

Rate-Splitting Multiple Access for Coordinated Multi-Point Joint Transmission with Partial CSIT

Author:

Liang Ma

Supervisor:

Dr. Bruno Clerckx

CID:

01117551

Second Marker:

Dr. Deniz Gündüz

A Thesis submitted in fulfilment of requirements for the degree of
Master of Science
Communications and Signal Processing
of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College London
5th September 2019

Abstract

The aim of this work is to implement Rate-Splitting Multiple Access (RSMA) in the Coordinated Multi-Point Joint Transmission (CoMP-JT) networks with partial CSIT, solve the sum-rate maximization problem and evaluate the performance. RSMA has recently been proven to be a promising downlink multiple access strategy due to its high spectral efficiency and energy efficiency. The key feature of Rate Spitting (RS) is that the message intended for each user is split into a common part and a private part, all the common parts are combined together as one common message and is transmitted alongside with all private messages. At the receiver, the common message is decoded by all users and the private message is only decoded by its corresponding user. By utilizing the idea of rate spitting with linear precoding at the transmitter and Successive Interference Cancellation (SIC) at the receiver, RSMA is able to partially treat interference as noise and partially decode interference, which contributes to performance enhancement.

In this work, we further investigate the use of RSMA in CoMP-JT with imperfect Channel State Information at the Transmitter (CSIT), which hasn't been considered in the literature. By designing optimal precoders using Sample Average Approximation (SAA) together with the Weighted Minimum Mean Square Error (WMMSE) method, the sum-rate maximization problem can be solved using Alternating Optimization (AO) algorithm. We show that compared with SDMA, RSMA is capable of delivering higher rate and Degrees of Freedom (DoF), more robustness and Quality of Service (QoS) enhancement under a wide range of CSIT quality, inter-cell channel strength disparity and inter-user channel strength disparity conditions.

Abstract

Acknowledgements

I would like to express my deep gratitude to Dr. Bruno Clerckx, my project supervisor, for his patient guidance and constructive advice. I also want to thank PhD candidate Longfei and Research Associate Lina, for their support and productive discussions with me. Furthermore, I would like to express my sincere appreciation to my friends and family, especially to my girlfriend Sheryl, for always being there for me through my ups and downs.

Acknowledgements

Contents

Abstract	3
Acknowledgements	5
Contents	8
List of Figures	11
List of Notations	13
List of Abbreviations	16
1 Introduction	17
2 Background	21
2.1 NOMA	21
2.2 SDMA	23
2.3 An Overview of RSMA	23
2.4 DoF and CSIT	25
3 RSMA	27
3.1 System Model	27
3.2 Two-user Network	28
3.3 Three-user Network	30
3.4 WMMSE-AO Algorithm	33
3.5 1-layer RSMA	38
3.6 Results and Discussions	39
3.6.1 Two-User Network with Random Channel Realizations	39
3.6.2 Two-User Network with Specific Channel Realizations	41
3.6.3 Three-User Network with Random Channel Realizations	45

3.6.4	Three-User Network with Specific Channel Realizations	48
4	RSMA for CoMP-JT	53
4.1	System Model	53
4.2	SR Maximization and WMMSE-AO Algorithm	54
4.3	Results and Discussions	57
4.3.1	Two-Cell Network	57
4.3.2	Three-Cell Network	59
5	RSMA with Partial CSIT	61
5.1	System Model	61
5.2	ASR Maximization	61
5.3	WMMSE-AO with SAA algorithm	64
5.4	Results and Discussions	68
5.4.1	Two-User Network	68
5.4.2	Three-User Network	69
5.4.3	DoF Analysis	72
6	RSMA for CoMP-JT with Partial CSIT	77
6.1	System Model	77
6.2	ASR Maximization and WMMSE-AO with SAA Algorithm	78
6.3	Results and Discussions	82
6.3.1	Two-Cell Network	82
6.3.2	Three-Cell Network	85
7	Conclusions and Future Work	91
	Appendix	93
	References	95

List of Figures

3.1	Two-user model using RSMA [1].	28
3.2	Three-user model using RSMA [1].	30
3.3	Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT averaged over 100 random channel realizations, $N_t = 2, K = 2$	40
3.4	Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 4, K = 2, \lambda_2 = 1$, SNR = 20dB. .	42
3.5	Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 4, K = 2, \lambda_2 = 0.3$, SNR = 20dB. .	43
3.6	Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 2, K = 2, \lambda_2 = 1$, SNR = 20dB. .	44
3.7	SR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA averaged over 100 channel realizations with perfect CSIT, $N_t = 3, K = 3$	47
3.8	WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 4, K = 3, \mathbf{u} = [0.4, 0.3, 0.3]$	49
3.9	WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 4, K = 3, \mathbf{u} = [0.2, 0.3, 0.5]$	50
3.10	WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 2, K = 3, \mathbf{u} = [0.4, 0.3, 0.3], \mathbf{r}_{th} = [0.02, 0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bit/s/Hz.	51
4.1	Wyner model, $L = 3, K_l = 1, l \in \{1, 2, 3\}$	57
4.2	Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA for CoMP-JT with perfect CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}$, SNR = 20dB.	58

4.3 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with perfect CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}$, $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, 0.3]$ bits/s/Hz.	60
5.1 Achievable rate region of MU-LP based SDMA and RSMA with partial CSIT, $N_t = 2, K = 2, \sigma_e^2 = P_t^{-\alpha}$, SNR = 20dB.	68
5.2 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 3, K = 3, \sigma_e^2 = P_t^{-\alpha}, \mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz.	70
5.3 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 3, K = 3, \sigma_e^2 = P_t^{-\alpha}, \mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, 0.3]$ bit/s/Hz.	71
5.4 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 2, K = 2, \sigma_e^2 = P_t^{-\alpha}$. (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, \dots, 0.3]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.	73
5.5 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 2, K = 4, \sigma_e^2 = P_t^{-\alpha}$. (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, \dots, 0.3]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.	75
6.1 Achievable rate region of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-0.6}$, SNR = 20dB.	82
6.2 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-0.6}, a = 0.1, b = 0.1$. (a) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b) $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, \dots, 0.1]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.	83
6.3 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-\alpha}, a = 0.5, b = 0.5$. (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, \dots, 0.1]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.	84
6.4 SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.3}, \mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.	85

