{12}^\dagger\Sigma_{12}V^\dagger\right)^{-1}V\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log 2\pi e \\
&= \log\Bigg|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}V^\dagger\left(I_{M_1} + \rho^{\alpha_{12}}U\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}\Sigma_{12}^\dagger\Sigma_{12}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}^\dagger U^\dagger\right)^{-1} \\
&\quad \cdot V\Lambda_{13}^{(1:r_{123})\dagger}\Bigg| + r_{123}\log 2\pi e \\
&= \log\left|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}V^\dagger U\begin{pmatrix} I_r + \rho^{\alpha_{12}}V_r\Sigma_{12}^\dagger\Sigma_{12}V_r^\dagger & \\ & I_{(M_1-r)^+}\end{pmatrix}^{-1}U^\dagger V\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log 2\pi e \\
&\overset{(b)}{\leq} \log\left|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}V^\dagger U\begin{pmatrix} \frac{1}{\lambda_{\max}^2(V_r)}V_rV_r^\dagger + \rho^{\alpha_{12}}V_r\Sigma_{12}^\dagger\Sigma_{12}V_r^\dagger & \\ & I_{(M_1-r)^+}\end{pmatrix}^{-1}U^\dagger V\Lambda_{13}^{(1:r_{123})\dagger}\right| \\
&\quad + r_{123}\log 2\pi e \\
&= \log\Bigg|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}^\dagger\begin{pmatrix} V_r^{\dagger-1}\left(\frac{1}{\lambda_{\max}^2(V_r)}I_r + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12}\right)^{-1}V_r^{-1} & \\ & I_{(M_1-r)^+}\end{pmatrix} \\
&\quad \cdot\begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r}\end{pmatrix}\Lambda_{13}^{(1:r_{123})\dagger}\Bigg| + r_{123}\log(2\pi e) \\
&= \log\left|I_{r_{123}} + \rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}\left(\frac{1}{\lambda_{\max}^2(V_r)}I_r + \rho^{\alpha_{12}}\Sigma_{12}^\dagger\Sigma_{12}\right)^{-1}\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log(2\pi e)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \log\left|I_{r_{123}} + \max\left\{\lambda^2_{\max}(V_r), 1\right\}\rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}\left(I_r + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12}\right)^{-1}\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log(2\pi e)\\
&= \log\left|I_{r_{123}} + \max\left\{\lambda^2_{\max}(V_r), 1\right\}\rho^{\alpha_{13}}\Lambda_{13}^{(1:r_{123})}\begin{pmatrix} (1+\rho^{\alpha_{12}})^{-1}I_{r-r_{13}} & & \\ & \left(I_{r_{123}} + \rho^{\alpha_{12}}C^{\dagger}C\right)^{-1} & \\ & & I_{r-r_{12}}\end{pmatrix}\right.\\
&\qquad \left.\cdot\Lambda_{13}^{(1:r_{123})\dagger}\right| + r_{123}\log(2\pi e)\\
&= \log\left|I_{r_{123}} + \max\left\{\lambda^2_{\max}(V_r), 1\right\}\rho^{\alpha_{13}}S\left(I_{r_{123}} + \rho^{\alpha_{12}}C^{\dagger}C\right)^{-1}S^{\dagger}\right| + r_{123}\log(2\pi e)\\
&\leq \log\left|I_{r_{123}} + \max\left\{\lambda^2_{\max}(V_r), 1\right\}\frac{\rho^{\alpha_{13}}\sigma^2_{\max}(\Lambda_{13})}{1+\rho^{\alpha_{12}}\sigma^2_{\min}(\Lambda_{12})}I_{r_{123}}\right| + r_{123}\log(2\pi e)\\
&\leq \log\left|\max\left\{\lambda^2_{\max}(V_r), 1\right\}\left(1 + \frac{\rho^{\alpha_{13}}\sigma^2_{\max}(\Lambda_{13})}{1+\rho^{\alpha_{12}}\sigma^2_{\min}(\Lambda_{12})}\right)I_{r_{123}}\right| + r_{123}\log(2\pi e)\\
&\leq \log\left|\max\left\{\lambda^2_{\max}(V_r), 1\right\}\left(1 + \frac{\sigma^2_{\max}(\Lambda_{13})}{\sigma^2_{\min}(\Lambda_{12})}\right)I_{r_{123}}\right| + r_{123}\log(2\pi e)\\
&= r_{123}\log\left|\max\left\{\lambda^2_{\max}(V_r), 1\right\}\right| + r_{123}\log\left(1 + \frac{\sigma^2_{\max}(\Lambda_{13})}{\sigma^2_{\min}(\Lambda_{12})}\right) + r_{123}\log(2\pi e)\\
&= \eta + r_{123}\log(2\pi e)
\end{aligned}
\tag{A.30}
$$

It remains to compute the negative terms. After relaxing Q, the negative entropy terms become,

$$
\begin{aligned}
h(Y_1|X_1) &= N_1 \log 2\pi e\\
h(S_{12}|X_1) &= N_2 \log 2\pi e\\
h(S_{123}|X_1) &= r_{123} \log 2\pi e\\
h(S_{123}, S_{12}|X_1) &= (N_2 + r_{123}) \log 2\pi e\\
h(S_{123}, S_{13}|X_1) &= N_3 \log 2\pi e\\
h(S_{123}, S_{12}, S_{13}|X_1) &= (N_2 + N_3) \log 2\pi e.
\end{aligned}
$$

We verify $h(S_{123}|X_1)$ as an example,

$$
\begin{aligned}
&h(S_{123}|X_1)\\
&= h\left(U_{13}\begin{pmatrix}(U_{13}^{-1}Z_3)^{(1:r_{123})}\\ \mathbf{0}_{(N_3-r_{123})\times 1}\end{pmatrix}\right)
\end{aligned}
$$

$$= h\begin{pmatrix} U_{13}^{-1(1:r_{123})} Z_3 \\ \\ \mathbf{0}_{(N_3 - r_{123})\times 1} \end{pmatrix} + \log|U_{13}|$$

$$= h(U_{13}^{-1(1:r_{123})} Z_3)$$

$$= \log|I_{r_{123}}| + r_{123}\log 2\pi e$$

$$= r_{123}\log 2\pi e.$$

The other negative entropy terms can be verified similarly. Finally, the outer bound can be determined by replacing the positive entropy terms in $\mathcal{R}_{\mathrm{o}}^{'}(P)$ with their upper bound of the entropy terms and the negative entropy terms with their computed values obtained above, which completes the proof.

A.4 Proof of Theorem 3.4

We first demonstrate the relationship between the matrices Q_{1p}, $Q_{12}+Q_{1p}$, $Q_{13}+Q_{1p}$ and $Q_{12}+Q_{13}+Q_{1p}$ by (3.49)-(3.52) with matrices K_{1p}, $K_{12,1p}$, $K_{13,1p}$ and $K_{12,13,1p}$ by (3.76)-(3.79), which is stated in Lemma A.2.

Lemma A.2. *The identities given by* (A.36)-(A.40) *hold. Furthermore, the covariance matrices* Q_{1p}, $Q_{12}+Q_{1p}$, $Q_{13}+Q_{1p}$ *and* $Q_{12}+Q_{13}+Q_{1p}$ *can be lower bounded as follows.*

$$Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}}K_{1p} \tag{A.31}$$

$$Q_{12}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}}K_{12,1p} \tag{A.32}$$

$$Q_{13}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}}K_{13,1p} \tag{A.33}$$

$$Q_{12}+Q_{13}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}}K_{12,13,1p} \tag{A.34}$$

$$Q_{1} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}}I_{M_1} \tag{A.35}$$

$$Q_{1p} = \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix} V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger + \rho^{\alpha_{12}}H_{12}^\dagger H_{12} + \rho^{\alpha_{13}}H_{13}^\dagger H_{13}\right)^{-1} \tag{A.36}$$

$$Q_{12}+Q_{1p} = \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix} V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger + \rho^{\alpha_{13}}H_{13}^\dagger H_{13}\right)^{-1} \tag{A.37}$$

$$Q_{13} + Q_{1p} = \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)} \end{pmatrix} U^\dagger + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} G_{13}^\dagger G_{13} \right)^{-1} \tag{A.38}$$

$$Q_{12} + Q_{13} + Q_{1p} = \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)} \end{pmatrix} U^\dagger + \rho^{\alpha_{13}} G_{13}^\dagger G_{13} \right)^{-1} \tag{A.39}$$

$$Q_1 = \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)} \end{pmatrix} U^\dagger \right)^{-1} \tag{A.40}$$

Proof. We show the truth of (A.36) in the sequence leading to (A.41), and the truth of (A.31) from (A.41) to (A.42). The other identities and inequalities can be shown in a similar way, which completes the proof.

$$\begin{aligned}
Q_{1p} &= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} V_p \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} & \\ & \mathbf{0}_{(M_1-r)^+ \times (M_1-r)^+} \end{pmatrix} \right)^{-1} V_P^\dagger \\
&= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} U^{\dagger -1} \begin{pmatrix} V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix}^{-1} \\
&\quad \cdot \left(I_{M_1} + \begin{pmatrix} \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} & \\ & \mathbf{0}_{(M_1-r)^+ \times (M_1-r)^+} \end{pmatrix} \right)^{-1} \begin{pmatrix} V_r & \\ & I_{(M_1-r)^+} \end{pmatrix}^{-1} U^{-1} \\
&= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix} U^\dagger + U \begin{pmatrix} V_r & \\ & I_{(M_1-r)^+} \end{pmatrix} \right. \\
&\quad \left. \begin{pmatrix} \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} & \\ & \mathbf{0}_{(M_1-r)^+ \times (M_1-r)^+} \end{pmatrix} \begin{pmatrix} V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix} U^\dagger \right)^{-1} \\
&= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix} U^\dagger \right. \\
&\quad \left. + U \begin{pmatrix} \rho^{\alpha_{12}} V_r \Sigma_{12}^\dagger \Sigma_{12} V_r^\dagger + \rho^{\alpha_{13}} V_r \Sigma_{13}^\dagger \Sigma_{13} V_r^\dagger & \\ & \mathbf{0}_{(M_1-r)^+ \times (M_1-r)^+} \end{pmatrix} U^\dagger \right)^{-1} \\
&= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix} U^\dagger + \rho^{\alpha_{12}} U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \Sigma_{12}^\dagger \Sigma_{12} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix}^\dagger U^\dagger \right.
\end{aligned}$$

$$+\rho^{\alpha_{13}} U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix} \Sigma_{13}^\dagger \Sigma_{13} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+\times r} \end{pmatrix}^\dagger U^\dagger \Bigg)^{-1}$$

$$= \frac{1}{\mathrm{Tr}(V_p V_p^\dagger)} \left(U \begin{pmatrix} V_r V_r^\dagger & \\ & I_{(M_1-r)^+} \end{pmatrix} U^\dagger + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1} \tag{A.41}$$

$$\succeq \zeta_{\max}^{-1} \left(U \begin{pmatrix} \lambda_{\max}^2(V_r) I_r & \mathbf{0}_{r\times(M_1-r)^+} \\ \mathbf{0}_{(M_1-r)^+\times r} & I_{(M_1-r)^+} \end{pmatrix} U^\dagger + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1}$$

$$\succeq \zeta_{\max}^{-1} \left(\max\left\{\lambda_{\max}^2(V_r), 1\right\} I_{M_1} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1}$$

$$\succeq \zeta_{\max}^{-1} \min\left\{\lambda_{\max}^{-2}(V_r), 1\right\} \left(I_{M_1} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1}$$

$$= \zeta_{\max}^{-1} \min\left\{\lambda_{\max}^{-2}(V_r), 1\right\} K_{1p} \tag{A.42}$$

□

As stated in the proof outline, there is a one-to-one correspondence between the positive entropy terms in both inner and outer bounds. Let $K_1 = I_{M_1}$, then the paired positive entropy terms will be identical if we replace for the sum of "Q" matrices with a corresponding K matrix in the entropy terms associated with R_1. By the lower bounds (A.31)-(A.35), we show the gap resulting from replacing the matrices K_{1p}, $K_{12,1p}$ and $K_{13,1p}$, $K_{12,13,1p}$ and K_1 with Q_{1p}, $Q_{12}+Q_{1p}$, $Q_{13}+Q_{1p}$, $Q_{12}+Q_{13}+Q_{1p}$ and Q_1 in the related positive entropy terms will not exceed δ_1. For example,

$$\begin{aligned} &\log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger\right| \\ \geq &\log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^\dagger\right| \\ &- \min\{M_1, N_1\} \\ &\cdot \left(\log\left(\zeta_{\max} \max\left\{\lambda_{\max}^2(V_r), 1\right\}\right)\right)^+ \\ = &\log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^\dagger\right| - \delta_1 \end{aligned}$$

There are three pair of positive entropy terms in the inner and outer bound associated with R_2. We bound them one by one. The gap between $\log\left|I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13}+Q_{1p}) H_{12}^\dagger + \frac{\rho^{\alpha_{22}}}{M_2} H_{22} H_{22}^\dagger\right|$ and

$\log\left|I_{N_2}+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right|$ are bounded in the following.

$$\begin{aligned}
&\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}(Q_{13}+Q_{1p})H_{12}^\dagger+\frac{\rho^{\alpha_{22}}}{M_2}H_{22}H_{22}^\dagger\right| \\
\geq\ &\log\left|I_{N_2}+\frac{\rho^{\alpha_{22}}}{M_2}H_{22}H_{22}^\dagger\right| \\
\geq\ &\log\left|I_{N_2}+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right|-\min\{M_2,N_2\}\log M_2
\end{aligned} \tag{A.43}$$

The gap between $\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}Q_1H_{12}^\dagger+\frac{\rho^{\alpha_{22}}}{M_2}H_{22}H_{22}^\dagger\right|$ and $\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}H_{12}^\dagger+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right|$ are thus bounded in the following.

$$\begin{aligned}
&\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}Q_1H_{12}^\dagger+\frac{\rho^{\alpha_{22}}}{M_2}H_{22}H_{22}^\dagger\right| \\
\geq\ &\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}H_{12}^\dagger+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right| \\
&-\min\left\{M_1+M_2,N_2\right\} \\
&\cdot\log\max\left\{\zeta_{\max}\max\left\{\lambda_{\max}^2(V_r),1\right\},M_2\right\}
\end{aligned} \tag{A.44}$$

It is obvious that the gap between $\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}(Q_{12}+Q_{13}+Q_{1p})H_{12}^\dagger+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right|$ and $\log\left|I_{N_2}+\rho^{\alpha_{12}}H_{12}K_{12,13,1p}H_{12}^\dagger+\rho^{\alpha_{22}}H_{22}H_{22}^\dagger\right|$ is also bounded by (A.44). The maximum should be the gap contributes to the individual gap n_2 on R_2, which is δ_2 given by (3.90). Similarly, the gap between the positive terms associated with R_3 in the inner and outer bounds is upper bounded by δ_3.

So far, we quantified the gap between the entropy terms in the inner and outer bounds. Lastly, we see there are negative terms β_2 and β_3 in the inner bound but not in the outer bound and the positive term η in the outer bound but not in the inner bound. We let β_2 and β_3 be absorbed by n_2 and n_3 respectively, and η be absorbed by n_1. The proof is completed.

A.5 Proof of Theorem 3.5

The first three bounds (3.94)-(3.96) on individual rate can be obtained by directly applying Fact 3.2 to the first three inequalities (3.81)-(3.83) in the outer bound. The inequality (3.97) is the sum of the limits of the two logarithm terms on the right hand side of (3.84). Note we can not directly apply Fact 3.2 to the first term $\log\left|I_{N_1}+\rho^{\alpha_{11}}H_{11}K_{13,1p}H_{11}^\dagger\right|$. However, this term can be asymptotically expressed in an alternative format as shown in (A.45) through a sequence of matrix operations that leads to it, and we define the matrix $H_{11}^{'} \triangleq H_{11}U \begin{pmatrix} V_r^{\dagger-1} & \\ & I_{M_1-r} \end{pmatrix}$ in step (a). Note since U is a unitary matrix and $\begin{pmatrix} V_r^{\dagger-1} & \\ & I_{M_1-r} \end{pmatrix}$ has full rank, we have $H_{11}^{'}$ has full rank w.p.1 as H_{11}, hence we can apply Fact 3.2 to (A.45), which leads to GDoF $g\left(N_1, ((\alpha_{11}-\alpha_{12})^+, r_{12}), (\alpha_{11}, M_1-r_{12})\right)$. The second logarithm term in the right hand side of (3.84) leads to GDoF $g\left(N_2, (\alpha_{12}, M_1), (\alpha_{22}, M_2)\right)$ with direct application of Fact (3.2). The inequality (3.97) is then proved. Using the same technique, other inequalities in Theorem 3.5 can be similarly verified, which completes the proof.

$$
\begin{aligned}
&\log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^{\dagger}\right| \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (I_{M_1} + \rho^{\alpha_{13}} G_{13}^{\dagger} G_{13} + \rho^{\alpha_{12}} H_{12}^{\dagger} H_{12})^{-1} H_{11}^{\dagger}\right| \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} (I_{M_1} + \rho^{\alpha_{13}} V \Lambda_{13}^{\dagger}\Lambda_{13} V^{\dagger} + \rho^{\alpha_{12}} V \Sigma_{12}^{\dagger}\Sigma_{12} V^{\dagger})^{-1} H_{11}^{\dagger}\right| \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} U \begin{pmatrix} I_r + V_r(\rho^{\alpha_{13}}\Lambda_{13}^{\dagger}\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12})V_r^{\dagger} & \\ & I_{(M_1-r)^+} \end{pmatrix}^{-1} U^{\dagger} H_{11}^{\dagger}\right| \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} U \begin{pmatrix} V_r^{\dagger-1} & \\ & I_{M_1-r} \end{pmatrix} \begin{pmatrix} I_r + \rho^{\alpha_{13}}\Lambda_{13}^{\dagger}\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12} & \\ & I_{(M_1-r)^+} \end{pmatrix}^{-1}\right. \\
&\qquad \left. \cdot \begin{pmatrix} V_r^{-1} & \\ & I_{M_1-r} \end{pmatrix} U^{\dagger} H_{11}^{\dagger}\right| \\
&\overset{(a)}{=} \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}^{'} \begin{pmatrix} I_r + \rho^{\alpha_{13}}\Lambda_{13}^{\dagger}\Lambda_{13} + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12} & \\ & I_{(M_1-r)^+} \end{pmatrix}^{-1} H_{11}^{'\dagger}\right| + O(1) \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}^{'} \begin{pmatrix} (1+\rho^{\alpha_{12}})^{-1} I_{r-r_{13}} & & & \\ & (1+\rho^{\alpha_{13}}+\rho^{\alpha_{12}})^{-1} I_{r_{123}} & & \\ & & I_{r-r_{12}} & \\ & & & I_{(M_1-r)^+} \end{pmatrix} H_{11}^{'\dagger}\right| + \mathcal{O}(1) \\
&= \log\left|I_{N_1} + \frac{\rho^{\alpha_{11}}}{1+\rho^{\alpha_{12}}} H_{11}^{'[1:r-r_{13}]} H_{11}^{'[1:r-r_{13}]\dagger} + \frac{\rho^{\alpha_{11}}}{1+\rho^{\alpha_{13}}+\rho^{\alpha_{12}}} H_{11}^{'[r-r_{13}+1:r_{12}]} H_{11}^{'[r-r_{13}+1:r_{12}]\dagger}\right. \\
&\qquad \left. + \rho^{\alpha_{11}} H_{11}^{'[r_{12}+1:r]} H_{11}^{'[r_{12}+1:r]\dagger} + \rho^{\alpha_{11}} H_{11}^{'[r+1:M_1]} H_{11}^{'[r+1:M_1]\dagger}\right| + \mathcal{O}(1)
\end{aligned}
$$

$$= \log \left| I_{N_1} + \rho^{\alpha_{11}-\alpha_{12}} H_{11}^{'[1:r-r_{13}]} H_{11}^{'[1:r-r_{13}]\dagger} + \rho^{\alpha_{11}-\alpha_{12}} H_{11}^{'[r-r_{13}+1:r_{12}]} H_{11}^{'[r-r_{13}+1:r_{12}]\dagger} \right.$$

$$\left. + \rho^{\alpha_{11}} H_{11}^{'[r_{12}+1:r]} H_{11}^{'[r_{12}+1:r]\dagger} + \rho^{\alpha_{11}} H_{11}^{'[r+1:M_1]} H_{11}^{'[r+1:M_1]\dagger} \right| + o(1) \tag{A.45}$$

$$= g\left(N_1, \left((\alpha_{11}-\alpha_{12})^+, r_{12}\right), (\alpha_{11}, M_1 - r_{12})\right) + \mathcal{O}(1) \tag{A.46}$$

Appendix B

Proofs for the Results on MIMO IC-ZIC

B.1 Proof of Theorem 4.1

We prove the achievability through a random coding argument. As stated in the proof outline, we employ three level superposition coding at Tx1, two level superposition coding (the CMG type coding) Tx2 and single user random coding at Tx3. More specifically, the achievable scheme can be described in the following steps.