6.5	SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.6}$, $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.	86
6.6	SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.9}$, $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.	87
6.7	SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.3}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.	88
6.8	SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.6}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.	89
6.9	SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.9}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.	90
B.1	SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 5, N_t = 1, K_l = 1, l \in \{1, 2, 3, 4, 5\}, \sigma_e^2 = P_t^{-0.6}$, $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.	94

List of Figures

List of Notations

Bold Lowercase Letters	Vector
Bold Uppercase Letters	Matrix
\mathbb{E}	Statistical expectation
$(\cdot)^T$	Transpose
$(\cdot)^H$	Hermitian transpose
$\ \cdot\ $	Euclidean norm
$\text{tr}(\cdot)$	Trace of a matrix
$\text{diag}(\cdot)_n$	The n^{th} entry of the diagonal elements
\mathbf{I}	Identity matrix
$\mathbf{0}$	All-zero vector
\mathbb{C}	Complex space
$\mathcal{CN}(\mu, \sigma^2)$	Complex Gaussian distribution with mean μ and variance σ^2

List of Notations

List of Abbreviations

AO Alternating Optimization.

ASR Average Sum-Rate.

AWGN Additive White Gaussian Noise.

BC Broadcast Channel.

BS Base Station.

CoMP-JT Coordinated Multi-Point Joint Transmission.

CSIR Channel State Information at the Receiver.

CSIT Channel State Information at the Transmitter.

DL Downlink.

DoF Degrees of Freedom.

DPC Dirty Paper Coding.

ESR Ergodic Sum-Rate.

FDMA Frequency-Division Multiple Access.

i.i.d. Independent and Identically Distributed.

MIMO Multiple-Input Multiple-Output.

MISO Multiple-Input Single-Output.

MRT Maximum Ratio Transmission.

MU Multi-User.

MU-LP Multi-User Linear Precoding.

NOMA Non-Orthogonal Multiple Access.

OMA Orthogonal Multiple Access.

QCQP Quadratically Constrained Quadratic Program.

QoS Quality of Service.

RS Rate-Splitting.

RSMA Rate-Splitting Multiple Access.

SAA Sample Average Approximation.

SAF Sample Average Function.

SC Superposition Coding.

SCMA Sparse Code Multiple Access.

SDMA Space-Division Multiple Access.

SIC Successive Interference Cancellation.

SISO Single-Input Single-Output.

SR Sum-Rate.

SVD Singular Value Decomposition.

TDMA Time-Division Multiple Access.

WMMSE Weighted Minimum Mean Square Error.

WSR Weighted Sum-Rate.

ZFBF Zero-Forcing Beamforming.

Chapter 1

Introduction

The increasing demands of high-speed data transmission and large capacity of the mobile network have been driving the telecommunication industry to design a next generation of wireless networks, which should be able to provide high spectral efficiency and energy efficiency. By utilizing multiple antennas at the Base Station (BS) to serve multiple users, the spectral efficiency would be boosted dramatically if the Multi-User (MU) interference can be handled properly. In fact, the quality of Channel State Information at the Transmitter (CSIT) is essential to the MU interference management [2][3]. Besides MU interference, the inter-cell interference also needs to be addressed as there will be massive BSs in 5G and beyond. One of the best approaches to handle inter-cell interference is to deploy the Coordinated Multi-Point Joint Transmission (CoMP-JT) network [4][5]. Rate-Splitting Multiple Access (RSMA), a novel multiple access strategy has recently been proposed and has demonstrated outstanding Sum-Rate (SR) performance in Downlink (DL) systems [3].

RSMA embraces Non-Orthogonal Multiple Access (NOMA) and Space-Division Multiple Access (SDMA) as two extreme cases and outperforms those two. In NOMA, Superposition Coding (SC) is implemented at the transmitter and Successive Interference Cancellation (SIC) is used at the receiver (together termed SC-SIC); the interference is fully decoded in NOMA. SDMA implements Multi-User Linear Precoding (MU-LP) at the transmitter, and treats the interference as noise at the decoding stage. RS incorporates NOMA and SDMA, and works as follows: the message intended for each user is split into a common message and a private message, all the common messages are combined together as one common message and is transmitted alongside with all users' private messages after linear precoding.

At the user side, the received signal is decoded using SIC by first decoding the common message. The common message is decoded by all users and the private message is only decoded by its corresponding user. As we will see later, such rate splitting with linear precoding at transmitter and SIC at the receiver, is able to partially treat interference as noise and partially decode interference [1]. This unique feature will ultimately leads to performance enhancement.

A number of works have been done to investigate the effectiveness of Rate-Splitting (RS) in various transmission network configurations using different performance measurement matrices. To name a few, [1] analyzed thoroughly the benefits of RS by considering different network loads including underloaded and overloaded regimes¹. In [6], implementing RS is proven to be able to deliver higher SR for CoMP-JT networks. Using RS to tackle the CSIT inaccuracy has also been studied in a couple of works: [7] addressed the max-min fairness problem in DL MU Multiple-Input Single-Output (MISO) systems, [2] investigated the SR maximization by using RS in MISO networks and [8] examined the use of hierarchical RS (HRS) with partial CSIT in massive MIMO.

The aim of this work is to implement RS in the CoMP-JT networks with partial CSIT, solve the SR maximization problem and evaluate the performance of RS. Such topic is novel and has not been proposed in the literature (to the best of the author's knowledge). The 2-user rate region and 3-user SR will be used as the main measures of performance² throughout the report; in addition, Degrees of Freedom (DoF) analysis will also be conducted. In order to provide a comprehensive view on the structure as well as the benefits of RS in CoMP-JT networks with partial CSIT, the rest of the report is organized as follows:

- Chapter 2 Background: Review NOMA and SDMA, discuss their limitations and the benefits of RSMA over them. Discuss DoF under perfect and imperfect CSIT.

¹Overloaded (or underloaded) means that there are more (or less) users than the transmit antennas at each BS. Underloaded networks also contains the case where the number of users and the number of transmit antennas are the same.

²The 2-cell/user and 3-cell/user networks are typical in CoMP-JT, and the benefits gained from the cooperation between more cells is little [6]. Therefore, in order to have a consistent analysis throughout the report, 2-user and 3-user networks are considered.

- Chapter 3 RSMA: Introduce the mechanism of RSMA and the Weighted Minimum Mean Square Error (WMMSE) Alternating Optimization (AO) algorithm, which is implemented to find the optimal precoder that maximize the SR. A low-complexity version of RS (1-layer RS) is also introduced, and will be used in Chapter 4-6.
 - ▷ We demonstrate the SR improvement of RSMA over NOMA and SDMA through various simulations (including considering underloaded/overloaded networks, different channel strengths and channel angles, different Quality of Service (QoS) requirements).
- Chapter 4 RSMA for CoMP-JT: Introduce CoMP-JT and deploy RS in such networks. Examine the rate enhancement of RSMA over NOMA and SDMA.
 - ▷ We show that compared with NOMA and SDMA, RSMA is more robust to any inter-cell channel strength disparity and inter-user channel strength disparity.
- Chapter 5 RSMA with partial CSIT: Introduce WMMSE-AO with Sample Average Approximation (SAA) algorithm to tackle partial CSIT. Evaluate the SR and DoF improvement of RSMA over SDMA.
 - ▷ We show that under imperfect CSIT, RSMA outperforms SDMA in terms of both SR and DoF with or without QoS constraints.
- Chapter 6 RSMA for CoMP-JT with partial CSIT: Bring Chapter 3-5 together, propose the new framework which is based on RS to maximize the SR in CoMP-JT networks with partial CSIT.
 - ▷ We demonstrate that RSMA can provide SR and DoF improvement over SDMA (with or without QoS requirements), and is more robust to any inter-cell channel strength disparity and inter-user channel strength disparity.
- Chapter 7: Conclusions and Future Work.

Chapter 2

Background

2.1 NOMA

The Orthogonal Multiple Access (OMA) schemes share the wireless resources in the same domain by exploiting the orthogonality of user signals; various domains including time domain (Time-Division Multiple Access (TDMA)) or frequency domain (Frequency-Division Multiple Access (FDMA)). In these multiple access techniques, only one user is served in a given time-frequency resource block, which is not suitable for the next generations wireless networks due to the enormous number of devices that need to be served [9]. On the other hand, NOMA superposes different users in the power domain [10] or code domain (e.g. Sparse Code Multiple Access (SCMA) [11]) within the same time-frequency block. Power-domain NOMA¹ utilizes SC-SIC, and is motivated by the fact that the capacity region of the Single-Input Single-Output (SISO) Broadcast Channel (BC) can be achieved using SC-SIC [12].

It's important to note that when users experience different (or similar) channel strengths, SI-SIC performs better (or worse) than OMA strategies. Therefore, the advantage of single-antenna NOMA is the capability to deal with overloaded scenarios when the users experience (very) different channel strengths. The main drawback of NOMA is that the complexity increases with the number of users, which is caused by the nature of SIC. By using SIC, the users with strongest channel strength has to decode all other users' messages before its own message; the user with second strongest channel strength has to decode messages of users that have weaker channel strengths than it, and so on. When the number of users is large, such process

¹Power-domain NOMA is simply referred as NOMA in other parts of the report.

introduces high complexity as well as the high probability of error propagation [1]. One way to reduce complexity and error propagation is to group users, use SC-SIC only within the group, and use other techniques to serve each group².

Driven by the advantage of NOMA in SISO BC, a number of works (e.g. [13][14]) have used NOMA in MISO BC by implementing SC-SIC. The users are ordered based on their channel strengths, and (at least) one user has to decode all other users' messages. As a result, the multi-antenna non-degraded channel effectively becomes a single-antenna degraded channel when NOMA is used [1]. The consequence of this is that the DoF³ (also known as spatial multiplexing gain) now reduces to one⁴. There are also other drawbacks of NOMA. Firstly, NOMA is only suitable for the user deployment where different users are sufficiently aligned and experience different channel strengths; NOMA in a more general user deployment will lead to unsatisfactory performance. Secondly, the accuracy of CSIT is also a limiting factor when using NOMA, as the initial design requirements of NOMA assume that perfect CSIT is available. Thirdly, NOMA introduces complexity both at the receiver end and at the transmitter. Regarding the receivers, as mentioned earlier, the number of layers of SIC increases with the number of users. Besides this, at the transmitter, all the possible decoding orders need to be considered since a natural order doesn't exist (in SISO BC the user channels are scalars thus we have a natural order, while here in MISO BC the user channels are vectors). Therefore, the decoding order is jointly optimized with the precoders at the transmitter, which introduces an exponentially increasing complexity (with the number of users)⁵. The advantage of NOMA is the capability of handling an overloaded network where users are aligned with each other and experience a disparity of channel strengths [1].

²The 'SC-SIC per group' strategy (for MISO BC) is thoroughly discussed and compared with RS in [1].

³DoF can be viewed as the number of interference-free streams that can be sent.

⁴A more detailed analysis on DoF of SIC can be found in [15].

⁵More discussions on the decoding orders will be presented in Section 3.3.

2.2 SDMA

SDMA exploits the spatial domain by superposing multiple users in the same time-frequency resource block. When using SDMA, the MISO BC is non-degraded (in NOMA it's degraded). The degraded channel is the reason that although using NOMA can achieve the SISO BC capacity region, in MISO BC, Dirty Paper Coding (DPC) is the only strategy to achieve capacity region when perfect CSIT is available [16]. The main philosophy of DPC is that the interference is considered and suppressed at the transmitter (not at the receiver), which requires perfect CSIT and complex transmitter design (non-linear precoding) to achieve that. However, the computational complexity of DPC has forced people to use other strategies; in MISO BC, MU-LP is considered to be an effective alternative (although suboptimal) when users experience (semi-) orthogonal channels with similar strengths [17]. MU-LP based SDMA is therefore widely used in a significant number of schemes in 4G and 5G [1].