(1) Generate time sharing sequence q^n according to $p(q^n) = \prod_{t=1}^n p(q_t)$.

(2) Tx1 generates $2^{nR_{123}}$ sequences w_{123}^n according to $p(w_{123}^n|q^n) = \prod_{t=1}^n p(w_{123,t}|q_t)$ and indexes them by $k_{123} \in \{1, \cdots, 2^{nR_{123}}\}$. For each $w_{123}^n(k_{123})$, it generates $2^{nR_{12}}$ sequences w_{12}^n according to $p(w_{12}^n|w_{123}^n(k_{123}), q^n) = \prod_{t=1}^n p(w_{12,t}|w_{123,t}(k_{123}), q_t)$ and indexes them by $(k_{123}, k_{12}) \in \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{12}}\}$ as well as $2^{nR_{13}}$ sequences w_{13}^n according to

$$p(w_{13}^n|w_{123}^n(k_{123}), q^n) = \prod_{t=1}^n p(w_{13,t}|w_{123,t}(k_{123}), q_t)$$

indexed by $(k_{123}, k_{13}) \in \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{13}}\}$. Finally Tx1 generates $2^{n(R_1 - R_{123} - R_{12} - R_{13})}$ sequences x_1^n according to

$$p(x_1^n|w_{123}^n(k_{123}), w_{12}^n(k_{12}), w_{13}^n(k_{13}), q^n) = \prod_{t=1}^n p(x_{1t}|w_{123,t}(k_{123}), w_{12.t}(k_{12}), w_{13,t}(k_{13}), q_t)$$

and indexes them as

$$(k_{123}, k_{12}, k_{13}, k_{1p}) \in \{1, \cdots 2^{nR_{12}}\} \times \{1, \cdots, 2^{nR_{123}}\} \times \{1, \cdots 2^{nR_{13}}\} \times \{1, \cdots, 2^{n(R_1 - R_{123} - R_{12} - R_{13})}\}.$$

(3) Tx2 generates $2^{nR_{21}}$ sequences w_{21}^n according to $p(w_{21}^n|q^n) = \prod_{t=1}^n p(w_{21,t}|q_t)$ and indexes them by $k_{21} \in \{1, \cdots, 2^{nR_{21}}\}$. For each $w_{21}^n(k_{21})$, it generates $2^{n(R_2 - R_{21})}$ sequences w_{21}^n according to $p(w_{2p}^n|w_{21}^n(k_{21}), q^n) = \prod_{t=1}^n p(w_{2p,t}|w_{21,t}(k_{21}), q_t)$ and indexes them by $(k_{21}, k_{2p}) \in \{1, \cdots, 2^{nR_{21}}\} \times \{1, \cdots 2^{nR_{2p}}\}$.

(4) Tx3 independently generates 2^{nR_3} sequences x_3^n according to $p(x_3^n|q^n) = \prod_{t=1}^n p(x_{3t}|q_t)$ and indexes them by $k_3 \in \{1, \cdots, 2^{nR_3}\}$.

(5) Once the codebooks are generated, they are fixed for the duration of communication and revealed to receivers Rx1-Rx3.

(6) A 4-tuple message $m_1 = (m_{123}, m_{12}, m_{13}, m_{1p}) = (k_{123}, k_{12}, k_{13}, k_{1p})$ at Tx1 is encoded to $x_1^n(k_{123}, k_{12}, k_{13}, k_{1p})$ at Tx1 and sent over the channel. A message $m_2 = (k_{21}, k_{2p})$ at Tx2 is encoded to $x_2^n(k_{21}, k_{2p})$ and sent over the channel. A message $m_3 = k_3$ is encoded to $x_3^n(m_3)$ and sent over the channel.

(7) Upon receiving y_1^n, Rx1 declares its decoded messages $(\hat{m}_{123}, \hat{m}_{12}, \hat{m}_{13}, \hat{m}_{1p}, \hat{m}_{21})$ as the unique index-tuple $(\hat{k}_{123}, \hat{k}_{12}, \hat{k}_{13}, \hat{k}_{1p}, \hat{k}_{21})$ for which q^n, $w_{123}^n(\hat{k}_{123})$, $w_{12}^n(\hat{k}_{123}, \hat{k}_{12})$, $w_{13}^n(\hat{k}_{123}, \hat{k_{13}})$, $x_1^n(\hat{k}_{123}, \hat{k}_{12}, \hat{k}_{13}, \hat{k}_{1p})$, $w_{21}^n(\hat{k}_{21})$ and y_i^n are jointly typical, for some $\hat{k}_{21}$. If such an index-tuple cannot be found, Rx1 declares an error.

(8) Upon receiving y_2^n, Rx2 declares its decoded messages $(\hat{m}_{123}, \hat{m}_{12}, \hat{m}_{21}, \hat{m}_{2p})$ as the unique index-tuple $(\hat{k}_{123}, \hat{k}_{12}, \hat{k}_{21}, \hat{k}_{2p})$ for which q^n, $w_{123}^n(\hat{k}_{123})$, $w_{12}^n(\hat{k}_{123}, \hat{k}_{12})$, $w_{21}^n(\hat{k}_{21})$, $x_2^n(\hat{k}_{21}, \hat{k}_{2p})$ and y_2^n are jointly typical, for some $\hat{k}_{123}$ and $\hat{k}_{12}$. If such an index-tuple cannot be found, Rx2 declares an error.

(9) Upon receiving y_3^n, Rx3 declares its decoded messages $(\hat{m}_{123}, \hat{m}_{13}, \hat{m}_3)$ as the unique index-tuple $(\hat{k}_{123}, \hat{k}_{13}, \hat{k}_3)$ for which q^n, $w_{123}^n(\hat{k}_{123})$, $w_{13}^n(\hat{k}_{123}, \hat{k}_{13})$, $x_3^n(\hat{k}_3)$ and y_3^n are jointly typical, for some $\hat{k}_{123}$ and $\hat{k}_{13}$. If such an index-tuple cannot be found, Rx3 declares an error.

Suppose $m_1 = (1, 1, 1, 1)$, $m_2 = (1, 1)$, and $m_3 = 1$ are sent. The reliability condition can be obtained from the typical decoding argument [13, Chapter 7],

$$R_1 - R_{1c} \leq I(X_1; Y_1|W_{1c}, W_{21}, Q)$$

$$R_1 - R_{123} - R_{12} \leq I(X_1; Y_1 | W_{123}, W_{12}, W_{21}, Q)$$

$$R_1 - R_{123} - R_{13} \leq I(X_1; Y_1 | W_{123}, W_{13}, W_{21}, Q)$$

$$R_1 - R_{123} \leq I(X_1; Y_1 | W_{123}, W_{21}, Q)$$

$$R_1 \leq I(X_1; Y_1 | W_{21}, Q)$$

$$R_1 - R_{1c} + R_{21} \leq I(X_1, W_{21}; Y_1 | W_{1c}, Q)$$

$$R_1 - R_{123} - R_{12} + R_{21} \leq I(X_1, W_{21}; Y_1 | W_{123}, W_{12}, Q)$$

$$R_1 - R_{123} - R_{13} + R_{21} \leq I(X_1, W_{21}; Y_1 | W_{123}, W_{13}, Q)$$

$$R_1 - R_{123} + R_{21} \leq I(X_1, W_{21}; Y_1 | W_{123}, Q)$$

$$R_1 + R_{21} \leq I(X_1, W_{21}; Y_1 | Q)$$

$$R_2 - R_{21} \leq I(X_2; Y_2 | W_{123}, W_{12}, W_{21}, Q)$$

$$R_2 \leq I(X_2; Y_2 | W_{123}, W_{12}, Q)$$

$$R_2 - R_{21} + R_{12} \leq I(X_2, W_{12}; Y_2 | W_{123}, W_{21}, Q)$$

$$R_2 + R_{12} \leq I(X_2, W_{12}; Y_2 | W_{123}, Q)$$

$$R_2 - R_{21} + R_{123} + R_{12} \leq I(X_2, W_{123}, W_{12}; Y_2 | W_{21}, Q)$$

$$R_2 + R_{123} + R_{12} \leq I(X_2, W_{123}, W_{12}; Y_2 | Q)$$

$$R_3 \leq I(X_3; Y_3 | W_{123}, W_{13}, Q)$$

$$R_3 + R_{13} \leq I(X_3, W_{13}; Y_3 | W_{123}, Q)$$

$$R_3 + R_{123} + R_{13} \leq I(X_3, W_{123}, W_{13}; Y_3 | Q)$$

$$R_{123}, R_{12}, R_{13} \geq 0$$

$$R_{123} + R_{12} + R_{13} \leq R_1$$

$$R_{21} \geq 0$$

$$R_{21} \leq R_2$$

$$R_2, R_3 \geq 0.$$

Performing Fourier-Motzkin elimination to eliminate R_{123}, R_{12} and R_{13} in the reliability condition, the inner bound can be obtained, which completes the proof.

B.2 Proof of Theorem 4.2

We evaluate the DM IC-ZIC inner bound when specialized to the MIMO channel setting and for the coding scheme introduced in the proof outline. In what follows, we prove the set function $F(\mathtt{M}_{1p})$ and $F(\mathtt{M}_{2p})$ as two examples.

$$\begin{aligned}
\mathsf{F}_1(\mathtt{M}_{1p}) &= I(X_1; Y_1 | W_{1c}, W_{21}, Q) \\
&= h(Y_1 | W_{1c}, W_{21}) - h(Y_1 | X_1, W_{21}) \\
&= \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^{\dagger} + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^{\dagger}) \\
&\quad - \log(I_{N_1} + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^{\dagger}) \\
&\overset{(a)}{\geq} \log(I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^{\dagger} + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^{\dagger}) \\
&\quad - \min\{M_2, N_1\} \log\left(\frac{1 + M_2}{M_2}\right) \\
&= F_1(m_{1p}) - \beta_1
\end{aligned}$$

Step (a) is true due to [27, equation (66)].

$$\begin{aligned}
&\mathsf{F}_2(m_{2p}) \\
&= I(X_2; Y_2 | W_{123}, W_{12}, W_{21}, Q) \\
&= h(Y_2 | W_{123}, W_{12}, W_{21}) - h(Y_2 | W_{123}, W_{12}, X_2) \\
&= \log\left(I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} Q_{2p} H_{22}\right) \\
&\quad - \log\left(I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{21}^{\dagger}\right) \\
&\overset{(a)}{\geq} \log\left(I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{12}^{\dagger} + \rho^{\alpha_{22}} H_{22} Q_{2p} H_{22}\right)
\end{aligned}$$

$$
\begin{aligned}
&\quad - \log \left|\max\left\{\zeta_{\min}^{-1}, 1\right\}\right| \\
&\quad - (r - r_{13}) - r_{123} \log\left(1 + \frac{\sigma_{\max}^2(\Lambda_{12})}{\sigma_{\min}^2(\Lambda_{12})}\right) \\
&= F_1(\mathtt{M}_{2p}) - \beta_2
\end{aligned}
$$

Step (a) is true according to the lower bound of $\log\left(I_{N_2} + \rho^{\alpha_{12}} H_{12}(Q_{13} + Q_{1p}) H_{21}^{\dagger}\right)$ by (A.1) in Section A.2 when $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$. With a slight change of the sequence that leads to (A.1), the same lower bound holds when $\mathsf{INR}_{12} < \mathsf{INR}_{13}$. The rest of the set functions for $\mathcal{R}_{\text{in}}$ can be verified in a similar fashion. The proof is completed.

B.3 Proof of Theorem 4.3

As stated in the proof outline, the outer bound for the MIMO IC-ZIC is derived in three steps. In the first step, we define the genie information and derive a variety of individual rate upper bounds on R_1,R_2 and R_3. In step two, combining these individual rate restrictions, we obtain an intermediate outer bound with these genie informations and channel side informations. This intermediate outer bound is a union region outer bound over all admissible input distributions, Lastly, we optimize the input distribution to be vector Gaussian and characterize a single region outer bound with this optimized distribution.

We construct the four random vectors T_{123}, T_{12}, T_{13} and T_{21} given by (B.1)-(B.3) as the genie informations to help receivers decode their message, where $Z_i' \sim \mathcal{CN}(\mathbf{0}, N_i)$ $Z_i' \perp Z_i$ for $i \in \{1, 2, 3\}$. T_{123}, T_{12}, T_{13} and T_{21} have identical distributions as the channel side informations S_{123}, S_{12}, S_{13} and S_{21} (c.f. (4.23)-(4.25) and (4.22)), respectively, but each pair of corresponding "T" and "S" random variables are independent conditioned on X_1.

$$
T_{123} = \begin{cases}
h_{13} G_{13} X_1 + U_{13} \begin{pmatrix} (U_{13}^{-1} Z_3')^{(1:r_{123})} \\ \mathbf{0}_{(N_3 - r_{123})\times 1} \end{pmatrix} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\
h_{12} G_{12} X_1 + U_{12} \begin{pmatrix} \mathbf{0}_{(r - r_{13})\times 1} \\ U_{12}^{-1(r - r_{13}+1:r_{12})} Z_2' \\ \mathbf{0}_{(N_2 - r_{12})\times 1} \end{pmatrix} & \mathsf{INR}_{12} < \mathsf{INR}_{13}
\end{cases} \tag{B.1}
$$

$$T_{12} = \begin{cases} h_{12}H_{12}X_1 + Z_2' & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h_{12}J_{12}X_1 + U_{12}\begin{pmatrix} U_{12}^{-1(1:r-r_{13})}Z_2' \\ \mathbf{0}_{r_{123}\times 1} \\ U_{12}^{-1(r_{12}+1:N_2)}Z_2' \end{pmatrix} & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{B.2}$$

$$T_{13} = \begin{cases} h_{13}J_{13}X_1 + U_{13}\begin{pmatrix} \mathbf{0}_{r_{123}\times 1} \\ (U_{13}^{-1}Z_3')^{(r_{123}+1:N_3)} \end{pmatrix} & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h_{13}H_{13}X_1 + Z_3' & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases} \tag{B.3}$$

$$T_{21} = h_{21}H_{21}X_2 + Z_1'. \tag{B.4}$$

The key fact in deriving individual rate in the following three subsections is that providing genie information to the receiver makes the receiver more interference resilient, and therefore should not decrease the capacity of the channel.

B.3.1 Upper Bounds on R_1

If we do not feed any genie information, the individual rate R_1 is simply upper bounded by a P2P channel capacity

$$\begin{aligned} nR_1 &\leq I(X_1^n; Y_1^n) + n\epsilon \\ &= h(Y_1^n) - h(Y_1^n|X_1^n) + n\epsilon \\ &= h(Y_1^n) - h(S_{21}^n) + n\epsilon \\ &\leq nh(Y_1|Q) - h(S_{21}^n) + n\epsilon \\ &= n\left[h(Y_1|Q) - h(S_{21}|X_2, Q)\right] + nh(S_{21}|X_2, Q) \\ &\quad - h(S_{21}^n) + n\epsilon \\ &\triangleq n\bar{F}_1'(\mathtt{M}_1, \mathtt{M}_{21}) + nh(S_{21}|X_2, Q) - h(S_{21}^n) + n\epsilon. \end{aligned} \tag{B.5}$$

Providing genie information T_{123} to Rx1, we get another upper bound on the rate R_1

$$
\begin{aligned}
nR_1 &\overset{(a)}{\leq} I(X_1^n; Y_1^n, T_{123}^n) + n\epsilon \\
&\overset{(b)}{=} I(X_1^n, T_{123}^n) + I(X_1^n, Y_1^n | T_{123}^n) \\
&= h(T_{123}^n) - h(T_{123}^n | X_1^n) + h(Y_1^n | T_{123}^n) - h(Y_1^n | X_1^n) \\
&\quad + n\epsilon \\
&\overset{(c)}{=} h(T_{123}^n) - h(T_{123}^n | X_1^n) + h(Y_1^n | T_{123}^n) - h(S_{21}^n) + n\epsilon \\
&\overset{(d)}{\leq} nh(Y_1 | T_{123}, Q) - nh(S_{123} | X_1, Q) + h(S_{123}^n) \\
&\quad - h(S_{21}^n) + n\epsilon \\
&= n\left[h(Y_1 | T_{123}, Q) - h(S_{21} | X_2, Q)\right] + nh(S_{21} | X_2, Q) \\
&\quad - nh(S_{123} | X_1, Q) + h(S_{123}^n) - h(S_{21}^n) + n\epsilon \\
&\triangleq n\bar{F}_1'(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + nh(S_{21} | X_2, Q) \\
&\quad - nh(S_{123} | X_1, Q) + h(S_{123}^n) - h(S_{21}^n) + n\epsilon. \qquad \text{(B.6)}
\end{aligned}
$$

The inequalities or equations (a)-(c) hold true because: (a) providing genie information T_{123}^n to Rx1 will not decrease the channel capacity; (b) channel rule of mutual information; (c) $h(Y_1^n | X_1^n, T_{123}^n) = h(Y_1^n | X_1^n) = h(S_{21}^n)$ according to the definition of the channel side informations. (d) $h(S_{123}^n) = h(T_{123}^n)$ as S_{123} and T_{123} are identically distributed.