Using SDMA in MISO BC is beneficial from a DoF perspective when CSIT is perfect, and low complexity is also brought by MU-LP based SDMA. Again, the CSIT quality is crucial to the performance. The disadvantage of SDMA is that it can not handle an overloaded network properly due to the MU interference. The cancellation of MU interference relies on the design of orthogonal precoders at the transmitter, which only works well when there are more transmit antennas than users. Another drawback is that MU-LP based SDMA is not suitable for general user distribution as it works effectively only when the users experience orthogonal channels with similar channel strengths. Finally, using SDMA would yield poor performance (in terms of both SR and DoF⁶) if CSIT is not perfect. The CSIT quality is crucial to SDMA, as the design of it is motivated by the perfect CSIT [1].

2.3 An Overview of RSMA

Having investigated the benefits and limitations of NOMA and SDMA, we now move onto the analysis of RSMA. RSMA incorporates NOMA (fully decode interference) and SDMA (fully treat interference as noise), and is able to partially decode interference and partially treat interference as noise by implementing rate splitting with linear precoding at transmitter and SIC at the receiver. RSMA can reduce to SDMA

⁶More discussion on DoF of SDMA will be presented in Section 2.4.

and NOMA⁷ by implementing certain power allocations between the common and private messages.

The concept of RS (i.e. split the message into common parts and private parts) was firstly appeared in [18], one difference is that they implemented SC in a 2-user single antenna interference channel. The feature of partially decoding interference and partially treating interference as noise in [18] is essential, and has motivated a number of recent works. We have mentioned a few in Chapter 1, and will briefly review some of the other works here, which involve investigating the benefits of RS in various networks through different means of performance measurement including SR maximization, max-min fairness and DoF region.

When perfect CSIT is available, the benefits of RS has been investigated. For example, [1] examined the rate region and Weighted Sum-Rate (WSR) in underloaded and overloaded MISO BC; [15][19] investigated the max-min fairness in an overloaded co-channel multicast multigroup setup. In the case where only partial CSIT is available, assuming the CSIT error⁸ decreases as SNR increases, [2] studied sum DoF and SR of RS in underloaded MISO BC and [20] showed that RS is optimal for the DoF region in MISO BC. When the CSIT error is fixed (i.e. doesn't change with SNR), [21] shows that RS is beneficial in terms of max-min DoF. Besides underloaded MISO BC, when the CSIT error decays as SNR increases, DoF of RS in K -cell MISO interference channel [22] and in Multiple-Input Multiple-Output (MIMO) networks [23] have also been examined. Furthermore, the use of RS in millimeter wave systems [24] and in massive MIMO systems under certain hardware impairments [25] have also been investigated. We will review more works and further discuss the DoF of RS in the next section.

RSMA is able to tackle the limitations of NOMA and SDMA, including not suitable for a general network or a general user deployment, perfect CSIT is required, complexity in NOMA and the incapability of SDMA in the overloaded networks, etc. We will show the effectiveness of RS through extensive simulation results, demonstrate the advantages of RS such as being robust to the general network, SR gain with and without QoS requirements, low complexity and DoF benefits under partial CSIT.

⁷Both NOMA and SDMA are subsets of RSMA [1].

⁸More on CSIT error in Section 2.4 and Chapter 5.

2.4 DoF and CSIT

DoF⁹ can be viewed as the number of interference-free streams that can be sent. It is the pre-log factor of the rate at high SNR [1]:

$$\text{DoF} = \lim_{\text{SNR} \rightarrow \infty} \frac{R(\text{SNR})}{\log_2(\text{SNR})} \quad (2.1)$$

For example, if DoF is 2, then on the SR versus SNR plot, when SNR is increased by 3dB, the rate should increase by 2 bit/s/Hz at high SNR (which is the value of DoF).

Under perfect CSIT, using linear precoding techniques such as Zero-Forcing Beamforming (ZFBF)¹⁰ can achieve optimal DoF of MISO BC [27]. When only partial CSIT is available, full DoF can also be achieved as long as $\alpha \geq 1$, where α is the quality scaling exponent ($\alpha \geq 0$) [28]. The CSIT is quantified in a way that the CSIT error decays at a rate of $O(\text{SNR}^{-\alpha})$ as SNR increases. When $\alpha \rightarrow \infty$, this indicates CSIT is perfect; whereas $\alpha = 0$ means the CSIT error power is at a fixed level¹¹. The quality scaling exponent is chosen from the range $\alpha \in [0, 1]$, as in the DoF sense $\alpha = 1$ corresponds to the perfect CSIT [2][29].

As mentioned earlier, DPC is the only way to achieve capacity region in MISO systems under perfect CSIT. The capacity region under partial CSIT is still a open problem. The DoF region is used as an alternative measure to characterize the fundamental limits of MISO BC with partial CSIT [30]. It has been shown that RS is optimal to achieve DoF in underloaded MISO BC with partial CSIT [2][29]. Moreover, implementing RS under imperfect CSIT with a flexible power allocation is optimal in terms of not only DoF but also DoF region in underloaded MISO BC [20]. Regarding the overloaded networks, [31] considered two sets of CSIT qualities, and proved that by superimposing degraded symbols on top of the rate-split symbols, optimal DoF region can be achieved.

Under partial CSIT, using conventional multiple access strategies is not effective,

⁹In this report, DoF refers to sum-DoF.

¹⁰The ZFBF precoders are determined by the pseudo-inverse of the channel [26]. Therefore, although ZFBF is optimal from DoF point of view, it's not optimal regarding SR as the precoders are not optimized.

¹¹The quality scaling exponent α can also be interpreted as the number of feedback bits, i.e., $\alpha = 0$ means the number of feedback bits is fixed, $\alpha \neq 0$ means that the number of feedback bits is scaled with SNR [2][3].

since they are designed assuming perfect CSIT (as discussed in Section 2.1 NOMA and 2.2 SDMA). Therefore, we need to design new techniques which are initiated by the imperfect CSIT. From DoF perspective, partial knowledge of CSIT leads to distorted interference cancellation, leaving residual interference behind which harms DoF. The conventional multiple access schemes normally consider the residual interference as noise. However, decoding some of the interference is beneficial in terms of DoF [2][29][32].