Similarly, if we feed genie information (T_{123}^n, T_{12}^n), (T_{123}^n, T_{13}^n) and $(T_{123}^n, T_{12}^n, T_{13}^n)$, we obtain the following three outer bounds on R_1,

$$
\begin{aligned}
nR_1 &\leq n\left[h(Y_1 | T_{123}, T_{12}, Q) - h(S_{21} | X_2, Q)\right] \\
&\quad + nh(S_{21} | X_2, Q) - nh(S_{123}, S_{12} | X_1, Q) \qquad \text{(B.7)} \\
&\quad + h(S_{123}^n, S_{12}^n) - h(S_{21}^n) + n\epsilon \\
&\triangleq n\bar{F}_1'(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + nh(S_{21} | X_2, Q) \\
&\quad - nh(S_{123}, S_{12} | X_1, Q) + h(S_{123}^n, S_{12}^n) \qquad \text{(B.8)} \\
&\quad - h(S_{21}^n) + n\epsilon \qquad \text{(B.9)}
\end{aligned}
$$

$$
\begin{aligned}
nR_1 \leq\ & n\left[h(Y_1|T_{123}, T_{13}, Q) - h(S_{21}|X_2, Q)\right] \\
& + nh(S_{21}|X_2, Q) - nh(S_{123}, S_{13}|X_1, Q) \\
& + h(S_{123}^n, S_{13}^n) - h(S_{21}^n) + n\epsilon && \text{(B.10)} \\
\triangleq\ & n\bar{F}_1'(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + nh(S_{21}|X_2, Q) \\
& - nh(S_{123}, S_{13}|X_1, Q) + h(S_{123}^n, S_{13}^n) && \text{(B.11)} \\
& - h(S_{21}^n) + n\epsilon && \text{(B.12)} \\
nR_1 \leq\ & n\left[h(Y_1|T_{123}, T_{12}, T_{13}, Q) - h(S_{21}|X_2, Q)\right] \\
& + nh(S_{21}|X_2, Q) - nh(S_{123}, S_{12}, S_{13}|X_1, Q) \\
& + h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{21}^n) + n\epsilon && \text{(B.13)} \\
\triangleq\ & n\bar{F}_1'(\mathtt{M}_{1p}, \mathtt{M}_{21}) + nh(S_{21}|X_2, Q) \\
& - nh(S_{123}, S_{12}, S_{13}|X_1, Q) + h(S_{123}^n, S_{12}^n, S_{13}^n) \\
& - h(S_{21}^n) + n\epsilon && \text{(B.14)}
\end{aligned}
$$

We then obtain five other outer bounds on R_1 based on the configuration of genie informations in getting (B.5)-(B.14) but additionally feeding X_2^n to Rx1.

$$
\begin{aligned}
nR_1 \leq\ & nh(Y_1|X_2, Q) - nh(S_{21}|X_2, Q) + n\epsilon \\
\triangleq\ & n\bar{F}_1'(\mathtt{M}_1) + n\epsilon && \text{(B.15)} \\
nR_1 \leq\ & nh(Y_1|T_{123}, X_2, Q) - nh(S_{123}|X_1, Q) \\
& - nh(S_{21}|X_2, Q) + h(S_{123}^n) + n\epsilon \\
\triangleq\ & n\bar{F}_1'(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) - nh(S_{123}|X_1, Q) \\
& + h(S_{123}^n) + n\epsilon && \text{(B.16)} \\
nR_1 \leq\ & n(Y_1|T_{123}, T_{12}, X_2, Q) - nh(S_{123}, S_{12}|X_1, Q) \\
& - nh(S_{21}|X_2, Q) + h(S_{123}^n, S_{12}^n) + n\epsilon \\
\triangleq\ & n\bar{F}_1'(\mathtt{M}_{13}, \mathtt{M}_{1p}) - nh(S_{123}, S_{12}|X_1, Q) \\
& + h(S_{123}^n, S_{12}^n) + n\epsilon && \text{(B.17)}
\end{aligned}
$$

$$
\begin{aligned}
nR_1 &\leq n(Y_1|T_{123}, T_{13}, X_2, Q) - nh(S_{123}, S_{13}|X_1, Q) \\
&\quad - nh(S_{21}|X_2, Q) + h(S_{123}^n, S_{13}^n) + n\epsilon \\
&\triangleq n\bar{F}_1'(\mathtt{M}_{12}, \mathtt{M}_{1p}) - nh(S_{123}, S_{13}|X_1, Q) \\
&\quad + h(S_{123}^n, S_{13}^n) + n\epsilon && \text{(B.18)} \\
nR_1 &\leq nh(Y_1|T_{123}, T_{12}, T_{13}, X_2, Q) \\
&\quad - nh(S_{123}, S_{12}, S_{13}|X_1, Q) - nh(S_{21}|X_2, Q) \\
&\quad + h(S_{123}^n, S_{12}^n, S_{13}^n) + n\epsilon \\
&\triangleq n\bar{F}_1'(\mathtt{M}_{1p}) - nh(S_{123}, S_{12}, S_{13}|X_1, Q) \\
&\quad + h(S_{123}^n, S_{12}^n, S_{13}^n) + n\epsilon && \text{(B.19)}
\end{aligned}
$$

B.3.2 Upper Bounds on R_2

When $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, we obtain the following set of six outer bounds on R_2 by 1) not feeding any genie information, 2) T_{21}^n, 3) T_{123}^n, 4)(T_{123}^n, T_{21}^n), 5) X_1^n and 6) (X_1^n, T_{21}^n) to Rx2, respectively. The proof of these bounds is quite similar to the proof of the upper bounds on R_1. We directly state the result in the following.

$$
\begin{aligned}
nR_2 &\leq n\left[h(Y_2|Q) - h(S_{12}|X_1, Q)\right] \\
&\quad + nh(S_{12}|X_1, Q) - h(S_{12}^n) + n\epsilon && \text{(B.20)} \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + nh(S_{12}|X_1, Q) && \text{(B.21)} \\
&\quad - h(S_{12}^n) + n\epsilon && \text{(B.22)} \\
nR_2 &\leq n\left[h(Y_2|T_{21}, Q) - h(S_{12}|X_1, Q)\right] \\
&\quad + nh(S_{12}|X_1, Q) - nh(S_{21}|X_2, Q) \\
&\quad + h(T_{21}^n) - h(S_{12}^n) + n\epsilon && \text{(B.23)} \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + nh(S_{12}|X_1, Q) \\
&\quad - nh(S_{21}|X_2, Q) + h(T_{21}^n) - h(S_{12}^n) + n\epsilon && \text{(B.24)}
\end{aligned}
$$

$$\begin{aligned}
nR_2 &\leq n\left[h(Y_2|T_{123},Q) - h(S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{12}|X_1,Q) - h(S_{12}^n|S_{123}^n) + n\epsilon \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{12},\mathtt{M}_2) + nh(S_{12}|X_1,Q) && \text{(B.25)} \\
&\quad - h(S_{12}^n|S_{123}^n) + n\epsilon && \text{(B.26)} \\
nR_2 &\leq n\left[h(Y_2|T_{123},T_{21},Q) - h(S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{12}|X_1,Q) - nh(S_{21}|X_2,Q) + h(S_{21}^n) \\
&\quad - h(S_{12}^n|S_{123}^n) + n\epsilon && \text{(B.27)} \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{12},\mathtt{M}_{2p}) + nh(S_{12}|X_1,Q) - nh(S_{21}|X_2,Q) \\
&\quad + h(S_{21}^n) - h(S_{12}^n|S_{123}^n) + n\epsilon && \text{(B.28)} \\
nR_2 &\leq n\left[h(Y_2|X_1,Q) - h(S_{12}|X_1)\right] + n\epsilon \\
&\triangleq n\bar{F}_2'(\mathtt{M}_2) + n\epsilon && \text{(B.29)} \\
nR_2 &\leq nh(Y_2|X_1,T_{21},Q) - nh(S_{12}|X_1) - nh(S_{21}|X_2,Q) \\
&\quad + h(T_{21}^n) + n\epsilon \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{2p}) - nh(S_{21}|X_2,Q) + h(T_{21}^n) + n\epsilon && \text{(B.30)}
\end{aligned}$$

When $\mathsf{INR}_{12} < \mathsf{INR}_{13}$, we have another set of six outer bounds on R_2.

$$\begin{aligned}
nR_2 &\leq n\left[h(Y_2|Q) - h(S_{123},S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{123},S_{12}|X_1,Q) - h(S_{123}^n,S_{12}^n) + n\epsilon \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) + nh(S_{123},S_{12}|X_1,Q) \\
&\quad - h(S_{123}^n,S_{12}^n) + n\epsilon \\
nR_2 &\leq n\left[h(Y_2|T_{21},Q) - h(S_{123},S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{123},S_{12}|X_1,Q) - nh(S_{21}|X_2,Q) \\
&\quad + h(S_{21}^n) - h(S_{123}^n,S_{12}^n) + n\epsilon \\
&\triangleq n\bar{F}_2'(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2p}) + nh(S_{123},S_{12}|X_1,Q) \\
&\quad - nh(S_{21}|X_2,Q) + h(S_{21}^n) - h(S_{123}^n,S_{12}^n) + n\epsilon
\end{aligned}$$

$$\begin{aligned}
nR_2 &\leq n\left[h(Y_2|T_{123},Q) - h(S_{123},S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{123},S_{12}|X_1,Q) - h(S_{123}^n,S_{12}^n|T_{123}^n) + n\epsilon \\
&\triangleq \bar{F}_2^{'}(\mathtt{M}_{12},\mathtt{M}_2) + nh(S_{123},S_{12}|X_1,Q) \\
&\quad - h(S_{123}^n,S_{12}^n|T_{123}^n) + n\epsilon \\
nR_2 &\leq n\left[h(Y_2|T_{123},T_{21},Q) - h(S_{123},S_{12}|X_1,Q)\right] \\
&\quad + nh(S_{123},S_{12}|X_1,Q) - nh(S_{21}|X_2,Q) \\
&\quad + h(T_{21}^n) - h(S_{123}^n,S_{12}^n|T_{123}^n) + n\epsilon \\
&\triangleq n\bar{F}_2^{'}(\mathtt{M}_{12},\mathtt{M}_{2p}) + nh(S_{123},S_{12}|X_1,Q) \\
&\quad - nh(S_{21}|X_2,Q) + h(T_{21}^n) \\
&\quad - h(S_{123}^n,S_{12}^n|T_{123}^n) + n\epsilon \\
nR_2 &\leq nh(Y_2|X_1,Q) - nh(S_{123},S_{12}|X_1,Q) + n\epsilon \\
&\triangleq n\bar{F}_2^{'}(\mathtt{M}_2) + n\epsilon \\
nR_2 &\leq nh(Y_2|X_1,T_{21},Q) - nh(S_{21}|X_2,Q) \\
&\quad - nh(S_{123},S_{12}|X_1,Q) + h(T_{21}^n) + n\epsilon \\
&\triangleq n\bar{F}_2^{'}(\mathtt{M}_{2p}) - nh(S_{21}|X_2,Q) + h(T_{21}^n) + n\epsilon
\end{aligned}$$

With a slight abuse of the notation, we defined each set function $\bar{F}_2^{'}(\cdot)$ twice in two channel conditions. Therefore, $\bar{F}_2^{'}(\cdot)$ should be understood as a function with two mappings depending on the relationship between INR_{12} and INR_{13}. For example,

$$\begin{aligned}
&\bar{F}_2^{'}(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_2) \\
&= \begin{cases} h(Y_2|Q) - h(S_{12}|X_1,Q) & \mathsf{INR}_{12} \geq \mathsf{INR}_{13} \\ h(Y_2|Q) - h(S_{123},S_{12}|X_1,Q) & \mathsf{INR}_{12} < \mathsf{INR}_{13} \end{cases},
\end{aligned}$$

so is $\bar{F}_3^{'}(\cdot)$ in the next subsection.

B.3.3 Upper Bounds on R_3

When $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, the following three upper bounds on R_3 are obtained by 1) feeding no genie information, 2)T_{123}^n and 3) X_1^n to Rx3, respectively.

$$nR_3 \leq n\left[h(Y_3|Q) - h(S_{123}, S_{13}|X_1, Q)\right]$$

$$+ nh(S_{123}, S_{13}|X_1, Q) - h(S_{123}^n, S_{13}^n) + n\epsilon \tag{B.31}$$

$$\triangleq n\bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + nh(S_{123}, S_{13}|X_1, Q) \tag{B.32}$$

$$- h(S_{123}^n, S_{13}^n) + n\epsilon \tag{B.33}$$

$$nR_3 \leq n\left[h(Y_3|T_{123}, Q) - h(S_{123}, S_{13}|X_1, Q)\right]$$

$$+ nh(S_{123}, S_{13}|X_1, Q) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \tag{B.34}$$

$$\triangleq n\bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + nh(S_{123}, S_{13}|X_1, Q) \tag{B.35}$$

$$- h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon \tag{B.36}$$

$$nR_3 \leq nh(Y_3|X_1, Q) - nh(S_{123}, S_{13}|X_1, Q) + n\epsilon$$

$$\triangleq n\bar{F}_3^{'}(\mathtt{M}_3) + n\epsilon \tag{B.37}$$

When $\mathsf{INR}_{12} < \mathsf{INR}_{13}$, the three upper bounds on R_3 become the following.

$$nR_3 \leq n\left[(Y_3|Q) - h(S_{13}|X_1, Q)\right] + nh(S_{13}|X_1, Q)$$

$$- h(S_{13}^n) + n\epsilon$$

$$\triangleq n\bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + nh(S_{13}|X_1, Q)$$

$$- h(S_{13}^n) + n\epsilon$$

$$nR_3 \leq n\left[h(Y_3|T_{123}, Q) - h(S_{13}|X_1, Q)\right] + nh(S_{13}|X_1, Q)$$

$$- h(S_{13}^n|T_{123}^n) + n\epsilon$$

$$\triangleq n\bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + nh(S_{13}|X_1, Q) - h(S_{13}^n|T_{123}^n)$$

$$+ n\epsilon$$

$$nR_3 \leq n\left[h(Y_3|X_1, Q) - h(S_{13}|X_1)\right] + n\epsilon$$

$$= n\bar{F}_3^{'}(\mathtt{M}_3) + n\epsilon$$

B.3.4 An Intermediate Outer Bound

We first consider the case of $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$. Since the goal of deriving the outer bound is to quantify the gap from the inner bound to itself, we want our outer bound to have close identical algebraic structure to the inner bound. Note there is a one to one correspondence between the set functions in the inner and outer bounds. For each inequality in the inner bound (c.f. Theorem 4.2) except (4.109), (4.111), (4.116), (4.117) and (4.125), we construct a corresponding inequality by replacing the inner bound set function $F_i(\cdot)$ with the outer bound set function $\bar{F}_i(\cdot)$. By assembling these 28 inequalities, we may get an intermediate outer bound for the MIMO IC-ZIC expressed in terms of the genie and channel side informations. To accomplish such a task, we meed to make sure all the unsingle-letterized entropy terms in the individual upper bounds could be either bounded or eliminated. As a matter of fact, all the unsingle-letterized entropy terms can indeed be removed when we linearly combine these individual upper bounds according to the structure of the inner bound. Let us justify the derivation of the following two inequalities which belong to the intermediate outer bound to follow in Lemma.

The 6th inequality in the inner bound (c.f. Theorem 4.2) suggests an outer bound inequality with $\bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2(\mathtt{M}_{2p})$ on the right side; therefore we add inequalities (B.5) and (B.30) as follows.

$$\begin{aligned} n(R_1 + R_2) &\leq n\bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + nh(S_{21}|X_2, Q) - h(S_{21}^n) \\ &\qquad n\bar{F}_2(\mathtt{M}_{2p}) - nh(S_{21}|X_2, Q) + h(T_{21}^n) + n\epsilon^{'} \\ &= n\bar{F}_1(\mathtt{M}_1, \mathtt{M}_{21}) + n\bar{F}_2(\mathtt{M}_{2p}) + n\epsilon^{'} \end{aligned}$$

In the derivation of these inequalities, unsingle-letterized entropy terms eliminate each other. The inequality (4.119) in the inner bound suggests an outer bound inequality with $\bar{F}_1(\mathtt{M}_{1p}) + \bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3)$ on the right side, so this inequality must be the combination of inequalities (B.19), (B.21) and (B.35), which is

$$\begin{aligned} &n(R_1 + R_2 + R_3) \\ &\leq n\bar{F}_1(\mathtt{M}_{1p}) + n\bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + n\bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \end{aligned}$$

$$
\begin{aligned}
&- nh(S_{123}, S_{12}, S_{13}|X_1, Q) + h(S_{123}^n, S_{12}^n, S_{13}^n) \\
&+ nh(S_{12}|X_1, Q) - h(S_{12}^n) + nh(S_{123}, S_{13}|X_1, Q) \\
&- h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon^{'} \\
\stackrel{(a)}{=}\ & n\bar{F}_1(\mathtt{M}_{1p}) + n\bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + n\bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n) - h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon^{'} \\
\stackrel{(b)}{=}\ & n\bar{F}_1(\mathtt{M}_{1p}) + n\bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + n\bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n) \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon^{'} \\
\stackrel{(c)}{\leq}\ & n\bar{F}_1(\mathtt{M}_{1p}) + n\bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + n\bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n, X_1^n) \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon^{'} \\
\stackrel{(d)}{=}\ & n\bar{F}_1(\mathtt{M}_{1p}) + n\bar{F}_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + n\bar{F}_3(\mathtt{M}_{13}, \mathtt{M}_3) \\
&+ h(S_{123}^n, S_{12}^n, S_{13}^n, T_{123}^n) - r_{123} \log 2\pi e \\
&- h(S_{12}^n) - h(S_{123}^n, S_{13}^n|T_{123}^n) + n\epsilon^{'} \\
=\ & n\left[-r_{123} \log 2\pi e + h(T_{123}|S_{12})\right] - I(S_{123}^n, S_{13}^n; S_{12}^n|T_{123}^n) \\
\stackrel{(e)}{\leq}\ & n\eta + n\epsilon^{'}.
\end{aligned}
$$

The rationales of the steps (a)-(e) are as follows: (a) the value of entropy terms can be explicitly computed and $nh(S_{123}, S_{12}, S_{13}|X_1, Q) = nh(S_{12}|X_1, Q) + nh(S_{123}, S_{13}|X_1, Q)$; (b) chain rule of the conditional entropy; (c) conditioning reduces entropy, therefore the negative term $h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n, X_1^n) \geq h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n, X_1^n)$; (d) given the fact that T_{123} and S_{123} are identically distributed, we have

$$h(T_{123}^n|S_{123}^n, S_{12}^n, S_{13}^n, X_1^n) = h(T_{123}^n|X_1^n) = nh(T_{123}|X_1) = n(S_{123}|X_1) = r_{123} \log 2\pi e$$

. The value of $h(S_{123}|X_1)$ is given by #. (e) $h(T_{123}|S_{12}) = h(S_{123}|S_{12})$ by the distributions of genie and channel side informations, and the fact that $h(S_{123}|S_{12})$ is upper bounded by $n\eta + r_{123} \log 2\pi e$ according to (A.30) in Section A.3.