Using RS under partial CSIT has been shown to achieve better performance than NoRS techniques (NoRS categories the transmission purely rely on private messages [2]). In [2], it has been proven that in underloaded MISO BC, DoF of NoRS scheme is $\max\{1, K\alpha\}$ and DoF of RS scheme is $1 + (K - 1)\alpha$, where K is the number of users. When CSIT is perfect ($\alpha = 1$), both NoRS and RS achieve full DoF K . For NoRS, it's intuitive that a fraction of DoF is achieved (i.e. $K\alpha$) under partial CSIT as the interference is not fully nulled and residual interference is present [3]. When α is smaller than $\frac{1}{K}$, MU transmission can no longer to be supported, and single-user transmission starts which leads to DoF of 1. Regarding RS, the private parts still achieve DoF of $K\alpha$, and the common parts contributes an extra DoF of $1 - \alpha$. Therefore, besides SR, we will also use DoF as a performance measure to analyze the performance of RS with partial CSIT.

Chapter 3

RSMA

In this chapter¹ I will first introduce the mechanism of RSMA using a 2-user example, and then elaborate the main WMMSE-AO algorithm using a 3-user configuration. Such algorithm is implemented in order to find the best precoder which maximizes the SR. I will also discuss 1-layer RS, which is a simplified version with low complexity. The numerical results and data analysis will be presented to conclude the chapter.

3.1 System Model

Let's consider a single-cell MISO configuration where a BS (with N_t transmit antennas) serves K users (each with $N_r = 1$ receive antenna). The interference from other nearby cells are ignored. The transmitted signal for a given user channel is $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$, which is subject to the total transmit power constraint: $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq P_t$. The channel between BS and user- k is given by $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}$, $k \in \mathcal{K} = \{1, 2, \dots, K\}$. The received signal at user- k can be expressed as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \quad (3.1)$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is Additive White Gaussian Noise (AWGN) at the receiver end (we assume every user experiences the same noise variance, which is 1, this also applies to other chapters; thus the SNR is the same as the total transmit power P_t). CSIT and Channel State Information at the Receiver (CSIR) are also assumed to be perfect.

¹The main reference of this chapter is [1] unless otherwise specified.

3.2 Two-user Network

We first have a look at the use RS in a 2-user MISO network. Figure 3.1 demonstrates the block diagram of such configuration. There are two messages intended to receive by user-1 and user-2 respectively: W_1 and W_2 . Each message is split into two parts: for user-1 they are $\{W_1^{12}, W_1^1\}$ and for user-2 they are $\{W_2^{12}, W_2^2\}$. The common message W_{12} consists W_1^{12} and W_2^{12} , and is encoded into common stream s_{12} using a common codebook which is available to both users. Regarding private messages, W_1^1 is encoded into private stream s_1 and W_2^2 is encoded into private stream s_2 . The corresponding linear precoders for common stream and private streams are \mathbf{p}_{12} , \mathbf{p}_1 and \mathbf{p}_2 respectively. The transmitted signal is then given by:

$$\mathbf{x} = \mathbf{Ps} = [\mathbf{p}_{12}, \mathbf{p}_1, \mathbf{p}_2][s_{12}, s_1, s_2]^T = \mathbf{p}_{12}s_{12} + \mathbf{p}_1s_1 + \mathbf{p}_2s_2 \quad (3.2)$$

The total transmit power constraint $\mathbb{E}\{||\mathbf{x}||^2\} \leq P_t$ can be further expressed as $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$ assuming $\text{tr}(\mathbf{s}\mathbf{s}^H) = \mathbf{I}$.

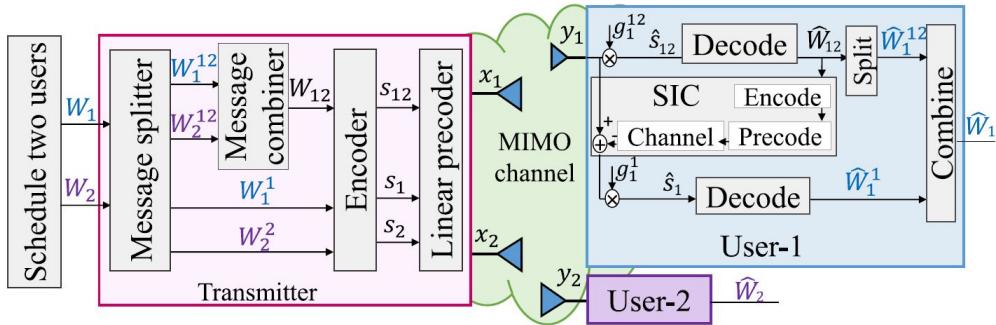


Figure 3.1: Two-user model using RSMA [1].

At the receiver end, both users will first decode the common stream s_{12} by considering private streams (s_1 and s_2) as noise. At user- k , the SINR of the common stream is:

$$\eta_k^{12} = \frac{|\mathbf{h}_k \mathbf{p}_{12}|^2}{|\mathbf{h}_k \mathbf{p}_1|^2 + |\mathbf{h}_k \mathbf{p}_2|^2 + 1} \quad (3.3)$$

Note here although named 'common stream', the term 'common' should not be confused with the common message in the multicast sense where the whole common message is intended to all users. In RS, for example, user-1 not only decodes it's corresponding common message (W_1^{12}) but also decodes user-2's common message (W_2^{12}), which is effectively part of the interference.