The other 25 inequalities can be justified in a similar fashion. For each desired combination of the individual upper bounds, the involved unsingle-letterized entropy terms are either upper bounded by 0 or $n\eta$.

When $\mathsf{INR}_{12} \leq \mathsf{INR}_{13}$, we have to use the same technique to derive another 28 inequalities for another intermediate outer bound with the same process as in the case of $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$. Again, for each desired combination of the individual upper bounds, the involved unsingle-letterized entropy terms are either upper bounded by 0 or $n\eta$ (with a different value when $\mathsf{INR}_{12} < \mathsf{INR}_{13}$, c.f. (4.185)). Thus we unify these two intermediate outer bounds in Lemma #, where all the inequalities except the individual rate upper bounds have η on the right hand side.

Lemma B.1. *Let $\mathcal{P}_{\text{o}}$ be the set of distributions P_{o} of joint random variables (Q, X_1, X_2, X_3) that can be factored as*

$$p(x_1, x_2, x_3) = p(q)p(x_1|q)p(x_2|q)p(x_3|q),$$

and define the following region $\mathcal{R}_{\text{o}}^{'}(P_{\text{o}})$ given by (B.38)-(B.65). *Then we have*

$$\mathcal{C} \subseteq \bigcup_{P_{\text{o}}} \mathcal{R}_{\text{o}}^{'}(P_{\text{o}}).$$

$$\mathcal{R}_{\text{o}}^{'}(P_{\text{o}}) \triangleq \big\{(R_1, R_2, R_3) \in \mathbb{R}_+^3 :$$

$$R_1 \leq \bar{F}_1^{'}(\mathtt{M}_1) \tag{B.38}$$

$$R_2 \leq \bar{F}_2^{'}(\mathtt{M}_2) \tag{B.39}$$

$$R_3 \leq \bar{F}_3^{'}(\mathtt{M}_3) \tag{B.40}$$

$$R_1 + R_2 \leq \bar{F}_1^{'}(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \eta \tag{B.41}$$

$$R_1 + R_2 \leq \bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \eta \tag{B.42}$$

$$R_1 + R_2 \leq \bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \eta \tag{B.43}$$

$$R_1 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.44}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.45}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.46}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.47}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.48}$$

$$R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.49}$$

$$R_1 + 2R_2 \leq \bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \eta \tag{B.50}$$

$$2R_1 + R_2 \leq \bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \eta \tag{B.51}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.52}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.53}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.54}$$

$$2R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.55}$$

$$R_1 + 2R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.56}$$

$$R_1 + 2R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.57}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.58}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.59}$$

$$2R_1 + 2R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.60}$$

$$2R_1 + R_2 + 2R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_{2p}) + 2\bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.61}$$

$$2R_1 + R_2 + 2R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{13}, \mathtt{M}_3) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.62}$$

$$3R_1 + R_2 + R_3 \leq \bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_1, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) + \eta \tag{B.63}$$

$$\begin{aligned} 3R_1 + 2R_2 + 2R_3 \leq\ & 2\bar{F}_1^{'}(\mathtt{M}_{1p}) + \bar{F}_1^{'}(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) + \bar{F}_2^{'}(\mathtt{M}_{12}, \mathtt{M}_{2p}) + \bar{F}_2^{'}(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_2) + 2\bar{F}_3^{'}(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \\ & + \eta \end{aligned} \tag{B.64}$$

$$2R_1+3R_2+R_3 \le \bar{F}_1^{'}(\mathtt{M}_{1p},\mathtt{M}_{21})+\bar{F}_1^{'}(\mathtt{M}_{12},\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21})+2\bar{F}_2^{'}(\mathtt{M}_{2p})+\bar{F}_2^{'}(\mathtt{M}_{123},\mathtt{M}_{12},\mathtt{M}_{2})+\bar{F}_3^{'}(\mathtt{M}_{123},\mathtt{M}_{13},\mathtt{M}_{3})$$
$$+\eta\} \tag{B.65}$$

Remark B.1. Recall there are five inequalities in the inner bound that do not have counterpoints in the outer bound. This is because if we add the individual upper bounds accordingly, the rate variable combination on the left hand side will not match the rate combination in the inner bound. For example, there are three set functions on the right hand side of inequality (4.116), but only two rate variables in the left hand side.

B.3.5 The Single Region Outer Bound

The intermediate upper bound is a union of polytopes over all admissible input distributions. To establish a single region outer bound, we maximize the set functions $\bar{F}_1^{'}(\cdot)$, $\bar{F}_2^{'}(\cdot)$ and $\bar{F}_3^{'}(\cdot)$ by optimizing the input distribution $p(x_1,x_2,x_3,q)$. First of all, the time sharing is disabled. The region $\mathcal{R}_{\mathrm{o}}^{'}(P_{\mathrm{o}})$ will not shrink because removing the random variable Q will not decrease the positive conditional entropy terms, and the negative entropy terms are entropies of the Gaussian noises which are independent of Q, for example, $h(Y_1|X_1,Q)=h(Z_1|Q)=h(Z_1)$. The positive entropy terms in the set functions are upper bounded below. Each term reaches its maximum value when X_1, X_2 and X_3 are independent Gaussian random vectors. For random vectors X and Y with zero mean and some fixed joint covariance, the conditional differential entropy of X given Y is maximized when X and Y are joint Gaussian [44, Lemma 1]. We also assumed the inputs have zero means, i.e., $E(X_i)=\mathbf{0}$ for $i\in\{1,2,3\}$, as non-zero means only contribute to power inefficiency. We prove the set function $\bar{F}_1(\mathtt{M}_{1p})$ when $\mathsf{INR}_{12}\ge\mathsf{INR}_{13}$ in the sequence of steps leading to (B.66). Steps labeled (a)-(c) hold true for the following rationale: (a1)-(a2), the covariance matrix $\mathrm{Cov}[U_{13}(U_{13}^{-1}Z_3)^{(1:r_{123})}]$ satisfies

$$\begin{aligned}
&\mathrm{Cov}[U_{13}U_{13}^{-1(1:r_{123})}Z_3]\\
&=U_{13}U_{13}^{-1(1:r_{123})}I_3\left(U_{13}^{-1(1:r_{123})}\right)^{\dagger}U_{13}^{\dagger}\\
&=U_{13}\begin{pmatrix} I_{r_{123}} & \mathbf{0}_{r_{123}\times(N_3-r_{123})}\\ \mathbf{0}_{(N_3-r_{123})\times r_{123}} & \mathbf{0}_{(N_3-r_{123})\times N_3-r_{123})}\end{pmatrix}U_{13}^{-1}\\
&\preceq U_{13}I_{N_3}U_{13}^{-1}=I_{N_3}.
\end{aligned}$$

For the two p.s.d. matrices A and B, if $A \preceq B$, then $B^{-1} \preceq A^{-1}$ and $-A^{-1} \preceq -B^{-1}$. Therefore, the inverse matrix term with a minus sign will be "greater" if we replace $\mathrm{Cov}[U_{13}(U_{13}^{-1}Z_3)^{(1:r_{123})}]$ with I_{N_3}. Since $\log|\cdot|$ is a monotonically non-decreasing function on the cone of p.s.d. matrices, the value of the entire entropy term will increase after this replacement. (b) follows from the Woodbury's identity. (c) $\mathrm{Tr}(Q_1) \leq P_1$ implies $Q_1 \preceq P_1 I_{M_1}$ and Lemma 6 in [27], also $\mathrm{Tr}(Q_2) \leq P_1$ implies $Q_2 \preceq P_2 I_{M_2}$.

$$
\begin{aligned}
&\bar{F}_1^{'}(\mathtt{M}_{13}, \mathtt{M}_{1p}, \mathtt{M}_{21}) \\
&\leq h(Y_1|T_{123}, T_{12}, Q) - h(S_{21}|X_2, Q) \\
&\leq h(Y_1|T_{123}, T_{12}) - h(Z_1) \\
&= h(Y_1, T_{123}, T_{12}) - h(T_{123}, T_{12}) - N_1 \log 2\pi e \\
&\leq \log \begin{vmatrix} \mathrm{Var}[Y_1] & \mathrm{Cov}[Y_1, T_{123}] & \mathrm{Cov}[Y_1, T_{12}] \\ \mathrm{Cov}[T_{123}, Y_1] & \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, Y_1] & \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{vmatrix} - \log \begin{vmatrix} \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{vmatrix} \\
&= \log \left| \mathrm{Var}[Y_1] - \begin{pmatrix} \mathrm{Cov}[Y_1, T_{123}] & \mathrm{Cov}[Y_1, T_{12}] \end{pmatrix} \cdot \begin{pmatrix} \mathrm{Var}[T_{123}] & \mathrm{Cov}[T_{123}, T_{12}] \\ \mathrm{Cov}[T_{12}, T_{123}] & \mathrm{Var}[T_{12}] \end{pmatrix}^{-1} \right. \\
&\qquad \left. \cdot \begin{pmatrix} \mathrm{Cov}[T_{123}, Y_1] \\ \mathrm{Cov}[T_{12}, Y_1] \end{pmatrix} \right| \\
&\overset{(a1)}{\leq} \log \left| I_{N_1} + |h_{11}|^2 H_{11} Q_1 H_{11}^{\dagger} + |h_{21}|^2 H_{21} Q_2 H_{21}^{\dagger} - \begin{pmatrix} h_{11} h_{13}^* H_{11} Q_1 G_{13}^{\dagger} & h_{11} h_{12}^* H_{11} Q_1 H_{12}^{\dagger} \end{pmatrix} \right. \\
&\qquad \cdot \begin{pmatrix} U_{13} \begin{pmatrix} I_{r_{123}} & \\ & \mathbf{0}_{(N_3 - r_{123}) \times N_3 - r_{123})} \end{pmatrix} U_{13}^{-1} + |h_{13}|^2 G_{13} Q_1 G_{13}^{\dagger} & h_{13} h_{12}^* G_{13} Q_1 H_{12}^{\dagger} \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^{\dagger} & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^{\dagger} \end{pmatrix}^{-1} \\
&\qquad \left. \cdot \begin{pmatrix} h_{13} h_{11}^* G_{13} Q_1 H_{11}^{\dagger} \\ h_{12} h_{11}^* H_{12} Q_1 H_{11}^{\dagger} \end{pmatrix} \right| \\
&\overset{(a2)}{\leq} \log \left| I_{N_1} + |h_{11}|^2 H_{11} Q_1 H_{11}^{\dagger} + |h_{21}|^2 H_{21} Q_2 H_{21}^{\dagger} - \begin{pmatrix} h_{11} h_{13}^* H_{11} Q_1 G_{13}^{\dagger} & h_{11} h_{12}^* H_{11} Q_1 H_{12}^{\dagger} \end{pmatrix} \right.
\end{aligned}
$$

$$
\cdot \begin{pmatrix} I_{N_3} + |h_{13}|^2 G_{13} Q_1 G_{13}^\dagger & h_{13} h_{12}^* G_{13} Q_1 H_{12}^\dagger \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^\dagger & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^\dagger \end{pmatrix}^{-1} \begin{pmatrix} h_{13} h_{11}^* G_{13} Q_1 H_{11}^\dagger \\ h_{12} h_{11}^* H_{12} Q_1 H_{11}^\dagger \end{pmatrix} \Bigg|
$$

$$
= \log \Bigg| I_{N_1} + |h_{21}|^2 H_{21} Q_2 H_{21}^\dagger + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \Bigg[I_{M_1} - \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix}
$$

$$
\cdot \begin{pmatrix} I_{N_3} + |h_{13}|^2 G_{13} Q_1 G_{13}^\dagger & h_{13} h_{12}^* G_{13} Q_1 H_{12}^\dagger \\ h_{12} h_{13}^* H_{12} Q_1 G_{13}^\dagger & I_{N_2} + |h_{12}|^2 H_{12} Q_1 H_{12}^\dagger \end{pmatrix}^{-1} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \Bigg] Q_1^{\frac{1}{2}} H_{11}^\dagger \Bigg|
$$

$$
= \log \Bigg| I_{N_1} + |h_{21}|^2 H_{21} Q_2 H_{21}^\dagger + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \Bigg[I_{M_1} - \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix}
$$

$$
\cdot \left(I_{N_2+N_3} + \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \right)^{-1} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \Bigg] Q_1^{\frac{1}{2}} H_{11}^\dagger \Bigg|
$$

$$
\stackrel{(b)}{=} \log \Bigg| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \left(I_{M_1} + \begin{pmatrix} h_{13}^* Q_1^{\frac{1}{2}} G_{13}^\dagger & h_{12}^* Q_1^{\frac{1}{2}} H_{12}^\dagger \end{pmatrix} \begin{pmatrix} h_{13} G_{13} Q_1^{\frac{1}{2}} \\ h_{12} H_{12} Q_1^{\frac{1}{2}} \end{pmatrix} \right)^{-1} Q_1^{\frac{1}{2}} H_{11}^\dagger
$$

$$
+ |h_{21}|^2 H_{21} Q_2 H_{21}^\dagger \Bigg|
$$

$$
= \log \left| I_{N_1} + |h_{11}|^2 H_{11} Q_1^{\frac{1}{2}} \left(I_{M_1} + Q_1^{\frac{1}{2}} (|h_{13}|^2 G_{13}^\dagger G_{13} + |h_{12}|^2 H_{12}^\dagger H_{12}) Q_1^{\frac{1}{2}} \right)^{-1} Q_1^{\frac{1}{2}} H_{11}^\dagger + |h_{21}|^2 H_{21} Q_2 H_{21}^\dagger \right|
$$

$$
\stackrel{(c)}{\leq} \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} \left(I_{M_1} + \rho^{\alpha_{13}} G_{13}^\dagger G_{13} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} \right)^{-1} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger \right|
$$

$$
= \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{13,1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} H_{21}^\dagger \right| \tag{B.66}
$$

In the case of $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, the set function $\bar{F}_1(\mathtt{M}_{1p})$ and also the other set functions can be proved in a similar fashion. The proof is completed.

B.4 Proof of Theorem 4.4

The gap is quantified in two steps. In the first step, we quantify the 28 inequalities in $\mathcal{R}_{\text{in}}$ which can be seen as a result of replacing the set function $\bar{F}_i(\mathtt{M}_{\phi_i})$ with $F_i(\mathtt{M}_{\phi_i})$ and removing η from the inequalities of the outer bound. Then we quantity the gap resulting from the inequalities (4.109), (4.111), (4.116), (4.117) and (4.125) in $\mathcal{R}_{\text{in}}$. The overall gap will thus be determined.

B.4.1 The Gap between the Twenty Eight Inequalities in $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$

Fact B.1 states the following relationship between the matrices Q_{1p}, $Q_{12}+Q_{1p}$, $Q_{13}+Q_{1p}$ and $Q_{12}+Q_{13}+Q_{1p}$ by (4.141)-(4.144) with matrices K_{1p}, $K_{12,1p}$, $K_{13,1p}$ and $K_{12,13,1p}$ by (4.161)-(4.164). The result has been proved in Lemma A.2 for the case $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, the result for $\mathsf{INR}_{12} < \mathsf{INR}_{13}$ can be proved similarly.

Fact B.1. *The identities given by* (B.72)-(B.75) *hold. Furthermore, the covariance matrices* Q_{1p}, $Q_{12}+Q_{1p}$, $Q_{13}+Q_{1p}$ *and* $Q_{12}+Q_{13}+Q_{1p}$ *can be lower bounded as follows.*

$$Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}} K_{1p} \tag{B.67}$$

$$Q_{12}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}} K_{12,1p} \tag{B.68}$$

$$Q_{13}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}} K_{13,1p} \tag{B.69}$$

$$Q_{12}+Q_{13}+Q_{1p} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}} K_{12,13,1p} \tag{B.70}$$

$$Q_{1} \succeq \frac{1}{\zeta_{\max}\max\{\lambda^2_{\max}(V_r),1\}} I_{M_1} \tag{B.71}$$

$$Q_{1p} = \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix} V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger + \rho^{\alpha_{12}}H_{12}^\dagger H_{12} + \rho^{\alpha_{13}}H_{13}^\dagger H_{13}\right)^{-1} \tag{B.72}$$

$$Q_{12}+Q_{1p} = \begin{cases} \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix} V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger + \rho^{\alpha_{13}}H_{13}^\dagger H_{13}\right)^{-1} \\ \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix} V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger + \rho^{\alpha_{12}}G_{12}^\dagger G_{12} + \rho^{\alpha_{13}}H_{13}^\dagger H_{13}\right)^{-1} \end{cases} \tag{B.73}$$

$$Q_{13}+Q_{1p}=\begin{cases}\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix}V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger+\rho^{\alpha_{12}}H_{12}^\dagger H_{12}+\rho^{\alpha_{13}}G_{13}^\dagger G_{13}\right)^{-1}\\ \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix}V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger+\rho^{\alpha_{12}}H_{12}^\dagger H_{12}\right)^{-1}\end{cases} \tag{B.74}$$

$$Q_{12}+Q_{13}+Q_{1p}=\begin{cases}\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix}V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger+\rho^{\alpha_{13}}G_{13}^\dagger G_{13}\right)^{-1}\\ \frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix}V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger+\rho^{\alpha_{12}}G_{12}^\dagger G_{12}\right)^{-1}\end{cases} \tag{B.75}$$

$$Q_1=\frac{1}{\mathrm{Tr}(V_pV_p^\dagger)}\left(U\begin{pmatrix}V_rV_r^\dagger & \\ & I_{(M_1-r)}\end{pmatrix}U^\dagger\right)^{-1} \tag{B.76}$$

We quantify the gap between $F_1(\mathtt{M}_{1p})$ and $\bar{F}_1(\mathtt{M}_{1p})$ in the following.

$$\begin{aligned}
&F_1(\mathtt{M}_{1p})\\
&\overset{(a)}{\geq}\log\left|I_{N_1}+\rho^{\alpha_{11}}H_{11}Q_{1p}H_{11}^\dagger\right|-\beta_1\\
&\overset{(b)}{\geq}\log\left|I_{N_1}+\rho^{\alpha_{11}}H_{11}K_{1p}H_{11}^\dagger\right|-\min\left\{M_1,N_1\right\}\\
&\qquad\cdot\left(\log\left(\zeta_{\max}\max\left\{\lambda_{\max}^2(V_r),1\right\}\right)\right)^+-\beta_1\\
&\overset{(b)}{\geq}\log\left|I_{N_1}+\rho^{\alpha_{11}}H_{11}K_{1p}H_{11}^\dagger\right|-\min\{M_1+M_2,N_1\}\\
&\qquad\cdot\log\max\left\{\zeta_{\max}\max\left\{\lambda_{\max}^2(V_r),1\right\},M_2\right\}-\beta_1\\
&=\bar{F}(\mathtt{M}_{1p})-\delta_1-\beta_1
\end{aligned}$$

Step (a) holds true because $\log|\cdot|$ is a non-decreasing function over the cone of p.s.d. matrices so removing the term $\rho^{\alpha_{21}}H_{21}Q_{2p}H_{21}^\dagger$ will not increase the value of $F_1(\mathtt{M}_{1p})$. Step (b) is true due to (B.67) in Fact B.1.