After the common stream is decoded, the contribution of it will be subtracted from

the received signal using SIC. Then user- k will decode its corresponding private stream by viewing the private stream of the other user as noise. At user- k , the SINR of the private stream is:

$$\eta_k = \frac{|\mathbf{h}_k \mathbf{p}_k|^2}{|\mathbf{h}_k \mathbf{p}_j|^2 + 1}, \quad j \neq k \quad (3.4)$$

The achievable rates for common stream s_{12} is $R_k^{12} = \log_2 (1 + \eta_k^{12})$ and that for the private stream s_k is $R_k = \log_2 (1 + \eta_k)$. In order for s_{12} to be decoded successfully by both users, the common achievable rate should not be greater than $R_{12} = \min\{R_1^{12}, R_2^{12}\}$. We assume that the common achievable rate is shared by user-1 and user-2 such that $R_{12} = Z_1^{12} + Z_2^{12}$ ($Z_k^{12}, k \in \{1, 2\}$, is the fraction of common rate R_{12} assigned to user- k). We can then write the total achievable rate of user- k as: $R_{total,k} = Z_k^{12} + R_k, k \in \{1, 2\}$. If we assign a weight vector $\mathbf{u} = [u_1, u_2]$ to both users, the WSR can be expressed as:

$$\begin{aligned} \text{WSR } (\mathbf{u}) &= \max_{\mathbf{P}, \mathbf{c}} \sum_{k=1}^2 u_k R_{total,k} \\ \text{s.t. } &\text{tr}(\mathbf{P} \mathbf{P}^H) \leq P_t \\ &R_{total,k} \geq R_k^{th}, k \in \{1, 2\} \\ &Z_1^{12} + Z_2^{12} \leq R_{12} \\ &\mathbf{z} \geq \mathbf{0} \end{aligned} \quad (3.5)$$

where R_k^{th} is the rate threshold for user- k (i.e. the minimum rate required to achieve). $\mathbf{z} = [Z_1^{12}, Z_2^{12}]$ is the common achievable rate vector and $\mathbf{P} = [\mathbf{p}_{12}, \mathbf{p}_1, \mathbf{p}_2]$ is the linear precoder matrix which both need to be optimized in order to obtain the maximum WSR. Eq. 3.5 can be solved by using WMMSE-AO algorithm presented in Section 3.4. By applying a range of various weight vector \mathbf{u} , we can get the rate region, which will be demonstrated in Section 3.6.

From the analysis above, we can know that RS is able to partially decode interference and partially treat interference as noise. Take user-1 for an example, at the stage of decoding common message W_1^{12} , part of the interference (i.e. W_2^{12}) is also decoded; at the stage of decoding private message W_1^1 , the interference from user-2's private message W_2^2 is treated as noise. If we assign no power to the common stream, RSMA boils down to SDMA as now we fully treat interference which is the other user's private message as noise. If we assign no power to one of the private streams, say s_2 of user-2, RSMA reduces to NOMA as user-1 now needs to

fully decode interference from user 2 (i.e. W_2^{12}). Therefore, RSMA incorporates the features from both SDMA and NOMA and embraces those two as extreme cases. Instead of making a binary choice between SDMA and NOMA, RSMA provides more flexibility which we will see later is a key advantage.

3.3 Three-user Network

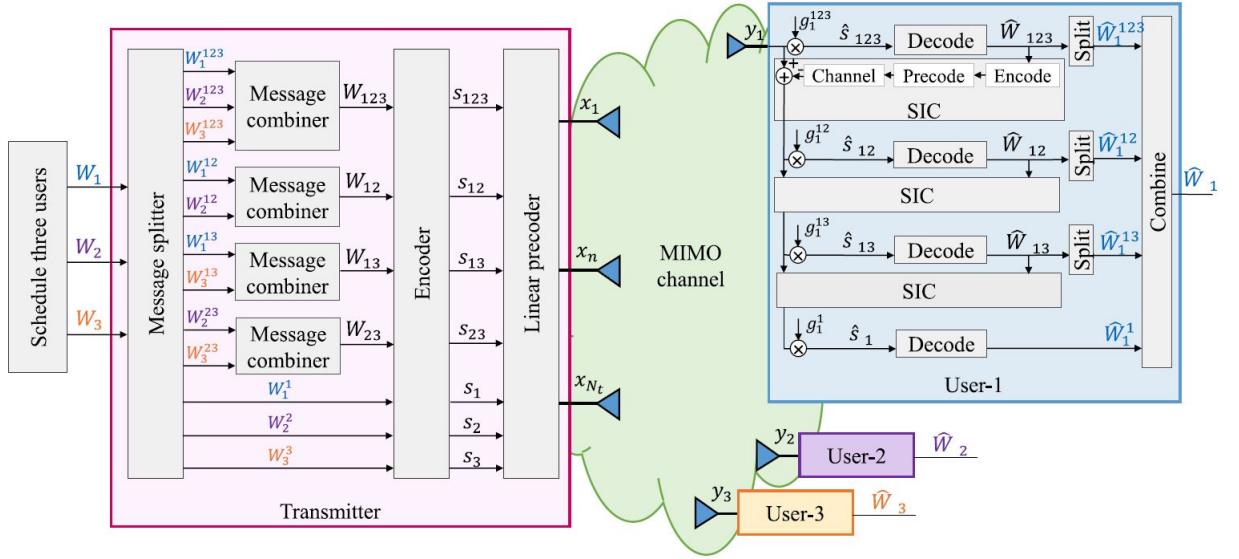


Figure 3.2: Three-user model using RSMA [1].

Now let's conduct a further analysis of RS by examining a 3-user configuration, which is more complicated than the 2-user case. Figure 3.2 shows the block diagram of a 3-user scenario. There are three messages intended for user-1, user-2 and user-3 : W_1 , W_2 and W_3 . W_1 is split into $\{W_1^{123}, W_1^{12}, W_1^{13}, W_1^1\}$, W_2 is split into $\{W_2^{123}, W_2^{12}, W_2^{13}, W_2^2\}$ and W_3 is split into $\{W_3^{123}, W_3^{13}, W_3^{23}, W_3^3\}$. Messages with same subscript belongs to a specific user and messages with same superscript belongs to a specific group of users. The messages inside the same group will be encoded together. This results in W_1^{123} , W_2^{123} and W_3^{123} being encoded into common stream s_{123} ; W_1^1 , W_2^2 and W_3^3 are encoded into private streams s_1 , s_2 and s_3 respectively. Unlike the 2-user case, besides common stream and private streams, we now have what's known as partial common streams s_{12} (encoded from W_1^{12} and W_2^{12}), s_{13} (encoded from W_1^{13} and W_3^{13}) and s_{23} (encoded from W_2^{23} and W_3^{23}). After applying linear precoder \mathbf{P} , the transmitted signal can be expressed as:

$$\mathbf{x} = \mathbf{Ps} = [\mathbf{p}_{123}, \mathbf{p}_{12}, \mathbf{p}_{13}, \mathbf{p}_{23}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3][s_{123}, s_{12}, s_{13}, s_{23}, s_1, s_2, s_3]^T \quad (3.6)$$

At the user end, decoding is more complicated than the 2-user case as now we have partial common streams. We define s_1 , s_2 and s_3 as 1-order streams, s_{12} , s_{13} and s_{23} as 2-order streams and s_{123} as 3-order stream (order number i is the amount of numbers shown in the subscript of the data stream, which is how many users will decode this message). We will analyze user-1 as an example, the decoding process of user-2 and user-3 follows the same pattern as that of user-1. For user-1, 4 streams s_{123} , s_{12} , s_{13} and s_1 will be decoded using SIC. The common (3-order) stream is firstly decoded, then the partial common (2-order) streams and finally the private (1-order) stream. At user-1, the SINR of the common stream s_{123} is:

$$\eta_1^{123} = \frac{|\mathbf{h}_1 \mathbf{p}_{123}|^2}{|\mathbf{h}_1 \mathbf{p}_{12}|^2 + |\mathbf{h}_1 \mathbf{p}_{13}|^2 + |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1} \quad (3.7)$$

Then user-1 decodes s_{12} and s_{23} by treating s_{23} as noise. Unlike the 2-user case, as we now have 2 streams to be decoded at a particular stream order, both decoding orders (i.e. first s_{12} then s_{23} or first s_{23} then s_{12}) are needed to be considered such that the WSR can be maximized using one of the two decoding orders. If we consider all three users, there will be 6 decoding orders in total. Figure 3.2 can be viewed as a system model with decoding order 12→13→23, which means for all three users, s_{13} is always decoded after s_{12} , and s_{23} is always decoded after s_{13} . With decoding order 12→13→23, the SINR of partial common streams s_{12} and s_{13} are given by:

$$\eta_1^{12} = \frac{|\mathbf{h}_1 \mathbf{p}_{12}|^2}{|\mathbf{h}_1 \mathbf{p}_{13}|^2 + |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1} \quad (3.8)$$

$$\eta_1^{13} = \frac{|\mathbf{h}_1 \mathbf{p}_{13}|^2}{|\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1} \quad (3.9)$$

Finally, user-1 decodes private stream s_1 by treating the remaining streams as noise. The SINR of the s_1 is:

$$\eta_1 = \frac{|\mathbf{h}_1 \mathbf{p}_1|^2}{|\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=2}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1} \quad (3.10)$$

The achievable rates for the common stream, partial common streams and private stream of user-1 can now be written as:

$$\begin{aligned} R_1^{123} &= \log_2 (1 + \eta_1^{123}) \\ R_1^{12} &= \log_2 (1 + \eta_1^{12}) \\ R_1^{13} &= \log_2 (1 + \eta_1^{13}) \\ R_1 &= \log_2 (1 + \eta_1) \end{aligned} \quad (3.11)$$

When we consider all three users, similar to the 2-user case, in order for s_{123} to be decoded successfully by all users, the common achievable rate should not be greater than $R_{123} = \min\{R_1^{123}, R_2^{123}, R_3^{123}\}$. This also applies to the partial common streams, and we have $R_{12} = \min\{R_1^{12}, R_2^{12}\}$, $R_{13} = \min\{R_1^{13}, R_3^{13}\}$ and $R_{23} = \min\{R_2^{23}, R_3^{23}\}$. Assuming that the common achievable rate is shared by users, we can get $R_{123} = Z_1^{123} + Z_2^{123} + Z_3^{123}$, $R_{12} = Z_1^{12} + Z_2^{12}$, $R_{13} = Z_1^{13} + Z_3^{13}$ and $R_{23} = Z_2^{23} + Z_3^{23}$. The total achievable rate for three users can then be expressed as: $R_{total,1} = Z_1^{123} + Z_1^{12} + Z_1^{13} + R_1$, $R_{total,2} = Z_2^{123} + Z_2^{12} + Z_2^{23} + R_2$ and $R_{total,3} = Z_3^{123} + Z_3^{13} + Z_3^{23} + R_3$. For one specific decoding order π , if we assign a weight vector $\mathbf{u} = [u_1, u_2, u_3]$ to three users, the WSR of 3-user RS is:

$$\begin{aligned} \text{WSR } (\mathbf{u}, \pi) &= \max_{\mathbf{P}, \mathbf{z}} \sum_{k=1}^3 u_k R_{total,k} \\ \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\ &R_{total,k} \geq R_k^{th}, k \in \{1, 2, 3\} \\ &Z_1^{123} + Z_2^{123} + Z_3^{123} \leq R_{123} \\ &Z_1^{12} + Z_2^{12} \leq R_{12} \\ &Z_1^{13} + Z_3^{13} \leq R_{13} \\ &Z_2^{23} + Z_3^{23} \leq R_{23} \\ &\mathbf{z} \geq \mathbf{0} \end{aligned} \tag{3.12}$$

where $\mathbf{z} = [Z_1^{123}, Z_2^{123}, Z_3^{123}, Z_1^{12}, Z_2^{12}, Z_1^{13}, Z_3^{13}, Z_2^{23}, Z_3^{23}]$ is the common rate vector and $\mathbf{P} = [\mathbf{p}_{123}, \mathbf{p}_{12}, \mathbf{p}_{13}, \mathbf{p}_{23}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]$ is the linear precoder matrix which both need to be optimized in order to get the maximum WSR. Eq. 3.12 can be solved by using WMMSE-AO algorithm in Section 3.4. In order to get the maximum WSR, all decoding orders needs to be considered (the WSR in the above equation is calculated under one specific decoding order). A generalized K -user RS [1] can be developed using the same concept. Relevant results and analysis will be demonstrated in Section 3.6.