The gap between $F_1(\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21})$ and $\bar{F}_1(\mathtt{M}_{13},\mathtt{M}_{1p},\mathtt{M}_{21})$ is bounded as follows.

$$\begin{aligned}
&F_1(\mathtt{M}_{1p},\mathtt{M}_{21})\\
&=\log|I_{N_1}+\rho^{\alpha_{11}}H_{11}Q_{1p}H_{11}^\dagger+\rho^{\alpha_{21}}H_{21}Q_2H_{21}^\dagger|-\beta_1
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\geq} \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger} + \rho^{\alpha_{21}} H_{21} H_{21}^{\dagger}\right| \\
&\qquad - \min\{M_1 + M_2, N_1\} \\
&\qquad \cdot \log\max\left\{\zeta_{\max}\max\left\{\lambda_{\max}^2(V_r), 1\right\}, M_2\right\} - \beta_1 \\
&= \bar{F}_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) - \delta_1 - \beta_1
\end{aligned}
$$

Step (a) is true because the rank of $\rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^{\dagger} + \rho^{\alpha_{21}} H_{21} H_{21}^{\dagger}$ should not exceed $\min\{M_1 + M_2, N_1\}$. In a similar way, the gap resulting from replacing $F_i(\mathtt{M}_{\phi_i})$ with $\bar{F}_i(\mathtt{M}_{\phi_i})$ is upper bounded by $\delta_i + \beta_i$. Note in each of these 28 inequalities, the coefficient of R_i in the left hand side is the same as the number of the set functions (with subscript i) $\bar{F}_i(\cdot)$ that appear on the right hand side; therefore, we let the gap $\delta_i + \beta_i$ be absorbed by the individual gap $n_i^{(1)}$.

Lastly, there is a constant η in all but the first three inequalities in the outer bound. Since R_1 always appears on the left hand side of these 25 inequalities, we let η be absorbed by $n_1^{(1)}$. So far, we quantified the gap between the 28 inequalities in the inner bound $\mathcal{R}_{\text{in}}$ to their counterpoints in the outer bound $\mathcal{R}_{\text{o}}$.

B.4.2 Preliminaries for the Proof of The Gap Induced From the Other Five Inequalities in the Inner Bound

Before we prove the gap induced from the other 5 inequalities in the inner bound, we need the following preliminary result to support the proof. Throughout, we assume $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, and the result of the case $\mathsf{INR}_{12} < \mathsf{INR}_{13}$ can be obtained in a similar fashion.

Lemma B.2. *The inner bound set functions $F_1(\mathtt{M}_{1p})$, $F_1(\mathtt{M}_{13}, \mathtt{M}_{1p})$, $F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p})$, $F_1(\mathtt{M}_{1p}, \mathtt{M}_{21})$, $F_2(\mathtt{M}_{2p})$, $F_2(\mathtt{M}_{12}, \mathtt{M}_{2p})$, $F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})$ and $F_3(\mathtt{M}_{13}, \mathtt{M}_{3})$ can be lower bounded as follows,*

$$
\begin{aligned}
&F_1(\mathtt{M}_{1p}) \\
&\geq h\left(h_{11} H_{11} X_1 + Z_1, h_{12} H_{12} X_1 + Z_2, h_{13} H_{13} X_1 + Z_3\right) \\
&\quad - h\left(h_{12} H_{12} X_1 + Z_2, h_{13} H_{13} X_1 + Z_3\right) \\
&\quad - N_1 \log(2\pi e) - \beta_1 - \gamma_{11} \qquad \text{(B.77)}
\end{aligned}
$$

$$
F_1(\mathtt{M}_{13}, \mathtt{M}_{1p})
$$

$$\geq h\left(h_{11}H_{11}X_1 + Z_1, h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right)$$
$$- h\left(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right)$$
$$- N_1 \log(2\pi e) - \beta_1 - \gamma_{11} \tag{B.78}$$

$$F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p})$$
$$\geq h\left(h_{11}H_{11}X_1 + Z_1, h_{13}G_{13}X_1 + Z_3\right)$$
$$- h\left(h_{13}G_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) - \beta_1 - \gamma_{11} \tag{B.79}$$

$$F_1(\mathtt{M}_{1p}, \mathtt{M}_{21})$$
$$\geq h(h_{21}H_{21}X_2 + Z_1) - N_1 \log(2\pi e) - \beta_1 - \gamma_{21} \tag{B.80}$$

$$F_2(\mathtt{M}_{2p})$$
$$\geq h(h_{22}H_{22}X_2 + Z_2, h_{21}H_{21}X_2 + Z_1)$$
$$- h(h_{21}H_{21}X_2 + Z_1) - N_2 \log(2\pi e) - \beta_2 - \gamma_{22} \tag{B.81}$$

$$F_2(\mathtt{M}_{12}, \mathtt{M}_{2p})$$
$$\geq h(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3)$$
$$- h(h_{13}G_{13}X_1 + Z_3) - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \tag{B.82}$$

$$F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})$$
$$\geq h(h_{12}H_{12}X_1 + Z_2) - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \tag{B.83}$$

$$F_3(\mathtt{M}_{13}, \mathtt{M}_3)$$
$$\geq h(h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3, h_{13}G_{13}X_1 + Z_3^{'})$$
$$- h(h_{13}G_{13}X_1 + Z_3^{'}) - N_3 \log(2\pi e) - \beta_3 - \delta_3, \tag{B.84}$$

where $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$, $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$, $Z_3^{'} \sim \mathcal{CN}(0, I_{N_3})$ *and* $Z_3 \perp Z_3^{'}$ *in the right hand side of above inequalities.*

Proof. The proof of lower bound on $F_1(\mathtt{M}_{1p})$ is demonstrated in the sequence of steps leading to (B.85). Step (a) holds true because of the Woodbury's identity; step (b) is true because for any matrices A, B, C and D, we have $\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D||A - BD^{-1}C|$ if D is invertible. Other lower bounds can be obtained similarly.

$$
\begin{aligned}
& F_1(\mathtt{M}_{1p}) \\
&= \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger + \rho^{\alpha_{21}} H_{21} Q_{2p} H_{21}^\dagger \right| - \beta_1 \\
&\geq \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} Q_{1p} H_{11}^\dagger \right| - \beta_1 \\
&\geq \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{1p} H_{11}^\dagger \right| - \beta_1 - \gamma_{11} \\
&= \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} \left(I_{M_1} + \rho^{\alpha_{12}} H_{12}^\dagger H_{12} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1} H_{11}^\dagger \right| - \beta_1 - \gamma_{11} \\
&= \log \left| I_{N_1} + \rho^{\alpha_{11}} H_{11} \left(I_{M_1} + \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \right)^{-1} H_{11}^\dagger \right| - \beta_1 - \gamma_{11} \\
&\overset{(a)}{=} \log \Bigg| I_{N_1} + \rho^{\alpha_{11}} H_{11} \Bigg(I_{M_1} - \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \\
&\quad \cdot \left(I_{N_2+N_3} + \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \right)^{-1} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \Bigg) H_{11}^\dagger \Bigg| \\
&\quad - \beta_1 - \gamma_{11} \\
&= \log \Bigg| I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger - \rho^{\alpha_{11}} H_{11} \Bigg(\begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \\
&\quad \cdot \left(I_{N_2+N_3} + \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \right)^{-1} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \Bigg) H_{11}^\dagger \Bigg| \\
&\quad - \beta_1 - \gamma_{11} \\
&\overset{(b)}{=} \log \begin{vmatrix} I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger & h_{11}\sqrt{P_1} H_{11} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \\ h_{11}\sqrt{P_1} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} H_{11}^\dagger & I_{N_2+N_3} + \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \end{vmatrix} \\
&\quad - \log \left| I_{N_2+N_3} + \begin{pmatrix} h_{12}\sqrt{P_1} H_{12} \\ h_{13}\sqrt{P_1} H_{13} \end{pmatrix} \begin{pmatrix} h_{12}\sqrt{P_1} H_{12}^\dagger & h_{13}\sqrt{P_1} \; H_{13}^\dagger \end{pmatrix} \right| - \beta_1 - \gamma_{11} \\
&= h\left(h_{11} H_{11} X_1 + Z_1, h_{12} H_{12} X_1 + Z_2, h_{13} H_{13} X_1 + Z_3\right) - h\left(h_{12} H_{12} X_1 + Z_2, h_{13} H_{13} X_1 + Z_3\right) - N_1 \log(2\pi e) \\
&\quad - \beta_1 - \gamma_{11}
\end{aligned} \tag{B.85}
$$

□

Lemma B.3. *The outer bound set functions $\bar{F}_1(\mathtt{M}_1)$, $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p})$, $\bar{F}_2(\mathtt{M}_2)$ and $\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$*

$$\bar{F}_1(\mathtt{M}_1) = h(h_{11}H_{11}X_1 + Z_1) - N_1 \log(2\pi e) \tag{B.86}$$

$$\begin{aligned}\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) &= h\left(h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3\right) \\ &\quad - h\left(h_{13}H_{13}X_1 + Z_3\right) - N_1 \log(2\pi e)\end{aligned} \tag{B.87}$$

$$\bar{F}_2(\mathtt{M}_2) = h(h_{22}H_{22}X_2 + Z_2) - N_2 \log(2\pi e) \tag{B.88}$$

$$\begin{aligned}\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) &= h(h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3) \\ &\quad - N_3 \log(2\pi e)\end{aligned} \tag{B.89}$$

where $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$, and $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$ in the right hand side of the inequalities above.

Proof. We prove the second identity as an example in the following. The rest of identities can be verified similarly. Recall the definition of the set function $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p})$ (c.f. (4.168)) and the fact that the outer bound is obtained by choosing X_1, X_2 and X_3 to be Gaussian vectors (c.f. Section B.3.5). We have

$$\begin{aligned}
&\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11} K_{12,1p} H_{11}^\dagger\right| \\
&= \log\left|I_{N_1} + \rho^{\alpha_{11}} H_{11}\left(I_{M_1} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13}\right)^{-1} H_{11}^\dagger\right| \\
&\stackrel{(a)}{=} \log|I_{N_1} + \rho^{\alpha_{11}} H_{11} \\
&\qquad \cdot \left(I_{M_1} - \rho^{\alpha_{13}} H_{13}^\dagger (I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger)^{-1} H_{13}\right) H_{11}^\dagger\Big| \\
&= \log\Big|I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger \\
&\qquad - \rho^{\alpha_{11}} H_{11} \rho^{\alpha_{13}} H_{13}^\dagger (I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger)^{-1} H_{13} H_{11}^\dagger\Big| \\
&= \log\begin{vmatrix} I_{N_1} + \rho^{\alpha_{11}} H_{11} H_{11}^\dagger & h_{11} h_{13}^\dagger P_1 H_{11} H_{13}^\dagger \\ h_{13} h_{11}^\dagger P_1 H_{13} H_{11}^\dagger & I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger \end{vmatrix} \\
&\qquad - \log\left|I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger\right|
\end{aligned}$$

$$= h(h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3)$$

$$- h(h_{13}H_{13}X_1 + Z_3) - N_1 \log(2\pi e).$$

Step (a) holds due to the Woodbury's identity. The proof is completed. □

Lemma B.4. *The following identities hold when $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$, and $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$.*

$$h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right)$$
$$= \log\left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} \right)^{-1} \Sigma_{12}^\dagger \right|$$
$$+ N_2 \log(2\pi e) \tag{B.90}$$

$$h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}G_{13}X_1 + Z_3\right)$$
$$= \log\left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13} \right)^{-1} \Sigma_{12}^\dagger \right|$$
$$+ N_2 \log(2\pi e) \tag{B.91}$$

$$h\left(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2\right)$$
$$= \log\left| I_{N_3} + \rho^{\alpha_{13}} \Lambda_{13} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{12}} \Sigma_{12}^\dagger \Sigma_{12} \right)^{-1} \Lambda_{13}^\dagger \right|$$
$$+ N_3 \log(2\pi e) \tag{B.92}$$

Proof. We compute the first equation in the sequence of steps leading to (B.93), and the rest of the identities can be shown similarly. The proof is completed.

$$h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right)$$
$$= h(h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3) - h(h_{13}H_{13}X_1 + Z_3)$$
$$= \log \begin{vmatrix} I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^\dagger & h_{12} h_{13}^\dagger P_1 H_{12} H_{13}^\dagger \\ h_{13} h_{12}^\dagger P_1 H_{13} H_{12}^\dagger & I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger \end{vmatrix} - \log\left| I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger \right| + N_2 \log(2\pi e)$$
$$= \log\left| I_{N_2} + \rho^{\alpha_{12}} H_{12} H_{12}^\dagger - h_{12} h_{13}^\dagger P_1 H_{12} H_{13}^\dagger (I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger)^{-1} h_{13} h_{12}^\dagger P_1 H_{13} H_{12}^\dagger \right| + N_2 \log(2\pi e)$$
$$= \log\left| I_{N_2} + \rho^{\alpha_{12}} H_{12} \left(I_{M_1} - \rho^{\alpha_{13}} H_{13}^\dagger (I_{N_3} + \rho^{\alpha_{13}} H_{13} H_{13}^\dagger)^{-1} H_{13} \right) H_{12}^\dagger \right| + N_2 \log(2\pi e)$$
$$= \log\left| I_{N_2} + \rho^{\alpha_{12}} H_{12} \left(I_{M_1} + \rho^{\alpha_{13}} H_{13}^\dagger H_{13} \right)^{-1} H_{12}^\dagger \right| + N_2 \log(2\pi e)$$

$$
\begin{aligned}
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} U_{12} \Sigma_{12} V^\dagger \left(I_{M_1} + \rho^{\alpha_{13}} V \Sigma_{13}^\dagger U_{13}^\dagger U_{13} \Sigma_{13} V^\dagger \right)^{-1} V \Sigma_{12}^\dagger U_{12}^\dagger \right| + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} V^\dagger \left(I_{M_1} + \rho^{\alpha_{13}} V \Sigma_{13}^\dagger \Sigma_{13} V^\dagger \right)^{-1} V \Sigma_{12}^\dagger \right| + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \begin{pmatrix} V_r^\dagger & \mathbf{0}_{r \times (M_1 - r)} \end{pmatrix} U^\dagger \right. \\
&\quad \left. \cdot \left(I_{M_1} + \rho^{\alpha_{13}} U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \Sigma_{13}^\dagger \Sigma_{13} \begin{pmatrix} V_r^\dagger & \mathbf{0}_{r \times (M_1 - r)} \end{pmatrix} U^\dagger \right)^{-1} U \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \Sigma_{12}^\dagger \right| \\
&\quad + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \begin{pmatrix} V_r^\dagger & \mathbf{0}_{r \times (M_1 - r)} \end{pmatrix} \left(I_{M_1} + \rho^{\alpha_{13}} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \Sigma_{13}^\dagger \Sigma_{13} \begin{pmatrix} V_r^\dagger & \mathbf{0}_{r \times (M_1 - r)} \end{pmatrix} \right)^{-1} \right. \\
&\quad \left. \cdot \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \Sigma_{12}^\dagger \right| + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \begin{pmatrix} V_r^\dagger & \mathbf{0}_{r \times (M_1 - r)} \end{pmatrix} \begin{pmatrix} \left(I_r + \rho^{\alpha_{13}} V_r \Sigma_{13}^\dagger \Sigma_{13} V_r^\dagger \right)^{-1} & \\ & I_{(M_1 - r)} \end{pmatrix} \begin{pmatrix} V_r \\ \mathbf{0}_{(M_1-r)^+ \times r} \end{pmatrix} \right. \\
&\quad \left. \cdot \Sigma_{12}^\dagger \right| + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} V_r^\dagger \left(I_r + \rho^{\alpha_{13}} V_r \Sigma_{13}^\dagger \Sigma_{13} V_r^\dagger \right)^{-1} V_r \Sigma_{12}^\dagger \right| + N_2 \log(2\pi e) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} \right)^{-1} \Sigma_{12}^\dagger \right| + N_2 \log(2\pi e)
\end{aligned}
\tag{B.93}
$$

□

Lemma B.5. *$h\left(h_{12} H_{12} X_1 + Z_2 | h_{13} H_{13} X_1 + Z_3\right) \leq h\left(h_{12} H_{12} X_1 + Z_2 | h_{13} G_{13} X_1 + Z_3\right)$ for $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$, and $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$.*

Proof. Using the fact that $\Sigma_{13}^\dagger \Sigma_{13} \succeq \Lambda_{13}^\dagger \Lambda_{13}$, we have

$$
\begin{aligned}
& h\left(h_{12} H_{12} X_1 + Z_2 | h_{13} H_{13} X_1 + Z_3\right) \\
&= \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{13}} \Sigma_{13}^\dagger \Sigma_{13} \right)^{-1} \Sigma_{12}^\dagger \right| \\
&\quad + N_2 \log(2\pi e) \\
&\leq \log \left| I_{N_2} + \rho^{\alpha_{12}} \Sigma_{12} \left(V_r^{-1} V_r^{\dagger -1} + \rho^{\alpha_{13}} \Lambda_{13}^\dagger \Lambda_{13} \right)^{-1} \Sigma_{12}^\dagger \right|
\end{aligned}
$$

$$+ N_2 \log(2\pi e)$$

$$= h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}G_{13}X_1 + Z_3\right).$$

□

Lemma B.6. *$h(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2) \le \eta + N_3 \log(2\pi e)$ for $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$ and $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$.*

Proof. The proof is demonstrated in the sequence leading to B.94. Step (a) is true because of the lower bound on $V_r^{-1}V_r^{\dagger -1}$ by (4.31); step (b) is true according to the structure of Σ_{12} by 4.15. The proof is completed.

$$
\begin{aligned}
& h(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2) \\
&= \log\left| I_{N_3} + \rho^{\alpha_{13}}\Lambda_{13}\left(V_r^{-1}V_r^{\dagger -1} + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12}\right)^{-1}\Lambda_{13}^{\dagger}\right| + N_3\log(2\pi e) \\
&\overset{(a)}{\le} \log\left| I_{N_3} + \rho^{\alpha_{13}}\Lambda_{13}\left(\lambda_{\mathtt{Max}}^{-2} I_r + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12}\right)^{-1}\Lambda_{13}^{\dagger}\right| + N_3\log(2\pi e) \\
&\le \log\left| I_{N_3} + \rho^{\alpha_{13}}\Lambda_{13}\left(I_r + \rho^{\alpha_{12}}\Sigma_{12}^{\dagger}\Sigma_{12}\right)^{-1}\Lambda_{13}^{\dagger}\right| + r_{123}\log\max\{\lambda_{\max}^2(V_r), 1\} + N_3\log(2\pi e) \\
&\overset{(b)}{=} \log\left| I_{N_3} + \rho^{\alpha_{13}}\Lambda_{13}\begin{pmatrix} (1+\rho^{\alpha_{12}})^{-1} I_{r-r_{13}} & & \\ & (I_{r_{123}} + \rho^{\alpha_{12}}C^{\dagger}C)^{-1} & \\ & & I_{r-r_{12}} \end{pmatrix}\Lambda_{13}^{\dagger}\right| \\
&\quad + r_{123}\log\max\{\lambda_{\max}^2(V_r), 1\} + N_3\log(2\pi e) \\
&= \log\left| I_{N_3} + \begin{pmatrix} \rho^{\alpha_{13}} S(I_{r_{123}} + \rho^{\alpha_{12}}C^{\dagger}C)^{-1}S^{\dagger} & \\ & \mathbf{0}_{(N_3-r_{123})\times(N_3-r_{123})} \end{pmatrix}\right| + r_{123}\log\max\{\lambda_{\max}^2(V_r), 1\} \\
&\quad + N_3\log(2\pi e) \\
&\le \log\left| I_{N_3} + \begin{pmatrix} \frac{\rho^{\alpha_{13}}\sigma_{\mathtt{Max}}^2(\Lambda_{13})}{1+\rho^{\alpha_{12}}\sigma_{\mathtt{Min}}^2(\Lambda_{12})} I_{r_{123}} & \\ & \mathbf{0}_{(N_3-r_{123})\times(N_3-r_{123})} \end{pmatrix}\right| + r_{123}\log\max\{\lambda_{\max}^2(V_r), 1\} + N_3\log(2\pi e) \\
&= \log\left| \begin{pmatrix} \left(1 + \frac{\rho^{\alpha_{13}}\sigma_{\mathtt{Max}}^2(\Lambda_{13})}{1+\rho^{\alpha_{12}}\sigma_{\mathtt{Min}}^2(\Lambda_{12})}\right) I_{r_{123}} & \\ & I_{(N_3-r_{123})} \end{pmatrix}\right| + r_{123}\log\max\{\lambda_{\max}^2(V_r), 1\} + N_3\log(2\pi e)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \log \left| \begin{pmatrix} \left(1+\frac{\sigma^2_{\mathtt{Max}}(\Lambda_{13})}{\sigma^2_{\mathtt{Min}}(\Lambda_{12})}\right) I_{r_{123}} & \\ & I_{(N_3-r_{123})} \end{pmatrix} \right| + r_{123}\log\max\{\lambda^2_{\max}(V_r), 1\} + N_3\log(2\pi e) \\
&= r_{123}\log\left(1+\frac{\sigma^2_{\mathtt{Max}}(\Lambda_{13})}{\sigma^2_{\mathtt{Min}}(\Lambda_{12})}\right) + r_{123}\log\max\{\lambda^2_{\max}(V_r), 1\} + N_3\log(2\pi e) \\
&= \eta + N_3\log(2\pi e)
\end{aligned} \tag{B.94}
$$

□

B.4.3 The Gap Induced From the Other Five Inequalities in the Inner Bound

We show the gap from (4.109) to (4.186), (4.111) to (4.187), (4.116) to (4.192), (4.117) to (4.192) and (4.125) to (4.186)+(4.192) one by one. Again, in what follows we assume $\mathsf{INR}_{12} \geq \mathsf{INR}_{13}$, and the gap when $\mathsf{INR}_{12} < \mathsf{INR}_{13}$ can be similarly shown. In the proof of each gap, we shall employ the lemmas that have been developed in the previous subsection, and the variables X_1, X_2 and X_3 have the desired distributions in all those lemmas, i.e., $X_1 \sim \mathcal{CN}(0, P_1 I_{M_1})$, $X_2 \sim \mathcal{CN}(0, P_2 I_{M_2})$ and $X_3 \sim \mathcal{CN}(\mathbf{0}, P_3 I_{N_3})$.

B.4.3.1 The Gap between $F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})$ and $\bar{F}_1(\mathtt{M}_1)$

$$
\begin{aligned}
&F_1(\mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \\
&\overset{(a)}{=} h\left(h_{11}H_{11}X_1 + Z_1, h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right) \\
&\quad - h\left(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right) \\
&\quad - N_1\log(2\pi e) - \beta_1 - \gamma_{11} \\
&\quad + h(h_{12}H_{12}X_1 + Z_2) - N_2\log(2\pi e) - \beta_2 - \gamma_{12} \\
&\overset{(b)}{\geq} h(h_{11}H_{11}X_1 + Z_1) - N_1\log(2\pi e) \\
&\quad + h\left(h_{12}H_{12}X_1 + Z_2,\right. \\
&\quad \left. h_{13}G_{13}X_1 + Z_3 | h_{11}H_{11}X_1 + Z_1, X_1\right) \\
&\quad - h\left(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right) \\
&\quad - \beta_1 - \gamma_{11} + h(h_{12}H_{12}X_1 + Z_2) - N_2\log(2\pi e)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \beta_2 - \gamma_{12} \\
&= \bar{F}_1(\mathtt{M}_1) + N_3 \log(2\pi e) \\
&\quad - h\left(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2\right) - \beta_1 - \gamma_{11} \\
&\quad - \beta_2 - \gamma_{12} \\
&\overset{(c)}{\geq} \bar{F}_1(\mathtt{M}_1) - \beta_1 - \gamma_{11} - \beta_2 - \gamma_{12} - \eta
\end{aligned}
$$

The steps (a)-(c) hold true for the following reasons: (a) by the lower bounds (B.77) and (B.83); (b) chain rule of joint entropy and conditioning reduces entropy; (c) Lemma B.6.

B.4.3.2 The Gap between $F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p})$ and $\bar{F}_2(\mathtt{M}_2)$

$$
\begin{aligned}
&F_1(\mathtt{M}_{1p}, \mathtt{M}_{21}) + F_2(\mathtt{M}_{2p}) \\
&\overset{(a)}{=} h(h_{21}H_{21}X_2 + Z_1) - N_1 \log(2\pi e) - \beta_1 - \gamma_{21} \\
&\quad + h(h_{22}H_{22}X_2 + Z_2, h_{21}H_{21}X_2 + Z_1) \\
&\quad - h(h_{21}H_{21}X_2 + Z_1) - N_2 \log(2\pi e) - \beta_2 - \gamma_{22} \\
&\overset{(b)}{\geq} h(h_{22}H_{22}X_2 + Z_2) - N_2 \log(2\pi e) \\
&\quad + h(h_{21}H_{21}X_2 + Z_1) - N_1 \log(2\pi e) - \beta_1 - \gamma_{21} \\
&\quad + h(h_{21}H_{21}X_2 + Z_1 | h_{22}H_{22}X_2 + Z_2, X_2) \\
&\quad - h(h_{21}H_{21}X_2 + Z_1) - \beta_2 - \gamma_{22} \\
&= \bar{F}_2(\mathtt{M}_2) - \beta_1 - \gamma_{21} - \beta_2 - \gamma_{22}
\end{aligned}
$$

The steps (a)-(b) hold true for the following reasons: (a) by the lower bounds (B.80) and (B.81); (b) chain rule of joint entropy and conditioning reduces entropy.

B.4.3.3 The gap between $F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$

We first quantify the gap between $F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p})$ and $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p})$.

$$
\begin{aligned}
& F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) \\
& \overset{(a)}{\geq} h\left(h_{11}H_{11}X_1 + Z_1, h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) \\
& \quad - h\left(h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) \\
& \quad + h(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3) \\
& \quad - h(h_{13}G_{13}X_1 + Z_3) - N_1 \log(2\pi e) - \beta_1 - \gamma_{11} \\
& \quad - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \\
& \overset{(b)}{=} \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + h\left(h_{12}H_{12}X_1 + Z_2\right. \\
& \quad \left. | h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3\right) \\
& \quad - h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right) \\
& \quad + h(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3) \\
& \quad - h(h_{13}G_{13}X_1 + Z_3) - \beta_1 - \gamma_{11} - N_2 \log(2\pi e) \\
& \quad - \beta_2 - \gamma_{12} \\
& \overset{(c)}{\geq} \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + h\left(h_{12}H_{12}X_1 + Z_2\right. \\
& \quad \left. | h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3, X_1\right) \\
& \quad - h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right) \\
& \quad + h(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3) \\
& \quad - h(h_{13}G_{13}X_1 + Z_3) - \beta_1 - \gamma_{11} - N_2 \log(2\pi e) \\
& \quad - \beta_2 - \gamma_{12} \\
& \overset{(d)}{\geq} \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) - h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right) \\
& \quad + h(h_{12}H_{12}X_1 + Z_2 | h_{13}G_{13}X_1 + Z_3) - \beta_1 - \gamma_{11} \\
& \quad - \beta_2 - \gamma_{12} \\
& = \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) - \beta_1 - \gamma_{11} - \beta_2 - \gamma_{12}
\end{aligned}
$$

The steps (a)-(b) hold true for the following reasons: (a) by the lower bounds (B.77) and (B.82); (b) chain rule of joint entropy; (c) conditioning reduces entropy; (d) Lemma B.5. The gap between $F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ has been bounded by $\beta_3 + \delta_3$ in Section B.4.1. Thus the gap between $F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ is quantified as $\beta_1 + \gamma_{11} + \beta_2 + \gamma_{12} + \beta_3 + \delta_3$.

B.4.3.4 The Gap between $F_1(\mathtt{M}_{1p})+F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})+F_3(\mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p})+\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$

$$
\begin{aligned}
& F_1(\mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{13}, \mathtt{M}_3) \\
\overset{(a)}{\geq}\ & h\left(h_{11}H_{11}X_1 + Z_1, h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) \\
& - h\left(h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) \\
& - \beta_1 - \gamma_{11} + h(h_{12}H_{12}X_1 + Z_2) \\
& - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \\
& + h(h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3, h_{13}G_{13}X_1 + Z_3^{'}) \\
& - h(h_{13}G_{13}X_1 + Z_3^{'}) - N_3 \log(2\pi e) - \beta_3 - \gamma_3 \\
\overset{(b)}{\geq}\ & h\left(h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3\right) \\
& - h\left(h_{13}H_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) \\
& + h(h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3) - N_3 \log(2\pi e) \\
& + h\left(h_{12}H_{12}X_1 + Z_2\right. \\
& \left.|h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3, X_1\right) \\
& - h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right) - \beta_1 - \gamma_{11} \\
& + h(h_{12}H_{12}X_1 + Z_2) - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \\
& + h(h_{13}G_{13}X_1 + Z_3^{'} | h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3) \\
& - h(h_{13}G_{13}X_1 + Z_3^{'}) - \beta_3 - \gamma_3 \\
=\ & \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)
\end{aligned}
$$

$$
\begin{aligned}
&- h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}H_{13}X_1 + Z_3\right) \\
&+ h(h_{12}H_{12}X_1 + Z_2) \\
&+ h(h_{13}G_{13}X_1 + Z_3^{'} | h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3) \\
&- h(h_{13}G_{13}X_1 + Z_3^{'}) - \beta_1 - \gamma_{11} - \beta_2 - \gamma_{12} - \beta_3 - \gamma_3 \\
\overset{(b)}{\geq}\ & \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \\
&- h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}G_{13}X_1 + Z_3\right) \\
&+ h(h_{12}H_{12}X_1 + Z_2) \\
&+ h(h_{13}G_{13}X_1 + Z_3^{'} | h_{13}H_{13}X_1 + h_{33}H_{33}X_3 + Z_3, X_1) \\
&- h(h_{13}G_{13}X_1 + Z_3^{'}) - \beta_1 - \gamma_{11} - \beta_2 - \gamma_{12} - \beta_3 - \gamma_3 \\
=\ & \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \\
&- h\left(h_{12}H_{12}X_1 + Z_2, h_{13}G_{13}X_1 + Z_3\right) \\
&+ h(h_{12}H_{12}X_1 + Z_2) + N_3 \log(2\pi e) - \beta_1 - \gamma_{11} - \beta_2 \\
&- \gamma_{12} - \beta_3 - \gamma_3 \\
=\ & \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) \\
&- h\left(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2\right) + N_3 \log(2\pi e) \\
&- \beta_1 - \gamma_{11} - \beta_2 - \gamma_{12} - \beta_3 - \gamma_3 \\
\overset{(c)}{\geq}\ & \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3) - \eta - \beta_1 - \gamma_{11} \\
&- \beta_2 - \gamma_{12} - \beta_3 - \gamma_3
\end{aligned}
$$

The steps (a)-(c) hold true for the following reasons: (a) by the lower bounds (B.77), (B.83) and (B.84); (b) chain rule of joint entropy and conditioning reduces entropy; (c) Lemma B.6.

B.4.3.5 The Gap between $F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_1(\mathtt{M}_1) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$

We first bound the gap between $F_1(\mathtt{M}_{1p})+F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p})+F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p})$ and $\bar{F}_1(\mathtt{M}_1)+\bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p})$ as follows.

$$
\begin{aligned}
& F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) \\
\stackrel{(a)}{=}\ & h\left(h_{11}H_{11}X_1 + Z_1, h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) \\
& \quad - h\left(h_{12}H_{12}X_1 + Z_2, h_{13}H_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) \\
& \quad - \beta_1 - \gamma_{11} + h\left(h_{11}H_{11}X_1 + Z_1, h_{13}G_{13}X_1 + Z_3\right) \\
& \quad - h\left(h_{13}G_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) - \beta_1 - \gamma_{11} \\
& \quad + h(h_{12}H_{12}X_1 + Z_2) - N_2 \log(2\pi e) - \beta_2 - \gamma_{12} \\
\stackrel{(b)}{\geq}\ & h(h_{11}H_{11}X_1 + Z_1) - N_1 \log(2\pi e) \\
& \quad + h\left(h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3\right) \\
& \quad - h\left(h_{13}H_{13}X_1 + Z_3\right) - N_1 \log(2\pi e) \\
& \quad + h\left(h_{12}H_{12}X_1 + Z_2\right. \\
& \quad \left.| h_{11}H_{11}X_1 + Z_1, h_{13}H_{13}X_1 + Z_3, X_1\right) \\
& \quad - h\left(h_{12}H_{12}X_1 + Z_2 | h_{13}G_{13}X_1 + Z_3\right) \\
& \quad + h\left(h_{13}G_{13}X_1 + Z_3 | h_{11}H_{11}X_1 + Z_1, X_1\right) \\
& \quad - h\left(h_{13}G_{13}X_1 + Z_3\right) + h(h_{12}H_{12}X_1 + Z_2) \\
& \quad - N_2 \log(2\pi e) - 2\beta_1 - 2\gamma_{11} - \beta_2 - \gamma_{12} \\
=\ & \bar{F}_1(\mathtt{M}_1) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) \\
& \quad - h\left(h_{13}G_{13}X_1 + Z_3 | h_{12}H_{12}X_1 + Z_2\right) \\
& \quad + N_3 \log(2\pi e) - 2\beta_1 - 2\gamma_{11} - \beta_2 - \gamma_{12} \\
\stackrel{(c)}{\geq}\ & \bar{F}_1(\mathtt{M}_1) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) - \eta - 2\beta_1 - 2\gamma_{11} - \beta_2 - \gamma_{12}
\end{aligned}
$$

The steps (a)-(c) hold true for the following reasons: (a) by the lower bounds (B.77), (B.79) and (B.83); (b) chain rule of joint entropy and conditioning reduces entropy; (c) Lemma B.6. The gap between $F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ has been bounded by $\beta_3 + \delta_3$ in Section B.4.1. Thus the gap between $F_1(\mathtt{M}_{1p}) + F_1(\mathtt{M}_{12}, \mathtt{M}_{13}, \mathtt{M}_{1p}) + F_2(\mathtt{M}_{123}, \mathtt{M}_{12}, \mathtt{M}_{2p}) + F_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ and $\bar{F}_1(\mathtt{M}_1) + \bar{F}_1(\mathtt{M}_{12}, \mathtt{M}_{1p}) + \bar{F}_3(\mathtt{M}_{123}, \mathtt{M}_{13}, \mathtt{M}_3)$ is quantified as $\eta + 2\beta_1 + 2\gamma_{11} + \beta_2 + \gamma_{12} + \beta_3 + \delta_3$.

At this point, we readily see that the gap $n^{(2)} = (\beta_1 + \beta_2 + \gamma_{11} + \gamma_{12} + \eta, \beta_1 + \beta_2 + \gamma_{21} + \gamma_{22}, \beta_3 + \gamma_3)$ settles the five sum rate gaps quantified in this subsection. Finally, for the gap between the $\mathcal{R}_{\text{in}}$ and $\mathcal{R}_{\text{o}}$, we pick each individual gap as the maximum of the individual gap in $n^{(1)}$ and $n^{(2)}$, i.e., $n_i = \max\{n_i^{(1)}, n_i^{(2)}\}$, and the desired gap $n = (\max\{n_1^{(1)}, n_1^{(2)}\}, \max\{n_1^{(1)}, n_1^{(2)}\}, \beta_3 + \delta_3)$ is justified. The proof is completed.

Appendix C

Proofs for the Results on MIMO MAC-IC-MAC

C.1 Details of the Proof of Theorem 5.1

The proof starts from the DM MAC-IC-MAC inner bound in [37, Theorem 1] which contains 9 classes of inequalities. In what follows, we refer the k-th class of inequalities as the k^{th}-inequality, for brevity. The first and second inequalities lead to the following two intra-cell sum rate inequalities regarding users in Ω_1 and Υ_1 for the DM MAC-IC-MAC (namely, the specifications of inequalities (14) and (15) of [37] in [37, Theorem 1]),

$$\sum_{1.j\in\Omega_1} R_{1.j} \leq \mathsf{B}_{1.j} \tag{C.1}$$

$$\sum_{1.j\in\Upsilon_1} R_{1.j} \leq \mathsf{A}_{\Upsilon_i} + \mathsf{E}_{\Upsilon_2}. \tag{C.2}$$

By Definition 5.2, it can be readily seen that for any Υ_1, there exists a corresponding Ω_1 such that $\Omega_1 = \Upsilon_1$, this observation implies the intra-cell sum rate $\sum_{1.j\in\Upsilon_1} R_{1.j}$ is bounded by both B_{Υ_1} and $\mathsf{A}_{\Upsilon_i} + \mathsf{E}_{\Upsilon_2}$, so we can replace the inequality (C.2) with

$$\sum_{1.j\in\Upsilon_1} R_{1.j} \leq \min\{\mathsf{B}_{\Upsilon_1}, \mathsf{A}_{\Upsilon_1} + \mathsf{E}_{\Upsilon_2}\} \tag{C.3}$$

without changing the inner bound. When $1.1 \notin \Omega_1$, inequality (C.2) takes no effect, $\sum_{1.j\in\Omega_1} R_{1.j}$ is upper bounded by (C.1) only. Hence we merge (C.1) and (C.6) as

$$\sum_{1.j\in\Omega_1} R_{1.j} \leq \mathsf{B}^{'}_{\Omega_1} \tag{C.4}$$

where

$$\mathsf{B}^{'}_{\Omega_i} \triangleq \begin{cases} \min\{\mathsf{B}_{\Omega_i}, \mathsf{A}_{\Omega_i} + \mathsf{E}_{\Upsilon_{i'}}\} & \text{if } i.1 \in \Omega_i \\ \\ \mathsf{B}_{\Omega_i} & \text{if } i.1 \notin \Omega_i \end{cases} \tag{C.5}$$

where $i, i^{'} \in \{1, 2\}$, $i \neq i^{'}$ and $(\Upsilon_1, \Omega_1, \Upsilon_2, \Omega_2) \in \Xi$.

By symmetry, we can coalesce the specifications of inequalities (16) and (17) of [37] in [37, Theorem 1] into the single inequality set

$$\sum_{2.j \in \Omega_2} R_{2.j} \leq \mathsf{B}^{'}_{\Omega_2} \tag{C.6}$$

At this point, we have the single region inner bound of [37, Theorem 1] to be effectively described by 7 sets of inequalities that are similar to the single region inner bound of Theorem 5.1 except that we have the quantities $\mathsf{A}_{\Upsilon_i}, \mathsf{B}_{\Omega_i}, \mathsf{E}_{\Upsilon_i}$, and G_{Ω_i} of [37] in place of $A_{\Upsilon_i}, B_{\Omega_i}, E_{\Upsilon_i}$, and G_{Ω_i} of Theorem 5.1, respectively.

Next we prove the set functions $\mathsf{A}_{\Upsilon_i}, \mathsf{B}_{\Omega_i}, \mathsf{E}_{\Upsilon_i}$, and G_{Ω_i} of [37, equations (31-34)] in [37, Definition 7], when specialized to the MIMO Gaussian case and for the coding scheme specified in the proof of Theorem 5.1 in the main text, are lower bounded by $A_{\Upsilon_i}, B_{\Omega_i}, E_{\Upsilon_i}$, and G_{Ω_i} of (5.16)-(5.19), respectively. We bound the set function A_{Υ_i} as an example.

$$\begin{aligned}
\mathsf{A}_{\Upsilon_i} &= I(X_{\Upsilon_i}; Y_i | X_{\bar{\Upsilon}_i}, X_{i.1c}, X_{i'.1c}, Q) \\
&= h(Y_i | X_{\bar{\Upsilon}_i}, X_{i.1c}, X_{i'.1c}) - h(S_{i'} | X_{i'.1c}) \\
&= h\Bigg(\sum_{i.j \in \Upsilon_i \setminus \{i.1\}} h_{i.j\to i} H_{i.j\to i} X_{i.j} \\
&\qquad + h_{i.1\to i} H_{i.1\to i} X_{i.1p} + h_{i'.1} H_{i'.1\to i} X_{i'.1p} + Z_i\Bigg) \\
&\qquad - h(h_{i'.1\to i} H_{i'.1\to i} X_{i'.1p} + Z_i) \\
&= \log \Bigg| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i} H^{\dagger}_{i.j\to i} \\
&\qquad + \frac{1}{M_{i.1}} \rho^{\alpha_{i.1\to i}} H_{i.1\to i} K_{i.1} H^{\dagger}_{i.1\to i} \\
&\qquad + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1\to i}} H_{i'.1\to i} K_{i'.1} H^{\dagger}_{i'.1\to i}\Bigg)
\end{aligned}$$

$$- \log \left| I_{N_i} + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} K_{i'.1} H^{\dagger}_{i'.1 \to i} \right|$$

$$\geq \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H^{\dagger}_{i.j \to i} \right.$$

$$+ \frac{1}{M_{i.1}} \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H^{\dagger}_{i.1 \to i}$$

$$\left. + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} K_{i'.1} H^{\dagger}_{i'.1 \to i} \right)$$

$$- \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}}$$

$$= A_{\Upsilon_1}$$

The rest of the bounds $\mathsf{B}_{\Omega_i} \geq B_{\Omega_i}$, $\mathsf{E}_{\Upsilon_i} \geq E_{\Upsilon_i}$ and $\mathsf{G}_{\Omega_i} \geq G_{\Omega_i}$ follow in a similar fashion. Hence the region of Theorem 5.1 is contained in the inner bound obtained from [37, Theorem 1] with the coding distribution specified as in the proof of Theorem 5.1 in the main text. The proof is hence completed.

C.2 Proof of Theorem 5.2

The starting point for the proof of Theorem 5.2 is Theorem 3 of [37] which gives an outer bound for the semi-deterministic DM MAC-IC-MAC. In particular, the latter bound is a union (over certain distributions) of polytopes of the form of $\mathcal{R}_{\mathrm{o}}$ given in the statement of Theorem 5.2 except that each of those polytopes in the union involves the set functions $\overline{\mathsf{A}}_{\Upsilon_i}, \overline{\mathsf{B}}_{\Omega_i}, \overline{\mathsf{E}}_{\Upsilon_i}$ and $\overline{\mathsf{G}}_{\Omega_i}$ defined in [37, Definition 11] (that depend on that distribution) in place of $\overline{A}_{\Upsilon_i}, \overline{B}_{\Omega_i}, \overline{E}_{\Upsilon_i}$ and $\overline{G}_{\Omega_i}$, respectively, in the definition of $\mathcal{R}_{\mathrm{o}}$ (which are explicitly computable). The idea here is to bound each of those set functions $\overline{\mathsf{A}}_{\Upsilon_i}, \overline{\mathsf{B}}_{\Omega_i}, \overline{\mathsf{E}}_{\Upsilon_i}$ and $\overline{\mathsf{G}}_{\Omega_i}$ universally (i.e., independently of the distribution involved) by $\overline{A}_{\Upsilon_i}, \overline{B}_{\Omega_i}, \overline{E}_{\Upsilon_i}$ and $\overline{G}_{\Omega_i}$, defined in (5.43)-(5.46), respectively. This will establish the explicit outer bound for the MIMO MAC-IC-MAC of Theorem 5.2.

We will first prove in detail that $\overline{\mathsf{A}}_{\Upsilon_i}$ can be upper bounded by $\overline{A}_{\Upsilon_i}$ of (5.43). The other three bounds follow in similar fashion. The fact that $\overline{\mathsf{A}}_{\Upsilon_i} \leq \overline{A}_{\Upsilon_i}$ is shown in the sequence of steps leading to (C.7) in the next page. The rationale for those steps is as follows: in the first inequality labeled (a), disabling time sharing will not shrink the outer bound because (i) reducing conditioning variable will not decrease the positive conditional entropy term of $\overline{\mathsf{A}}_{\Upsilon_i}$ and (ii) for its negative entropy term, we have $h(S_{i'}|X_{i'.1}, Q) = h(Z_{i'}|X_{i'.1}, Q) = h(Z_{i'})$, since the noise $Z_{i'}$ is independent of Q and $X_{i'.1}$. The second inequality labeled (b) holds since for random vectors X and Y with zero mean and joint covariance K, the conditional differential entropy of X given Y is maximized when X and Y are joint Gaussian [44, Lemma 1] We also assume the inputs have zero mean, i.e., $\mathrm{E}(X_{i.j}) = \mathbf{0}$, because non-zero means only contribute to power inefficiency. The equality (c) holds because of Woodbury's identity. Inequality (d) holds since (i) we have power constraint $\mathrm{Tr}(Q_{i.j}) \leq P_{i.j}$ for any $X_{i.j}$, and hence $Q_{i.j} \preceq P_{i.j} I_{M_{i.j}}$ (ii) using the matrix inequality of [27, Lemma 6] and (iii) $\log\det(\cdot)$ is increasing over the cone of positive definite matrices.

As stated previously, the upper bounds for $\overline{\mathsf{B}}_{\Omega_i}, \overline{\mathsf{E}}_{\Upsilon_i}$ and $\overline{\mathsf{G}}_{\Omega_i}$ defined in [37, Definition 11] can be similarly shown to be $\overline{B}_{\Omega_i}, \overline{E}_{\Upsilon_i}$ and $\overline{G}_{\Omega_i}$, respectively. The proof is completed.

C.3 Proof of Theorem 5.3

Following the proof outline, we prove the theorem in three steps. In the first step, we upper bound $\max\{\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}, \overline{E}_{\Upsilon_i} - E_{\Upsilon_i}\}$ for any $\Upsilon_i = \Upsilon_i^{'} \cup \{i.1\}$ and $\Upsilon_i^{'} \subseteq \Theta_i \backslash \{i.1\}$, and obtain an intra-cell sum rate gap β_{Υ_i} given by (C.10). In step two, we upper bound $\max\{\overline{B}_{\Omega_i} - B_{\Omega_i}^{'}, \overline{G}_{\Omega_i} - G_{\Omega_i}\}$ for any $\Omega_i \in 2^{\Theta_i} \backslash \emptyset$, and obtain an intra-cell sum rate gap $\max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}$ given by (C.17). In the last step, we show the desired individual rate gap $n_{i.j}$ given in the theorem satisfies that for any Υ_i,

$$\sum_{i.j \in \Upsilon_i} n_{i.j} \geq \beta_{\Upsilon_i} \tag{C.8}$$

and for any Ω_i,

$$\sum_{i.j \in \Omega_i} n_{i.j} \geq \max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}. \tag{C.9}$$

C.3.1 The Upper Bound on $\max\{\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}, \overline{E}_{\Upsilon_i} - E_{\Upsilon_i}\}$

We upper bound the difference $\overline{E}_{\Upsilon_i} - E_{\Upsilon_i}$ as an example. The upper bound on $\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}$ can be similarly derived.

$$\begin{aligned} E_{\Upsilon_i} = \log \Bigg| I_{N_i} &+ \sum_{i.j \in \Upsilon_i \backslash \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} \\ &+ \frac{1}{M_{i.1}} \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H_{i.1 \to i}^{\dagger} \\ &+ \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} H_{i'.1 \to i}^{\dagger} \Bigg| \end{aligned}$$

$$
\begin{aligned}
&\quad - \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
&\overset{(a)}{\geq} \log \Bigg| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} \\
&\quad + \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H_{i.1 \to i}^{\dagger} \\
&\quad + \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} H_{i'.1 \to i}^{\dagger} \Bigg| \\
&\quad - \min \left\{ \left(\sum_{i.j \in \Upsilon_i} M_{i.j} \right) + M_{i'.1}, N_i \right\} \\
&\quad \cdot \log \max \left\{ \max_{i.j \in \Upsilon_i} M_{i.j}, M_{i'.1} \right\} \\
&\quad - \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
&\triangleq \overline{E}_{\Upsilon_i} - \beta_{\Upsilon_i}
\end{aligned}
$$

The inequality (a) is true because the rank of the matrix $\sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger}$ $+ \frac{1}{M_{i.1}} \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H_{i.1 \to i}^{\dagger} + \frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} H_{i'.1 \to i}^{\dagger}$ cannot be greater than $\min\left\{ \left(\sum_{i.j \in \Upsilon_i} M_{i.j} \right) + M_{i'.1}, N_i \right\}$ and $\log\det(\cdot)$ is a monotonically increasing function over the cone of p.s.d. matrices. Similarly, we can show $\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}$ is upper bounded by β_{Υ_i} as well. Hence, we have

$$
\max\{\overline{A}_{\Upsilon_i} - A_{\Upsilon_i}, \overline{E}_{\Upsilon_i} - E_{\Upsilon_i}\} \leq \beta_{\Upsilon_i}. \tag{C.10}
$$

Thus we choose β_{Υ_i} as the intra-cell sum rate gap for any user subset Υ_i .

C.3.2 The Upper Bound on $\max\{\overline{B}_{\Omega_i} - B'_{\Omega_i}, \overline{G}_{\Omega_i} - G_{\Omega_i}\}$

Regarding the difference $\overline{B}_{\Omega_i} - B'_{\Omega_i}$, we shall have two different upper bounds depending on whether user $i.1$ belongs to Ω_i or not. If $i.1$ is not in Ω_i, then $B'_i = B_i$, and the gap from B_{Ω_i} to $\overline{B}_{\Omega_i}$ should be no bigger than β_{Ω_i}, which can be proved in the following,

$$
\begin{aligned}
B_{\Omega_i} &\geq \log \left| I_{N_i} + \sum_{i.j \in \Omega_i} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H_{i.j \to i}^{\dagger} \right| \\
&\quad - \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
&\geq \overline{B}_{\Omega_i} - \min \left\{ \left(\sum_{i.j \in \Omega_i} M_{i.j} \right), N_i \right\}
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \log \max \left\{ \max_{i.j \in \Omega_i} M_{i.j} \right\} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
\geq\ & \overline{B}_{\Omega_i} - \min \left\{ \left(\sum_{i.j \in \Omega_i} M_{i.j} \right) + M_{i'.1}, N_i \right\} \\
&\cdot \log \max \left\{ \max_{i.j \in \Omega_i} M_{i.j}, M_{i'.1} \right\} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
\triangleq\ & \overline{B}_{\Omega_i} - \beta_{\Omega_i}.
\end{aligned} \tag{C.11}
$$

The first inequality is true because $\log\det(\cdot)$ is a monotonically increasing function over the cone of p.s.d. matrices as well as the fact that the dropped term from the definition of B_{Ω_i} (c.f. (5.17)), i.e., $\frac{1}{M_{i'.1}} \rho^{\alpha_{i'.1 \to i}} H_{i'.1 \to i} K_{i'.1} H^{\dagger}_{i'.1 \to i}$, is p.s.d..

If $i.1 \in \Omega_i$, the intra-cell sum rate gap regarding $\overline{B}_{\Omega_i} - B'_{\Omega_i}$ should be an upper bound of the maximum value of $\overline{B}_{\Omega_i} - B_{\Omega_i}$ and $\overline{B}_{\Omega_i} - \min_{\Upsilon_{i'}} (A_{\Omega_i} + E_{\Upsilon_{i'}})$. The former difference has been bounded above. We compute the latter difference below. An intermediate lower bound on A_{Ω_i} is needed and obtained first. Let $J_{N_i} = I_{N_i} + \sum_{i.j \in \Omega_i \setminus \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H^{\dagger}_{i.j \to i}$, then we have

$$
\begin{aligned}
A_{\Omega_i} \geq\ & \Bigg| I_{N_i} + \sum_{i.j \in \Omega_i \setminus \{i.1\}} \frac{1}{M_{i.j}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H^{\dagger}_{i.j \to i} \\
&+ \frac{1}{M_{i.1}} \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H^{\dagger}_{i.1 \to i} \Bigg| \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
\geq \log & \Bigg| I_{N_i} + \sum_{i.j \in \Omega_i \setminus \{i.1\}} \rho^{\alpha_{i.j \to i}} H_{i.j \to i} H^{\dagger}_{i.j \to i} \\
&+ \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H^{\dagger}_{i.1 \to i} \Bigg| \\
&- \min \left\{ \sum_{i.j \in \Omega_i} M_{i.j}, N_i \right\} \log \max_{i.j \in \Omega_i} M_{i.j} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1 + M_{i'.1}}{M_{i'.1}} \\
= \log & \left| J_{N_i} + \rho^{\alpha_{i.1 \to i}} H_{i.1 \to i} K_{i.1} H^{\dagger}_{i.1 \to i} \right|
\end{aligned}
$$

$$
\begin{aligned}
&- \min\left\{\sum_{i.j\in\Omega_i} M_{i.j}, N_i\right\} \log \max_{i.j\in\Omega_i} M_{i.j} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \\
= &\log |J_{N_i}| + \log \left|I_{N_i} + \rho^{\alpha_{i.1\to i}} J_{N_i}^{-1} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger}\right| \\
&- \min\left\{\sum_{i.j\in\Omega_i} M_{i.j}, N_i\right\} \log \max_{i.j\in\Omega_i} M_{i.j} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}}. \qquad \text{(C.12)}
\end{aligned}
$$

The term $\log\left|I_{N_i} + \rho^{\alpha_{i.1\to i}} J_{N_i}^{-1} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger}\right|$ in (C.12) can be further lower bounded as

$$
\begin{aligned}
&\log \left|I_{N_i} + \rho^{\alpha_{i.1\to i}} J_{N_i}^{-1} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger}\right| \\
= &\log \left|I_{N_i} + \rho^{\alpha_{i.1\to i}} J_{N_i}^{-1} H_{i.1\to i} K_{i.1}^{\frac{1}{2}} K_{i.1}^{\frac{1}{2}} H_{i.1\to i}^{\dagger}\right| \\
\overset{(a)}{=} &\log \left|I_{M_{i.1}} + \rho^{\alpha_{i.1\to i}} K_{i.1}^{\frac{1}{2}} H_{i.1\to i}^{\dagger} J_{N_i}^{-1} H_{i.1\to i} K_{i.1}^{\frac{1}{2}}\right| \\
= &\log \left|K_{i.1}^{\frac{1}{2}} K_{i.1}^{-1} K_{i.1}^{\frac{1}{2}} \right. \\
&\quad \left. + \rho^{\alpha_{i.1\to i}} K_{i.1}^{\frac{1}{2}} H_{i.1\to i}^{\dagger} J_{N_i}^{-1} H_{i.1\to i} K_{i.1}^{\frac{1}{2}}\right| \\
= &\log \left|K_{i.1}^{-1} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i}^{\dagger} J_{N_i}^{-1} H_{i.1\to i}\right| - \log|K_{i.1}| \\
\geq &\log \left|I_{M_{i.1}} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i}^{\dagger} J_{N_i}^{-1} H_{i.1\to i}\right| - \log|K_{i.1}| \\
= &\log \left|I_{N_i} + \rho^{\alpha_{i.1\to i}} J_{N_i}^{-1} H_{i.1\to i} H_{i.1\to i}^{\dagger}\right| - \log|K_{i.1}|. \qquad \text{(C.13)}
\end{aligned}
$$

The equation (a) is true because of the fact that $\log\det(I_n + AB) = \log\det(I_m + BA)$ for any complex $n \times m$ matrix A and $m \times n$ matrix B. Substituting the lower bound (C.13) in (C.12), we end up with an intermediate lower bound on the set function A_{Ω_i} which is

$$
\begin{aligned}
A_{\Omega_i} \geq &\log \left|J_{N_i} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} H_{i.1\to i}^{\dagger}\right| - \log|K_{i.1}| \\
&- \min\left\{\sum_{i.j\in\Omega_i} M_{i.j}, N_i\right\} \log \max_{i.j\in\Omega_i} M_{i.j} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}}. \qquad \text{(C.14)}
\end{aligned}
$$

On the other hand, an intermediate lower bound on $E_{\Upsilon_{i'}}$ is obtained as follows,

$$
E_{\Upsilon_{i'}} \geq \log \left|I_{N_{i'}} + \frac{1}{M_{i.1}} \rho^{\alpha_{i.1\to i'}} H_{i.1\to i'} H_{i.1\to i'}^{\dagger}\right|
$$

$$
\begin{aligned}
&- \min\{M_{i.1}, N_{i'}\} \log \frac{1+M_{i.1}}{M_{i.1}} \\
&\geq \log |K_{i.1}| - \min\{M_{i.1}, N_{i'}\} \log M_{i.1} \\
&- \min\{M_{i.1}, N_{i'}\} \log \frac{1+M_{i.1}}{M_{i.1}} \\
&= \log |K_{i.1}| - \min\{M_{i.1}, N_{i'}\} \log(1+M_{i.1}).
\end{aligned} \tag{C.15}
$$

Note this lower bound of $E_{\Upsilon_{i'}}$ does not depend on the specific choice of $\Upsilon_{i'}$. Adding (C.14) and (C.15), we have

$$
\begin{aligned}
A_{\Omega_i} + E_{\Upsilon_{i'}} \geq &\log \left| J_{N_i} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} H_{i.1\to i}^{\dagger} \right| \\
&- \min\left\{ \sum_{i.j \in \Omega_i} M_{i.j}, N_i \right\} \log \max_{i.j\in\Omega_i} M_{i.j} \\
&- \min\{M_{i'.1}, N_i\} \log \frac{1+M_{i'.1}}{M_{i'.1}} \\
&- \min\{M_{i.1}, N_{i'}\} \log(1+M_{i.1}) \\
\triangleq &\overline{B}_{\Omega_i} - \gamma_{\Omega_i}.
\end{aligned} \tag{C.16}
$$

Hence, the difference $\overline{B}_{\Omega_i} - (A_{\Omega_i} + E_{\Upsilon_{i'}})$ is bounded. Combine the intra-cell sum rate gap for $i.1 \notin \Omega$ by (C.11) and for $i.1 \in \Omega$ by (C.16), we conclude for any user set Ω_i, the difference $\overline{B}_{\Omega_i} - B'_{\Omega_i}$ should be no more than $\max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}$ bits.

It can be verified with similar rationale that $\overline{G}_{\Omega_i} - G_{\Omega_i} \leq \beta_{\Omega_i}$. Hence, for any Ω_i, the term $\max\{\overline{B}_{\Omega_i} - B'_{\Omega_i}, \overline{G}_{\Omega_i} - G_{\Omega_i}\}$ is upper bounded by

$$
\max\{\overline{B}_{\Omega_i} - B'_{\Omega_i}, \overline{G}_{\Omega_i} - G_{\Omega_i}\} \leq \max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}. \tag{C.17}
$$

Thus we choose $\max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}$ as the intra-cell sum rate gap for any user subset Ω_i.

C.3.3 The Constant Gap

Now that we have intra-cell sum rate gap β_{Υ_i} for Υ_i and $\max\{\beta_{\Omega_i}, \gamma_{\Omega_i}\}$ for Ω_i. Returning to the discussion in the beginning of the proof, we are left to show the individual rate gap $n_{i.j}$ ensures (C.8) and

(C.9) for any $\varUpsilon_i$ and $\varOmega_i$. The former relationship can be seen in the following,

$$\begin{aligned}
\sum_{i.j\in\varUpsilon_i} n_{i.j} \geq \sum_{i.j\in\varUpsilon_i} \beta_{i.j} & \\
&= \sum_{i.j\in\varUpsilon_i} \min\left\{M_{i.j}+M_{i'.1},N_i\right\} \\
&\qquad \cdot \log\max\left\{\max_{i.j\in\varTheta_i} M_{i.j}, M_{i'.1}\right\} \\
&\qquad + |\varUpsilon_i| \min\{M_{i'.1},N_i\}\log\frac{M_{i'.1}}{1+M_{i'.1}} \\
&\geq \min\left\{\left(\sum_{i.j\in\varUpsilon_{i.j}} M_{i.j}\right)+M_{i'.1},N_i\right\} \\
&\qquad \cdot \log\max\left\{\max_{i.j\in\varUpsilon_i} M_{i.j}, M_{i'.1}\right\} \\
&\qquad + \min\{M_{i'.1},N_i\}\log\frac{M_{i'.1}}{1+M_{i'.1}} \\
&= \beta_{\varUpsilon_i}.
\end{aligned}$$

Similarly, it can verified that $\sum_{i.j\in\varOmega_i} n_{i.j} \geq \beta_{\varOmega_i}$ and $\sum_{i.j\in\varOmega_i} n_{i.j} \geq \gamma_{\varOmega_i}$ for any $\varOmega_i$. Hence, we have $\sum_{i.j\in\varOmega_i} n_{i.j} \geq \max\{\beta_{\varOmega_i},\gamma_{\varOmega_i}\}$ too, which completes the proof.

C.4 Proof of Lemma 5.1

We prove the Lemma by mathematical induction. Let us start from the case when $n=1$ and $H_1 = U\Sigma_1 V^\dagger$. We have

$$\begin{aligned}
&\log\det\left(I_u+\rho^{a_1}H_1H_1^\dagger\right) \\
&=\log\det\left(I_u+\rho^{a_1}\Sigma_1\Sigma_1^\dagger\right) \\
&=\log\det\begin{pmatrix} (1+\rho^{\alpha_1})I_{\min\{u.u_1\}} & \\ & I_{(u-u_1)^+}\end{pmatrix}+\mathcal{O}(1) \\
&=\alpha_1^+\min\{u_1,u\}\log(\rho)+\mathcal{O}(1) \\
&=g(u,(a_1,u_1))\log(\rho)+\mathcal{O}(1).
\end{aligned}$$

For $n=2$ and $n=3$, the results have been proved in [38] and [26]. Suppose the conclusion holds when $n=k$, in what follows we demonstrate the result for $n=k+1$. Without loss of generality, we assume

$a_1 \geq \max\{a_2, \cdots, a_{k+1}\}$, then

$$\begin{aligned}
&\log\left|I_u + \rho^{a_1} H_1 H_1^\dagger + \sum_{i=2}^{k+1} \rho^{a_i} H_i H_i^\dagger\right| \\
=& \log\left|I_u + \rho^{a_1} H_1 H_1^\dagger\right| \\
&+ \log\left|I_u + (I_u + \rho^{a_1} H_1 H_1^\dagger)^{-1} \sum_{i=2}^{k+1} \rho^{a_i} H_i H_i^\dagger\right|. \qquad \text{(C.18)}
\end{aligned}$$

Let the matrices Λ and $H_2^{(k+1)}$ be defined as

$$\Lambda = \begin{pmatrix} \rho^{a_2} I_{u_2} & & \\ & \ddots & \\ & & \rho^{\alpha_{k+1}} I_{u_k} \end{pmatrix}$$

and

$$H_2^{(k+1)} = (H_2, \cdots, H_{k+1}).$$

Note that the way $H_2^{(k+1)}$ is denoted temporarily violates our notation rule where $A^{(k)}$ denotes the k-th row of the matrix A, but no confusion will arise within the proof. Applying the identity that $\log\det(I_n + AB) = \log\det(I_m + BA)$ for any complex $n \times m$ matrix A and $m \times n$ matrix B, and substituting $H_1 H_1^\dagger$ with its SVD form $U\Sigma_1 V^\dagger$, the second term in the right hand side of (C.18) can be written as

$$\begin{aligned}
&\log\left|I_u + (I_u + \rho^{a_1} H_1 H_1^\dagger)^{-1} \sum_{i=2}^{k+1} \rho^{a_i} H_i H_i^\dagger\right| \\
=& \log\left|I_u + (I_u + \rho^{a_1} H_1 H_1^\dagger)^{-1} H_2^{(k+1)} \Lambda H_2^{(k+1)\dagger}\right| \\
=& \log\left|I_{\sum_{i=2}^{k+1} u_i} + \Lambda H_2^{(k+1)\dagger} (I_u + \rho^{a_1} H_1 H_1^\dagger)^{-1} H_2^{(k+1)}\right| \\
=& \log\left|I_{\sum_{i=2}^{k+1} u_i} + \Lambda H_2^{(k+1)\dagger} U\right. \\
&\cdot \begin{pmatrix} (I_{\min\{u,u_1\}} + \rho^{a_1} \Sigma_1 \Sigma_1^\dagger)^{-1} & \mathbf{0} \\ \mathbf{0} & I_{(u-u_1)^+} \end{pmatrix} \\
&\left.\cdot U^\dagger H_2^{(k+1)}\right|. \qquad \text{(C.19)}
\end{aligned}$$

We divide $H_2^{(k+1)}$ into two sub-matrices

$$G_1 = \left(U^\dagger H_2^{(k+1)}\right)^{(1:\min\{u,u_1\})}$$

and

$$G_2 = \left(U^\dagger H_2^{(k+1)}\right)^{(\min\{u,u_1\}+1:u)}$$

by extracting the first $\min\{u, u_1\}$ and the rest $(u-u_1)^+$ rows of $H_2^{(k+1)}$ respectively. Because the entries of $H_2^{(k+1)}$ are drawn i.i.d. from a continuous and unitarily invariant distribution, the product $U^\dagger H_2^{(k+1)}$ is identically distributed as $H_2^{(k+1)}$. This implies that the entries of the product $U^\dagger H_2^{(k+1)}$ are also drawn i.i.d. from a continuous and unitarily invariant distribution, so are the entries of matrices G_1 and G_2. Hence, both G_1 and G_2 are full rank w.p.1 and have the same property as the channel matrices. Continue from (C.19), we have

$$\begin{aligned}
&\log\left|I_u + (I_u + \rho^{a_1} H_1 H_1^\dagger)^{-1} \sum_{i=2}^{k+1} \rho^{a_i} H_i H_i^\dagger\right| \\
&= \log\left|I_{\sum_{i=2}^{k+1} u_i} + \Lambda G_1^\dagger (I_{\min\{u,u_1\}} + \rho^{a_1}\Sigma_1\Sigma_1^\dagger)^{-1} G_1\right. \\
&\qquad \left. + \Lambda G_2^\dagger G_2\right| \\
&\overset{(a)}{=} \log\left|I_{\sum_{i=2}^{k+1} u_i} + \Lambda G_2^\dagger G_2\right| + \mathcal{O}(1) \\
&= \log\left|I_{(u-u_1)^+} + G_2 \Lambda G_2^\dagger\right| + \mathcal{O}(1).
\end{aligned} \tag{C.20}$$

The equation (a) is true because we assumed $a_1 \geq \max\{a_2, \cdots, a_{k+1}\}$, and the matrix $\Lambda G_1^\dagger (I_{\min\{u,u_1\}} + \rho^{a_1}\Sigma_1\Sigma_1^\dagger)^{-1} G_1$ tends to be constant when $\rho \to \infty$. Plugging (C.20) in (C.18), we have for $n = k+1$,

$$\begin{aligned}
&\log\left|I_u + \rho^{a_1} H_1 H_1^\dagger + \sum_{i=2}^{k+1} \rho^{a_i} H_i H_i^\dagger\right| \\
&= \log\left|I_u + \rho^{a_1} H_1 H_1^\dagger\right| + \log\left|I_{(u-u_1)^+} + G_2 \Lambda G_2^\dagger\right| + \mathcal{O}(1) \\
&= f\left(u, (a_1, u_1)\right) \log(\rho) + f\left((u-u_1)^+, (a_2, u_2),\right. \\
&\qquad \left. \cdots, (a_{k+1}, u_{k+1})\right) \log(\rho) + \mathcal{O}(1) \\
&= f\left(u, (a_1, u_1), \cdots, (a_{k+1}, u_{k+1})\right) \log(\rho) + \mathcal{O}(1)
\end{aligned}$$

which completes the proof.

C.5 Proof of Theorem 5.4

We derive the GDoF region by taking the limiting value on both the left hand side and right hand side of the inequalities in the outer bound (c.f. Theorem 5.2). It is obvious that both b_{Ω_i} and g_{Ω_i} can be derived by directly using Lemma 5.1 on the outer bound set functions $\overline{B}_{\Omega_i}$ and $\overline{G}_{\Omega_i}$ (c.f. (5.44) and (5.46)), as the $\log\det(\cdot)$ terms in these two set functions are in the form $\log|I+\sum_i \rho^{\alpha_i} H_i H_i^\dagger|$. The computation of the asymptotic value of the set functions $\overline{A}_{\Upsilon_i}$ and $\overline{E}_{\Upsilon_i}$ in the form of $\log|I+\sum_i \rho^{\alpha_i} H_i H_i^\dagger + HK^{-1}H^\dagger|$ needs the SVD of the matrix K. In what follows, we compute the set function a_{Υ_i} as an example and e_{Υ_i} can be obtained similarly. Steps (a) and (b) hold true because (a) in this step, we let $G_1 = (H_{i.1\to i} V_{i.1\to i'})^{[1:\min\{M_{i.1}, N_{i'}\}]}$ and $G_2 = (H_{i.1\to i} V_{i.1\to i'})^{[\min\{M_{i.1}, N_{i'}\}+1:M_{i.1}]}$; (b) since $H_{i.1\to i} V_{i.1\to i'}$ is identically distributed as $H_{i.1\to i}$, the entries of the production $H_{i.1\to i} V_{i.1\to i'}$ are drawn i.i.d. from a continuous and unitarily invariant distribution, so are the entries of matrices G_1 and G_2 . Hence, G_1 and G_2 are full rank w.p.1 and have the same properties as the channel matrices, and thus Lemma 5.1 applies.

$$
\begin{aligned}
\overline{\mathsf{A}}_{\Upsilon_i} &= h(Y_i|X_{\bar{\Upsilon}_i}, T_i, X_{i'.1}, Q) - h(S_{i'}|X_{i'.1}, Q) \\
&\overset{(a)}{\leq} h\left(\sum_{i.j\in\Upsilon_i\backslash\{i.1\}} h_{i.j\to i}H_{i.j\to i}X_{i.j} + h_{i.1\to i}H_{i.1\to i}X_{i.1} + Z_i \,\middle|\, h_{i.1\to i'}H_{i.1\to i'}X_{i.1} + Z'_{i'}\right) - h(Z_i) \\
&\overset{(b)}{\leq} \log\left| I_{N_i} + \sum_{i.j\in\Upsilon_i\backslash\{i.1\}} |h_{i.j\to i}|^2 H_{i.j\to i}Q_{i.j}H^{\dagger}_{i.j\to i} + |h_{i.1\to i}|^2 H_{i.1\to i}Q_{i.1}H^{\dagger}_{i.1\to i} \right. \\
&\qquad \left. - |h_{i.1\to i}|^2|h_{i.1\to i'}|^2 H_{i.1\to i}Q_{i.1}H^{\dagger}_{i.1\to i'}\left(I_{N_{i'}} + |h_{i.1\to i'}|^2 H_{i.1\to i'}Q_{i.1}H^{\dagger}_{i.1\to i'}\right)^{-1} H_{i.1\to i'}Q_{i.1}H^{\dagger}_{i.1\to i}\right| \\
&= \log\left| I_{N_i} + \sum_{i.j\in\Upsilon_i\backslash\{i.1\}} |h_{i.j\to i}|^2 H_{i.j\to i}Q_{i.j}H^{\dagger}_{i.j\to i} + |h_{i.1\to i}|^2 H_{i.1\to i}Q^{\frac{1}{2}}_{i.1} \right. \\
&\qquad \left. \cdot\left[I_{M_{i.1}} - |h_{i.1\to i'}|^2 Q^{\frac{1}{2}}_{i.1}H^{\dagger}_{i.1\to i'}\left(I_{N_{i'}} + |h_{i.1\to i'}|^2 H_{i.1\to i'}Q_{i.1}H^{\dagger}_{i.1\to i'}\right)^{-1} H_{i.1\to i'}Q^{\frac{1}{2}}_{i.1}\right] Q^{\frac{1}{2}}_{i.1}H^{\dagger}_{i.1\to i}\right| \\
&\overset{(c)}{=} \log\left| I_{N_i} + \sum_{i.j\in\Upsilon_i\backslash\{i.1\}} |h_{i.j\to i}|^2 H_{i.j\to i}Q_{i.j}H^{\dagger}_{i.j\to i} + |h_{i.1\to i}|^2 H_{i.1\to i}Q^{\frac{1}{2}}_{i.1} \right. \\
&\qquad \left. \cdot\left(I_{M_{i.1}} + |h_{i.1\to i'}|^2 Q^{\frac{1}{2}}_{i.1}H^{\dagger}_{i.1\to i'}H_{i.1\to i'}Q^{\frac{1}{2}}_{i.1}\right)^{-1} Q^{\frac{1}{2}}_{i.1}H^{\dagger}_{i.1\to i}\right| \\
&\overset{(d)}{\leq} \log\left| I_{N_i} + \sum_{i.j\in\Upsilon_i\backslash\{i.1\}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i}H^{\dagger}_{i.j\to i} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i}\left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}}H^{\dagger}_{i.1\to i'}H_{i.1\to i'}\right)^{-1} H^{\dagger}_{i.1\to i}\right| \\
&= \log\left| I_{N_i} + \sum_{i.j\in\Upsilon_i\backslash\{i.1\}} \rho^{\alpha_{i.j\to i}} H_{i.j\to i}H^{\dagger}_{i.j\to i} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i}K_{i.1}H^{\dagger}_{i.1\to i}\right| \\
&= \overline{A}_{\Upsilon_i}
\end{aligned}
\tag{C.7}
$$

$$
\begin{aligned}
A_{\Upsilon_i} &= \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho_{i.j\to i} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} K_{i.1} H_{i.1\to i}^{\dagger} \right| \\
&= \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho_{i.j\to i} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} \left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}} H_{i.1\to i'}^{\dagger} H_{i.1\to i'} \right)^{-1} H_{i.1\to i}^{\dagger} \right| \\
&= \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho_{i.j\to i} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} V_{i.1\to i'} \left(I_{M_{i.1}} + \rho^{\alpha_{i.1\to i'}} \Sigma_{i.1\to i'}^{\dagger} \Sigma_{i.1\to i'} \right)^{-1} \right. \\
&\quad \left. V_{i.1\to i'}^{\dagger} H_{i.1\to i}^{\dagger} \right| \\
&= \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho_{i.j\to i} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i}} H_{i.1\to i} V_{i.1\to i'} \right. \\
&\quad \left. \cdot \begin{pmatrix} \left(I_{\min\{M_{i.1}, N_{i'}\}} + \rho^{\alpha_{i.1\to i'}} [\Sigma_{i.1\to i'}^{\dagger} \Sigma_{i.1\to i'}]^{+} \right)^{-1} & \\ & I_{(M_{i.1} - N_{i'})^{+}} \end{pmatrix} V_{i.1\to i'}^{\dagger} H_{i.1\to i}^{\dagger} \right| \\
&\overset{(a)}{=} \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho^{\alpha_{i,j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i}} G_1 \left(I_{\min\{M_{i.1}, N_{i'}\}} + \rho^{\alpha_{i.1\to i'}} [\Sigma_{i.1\to i'}^{\dagger} \Sigma_{i.1\to i'}]^{+} \right)^{-1} \right. \\
&\quad \left. \cdot G_1^{\dagger} + \rho^{\alpha_{i.1\to i}} G_2 G_2^{\dagger} \right| \\
&= \log \left| I_{N_i} + \sum_{i.j \in \Upsilon_i \setminus \{i.1\}} \rho^{\alpha_{i,j\to i}} H_{i.j\to i} H_{i.j\to i}^{\dagger} + \rho^{\alpha_{i.1\to i} - \alpha_{i.1\to i'}} G_1 G_1^{\dagger} + \rho^{\alpha_{i.1\to i}} G_2 G_2^{\dagger} \right| + o(1) \\
&\overset{(b)}{=} f \left(N_i, \bigcup_{i.j \in \Upsilon_i \setminus \{i.1\}} (\alpha_{i.j\to i}, M_{i.j}), \left((\alpha_{i.1\to i} - \alpha_{i.1\to i'})^{+}, \min\{M_{i.1}, N_{i'}\} \right), \left(\alpha_{i.1\to i}, (M_{i.1} - N_{i'})^{+} \right) \right) \\
&\quad \cdot \log(\rho) + \mathcal{O}(1) \qquad\qquad \text{(C.21)}
\end{aligned}
$$