As we can observe, more complexities are introduced to the 3-user system compared with the 2-user network since we need to consider all the decoding orders whichever maximize the rate. The complexity increases significantly with the number of users. The number of decoding orders of i -order streams ($i \in [2, K - 1]$) for a group of K users is $\binom{K}{i}!$. For example, for 2-order streams of a 4-user-system, there are $\binom{4}{2}! = \frac{4!}{2! \cdot 2!}! = 6! = 720$ different decoding orders, for 3-order streams of

the 4-user-system, there are $\binom{4}{3}! = 4! = 24$ different decoding orders. This means in total we need to consider $720 \times 24 = 17280$ decoding orders, which increases complexity dramatically. To resolve this issue, a simplified version of RS (1-layer RS) will be introduced in Section 3.5.

3.4 WMMSE-AO Algorithm

In this section, we will discuss the WMMSE-AO algorithm in details by continuing examining user-1 of the three-user system (with decoding order 12→13→23, same process can be applied to user-2 and user-3). The WMMSE-AO algorithm is adopted from [33] and can be easily extended to a generalized K -user RS. The core of WMMSE-AO algorithm is to transform WSR (or SR) problem into WMMSE problem and then solve it using AO algorithm. The optimal precoder is then found and the WSR (or SR) is maximized.

Firstly, the common stream s_{123} is decoded and its estimation is $\hat{s}_{123} = g_1^{123}y_1$ where y_1 is the received signal at user-1 and g_1^{123} is the combiner. Secondly, by subtracting \hat{s}_{123} from received signal, 2-orders streams can be decoded, and the estimates of them are: $\hat{s}_{12} = g_1^{12}(y_1 - \mathbf{h}_1\mathbf{p}_{123}\hat{s}_{123})$, $\hat{s}_{13} = g_1^{13}(y_1 - \mathbf{h}_1\mathbf{p}_{123}\hat{s}_{123} - \mathbf{h}_1\mathbf{p}_{12}\hat{s}_{12})$. Finally, private stream is decoded, and the estimate is: $\hat{s}_1 = g_1^1(y_1 - \mathbf{h}_1\mathbf{p}_{123}\hat{s}_{123} - \mathbf{h}_1\mathbf{p}_{12}\hat{s}_{12} - \mathbf{h}_1\mathbf{p}_{13}\hat{s}_{13})$. The received signal power X at each SIC layer is given by:

$$\begin{aligned}
 X_1^{123} &= |\mathbf{h}_1\mathbf{p}_{123}|^2 + |\mathbf{h}_1\mathbf{p}_{12}|^2 + |\mathbf{h}_1\mathbf{p}_{13}|^2 + |\mathbf{h}_1\mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1\mathbf{p}_k|^2 + 1 \\
 X_1^{12} &= |\mathbf{h}_1\mathbf{p}_{12}|^2 + |\mathbf{h}_1\mathbf{p}_{13}|^2 + |\mathbf{h}_1\mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1\mathbf{p}_k|^2 + 1 \\
 X_1^{13} &= |\mathbf{h}_1\mathbf{p}_{13}|^2 + |\mathbf{h}_1\mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1\mathbf{p}_k|^2 + 1 \\
 X_1^1 &= |\mathbf{h}_1\mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1\mathbf{p}_k|^2 + 1
 \end{aligned} \tag{3.13}$$

And the noise power Y are:

$$\begin{aligned}
 Y_1^{123} &= |\mathbf{h}_1 \mathbf{p}_{12}|^2 + |\mathbf{h}_1 \mathbf{p}_{13}|^2 + |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1 \\
 Y_1^{12} &= |\mathbf{h}_1 \mathbf{p}_{13}|^2 + |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1 \\
 Y_1^{13} &= |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=1}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1 \\
 Y_1^1 &= |\mathbf{h}_1 \mathbf{p}_{23}|^2 + \sum_{k=2}^3 |\mathbf{h}_1 \mathbf{p}_k|^2 + 1
 \end{aligned} \tag{3.14}$$

The definition of MSE is $e_1 = \mathbb{E}\{|s_1 - \hat{s}_1|^2\}$, and can be expressed as:

$$\begin{aligned}
 e_1^{123} &= |g_1^{123}|^2 X_1^{123} - 2\Re\{g_1^{123} \mathbf{h}_1 \mathbf{p}_{123}\} + 1 \\
 e_1^{12} &= |g_1^{12}|^2 X_1^{12} - 2\Re\{g_1^{12} \mathbf{h}_1 \mathbf{p}_{12}\} + 1 \\
 e_1^{13} &= |g_1^{13}|^2 X_1^{13} - 2\Re\{g_1^{13} \mathbf{h}_1 \mathbf{p}_{13}\} + 1 \\
 e_1^1 &= |g_1^1|^2 X_1^1 - 2\Re\{g_1^1 \mathbf{h}_1 \mathbf{p}_1\} + 1
 \end{aligned} \tag{3.15}$$

By taking the partial derivative with respect to the combiner, and solving $\frac{\partial e_1^{123}}{\partial g_1^{123}} = 0$, $\frac{\partial e_1^{12}}{\partial g_1^{12}} = 0$, $\frac{\partial e_1^{13}}{\partial g_1^{13}} = 0$, $\frac{\partial e_1^1}{\partial g_1^1} = 0$, we can get the optimal MMSE combiner:

$$\begin{aligned}
 (g_1^{123})^{\text{MMSE}} &= \mathbf{p}_{123}^H \mathbf{h}_1^H / X_1^{123} \\
 (g_1^{12})^{\text{MMSE}} &= \mathbf{p}_{12}^H \mathbf{h}_1^H / X_1^{12} \\
 (g_1^{13})^{\text{MMSE}} &= \mathbf{p}_{13}^H \mathbf{h}_1^H / X_1^{13} \\
 (g_1^1)^{\text{MMSE}} &= \mathbf{p}_1^H \mathbf{h}_1^H / X_1^1
 \end{aligned} \tag{3.16}$$

Substituting 3.16 into Eq.3.15, we can get MMSE:

$$\begin{aligned}
 (e_1^{123})^{\text{MMSE}} &= \min_{g_1^{123}} e_1^{123} = Y_1^{123} / X_1^{123} \\
 (e_1^{12})^{\text{MMSE}} &= \min_{g_1^{12}} e_1^{12} = Y_1^{12} / X_1^{12} \\
 (e_1^{13})^{\text{MMSE}} &= \min_{g_1^{13}} e_1^{13} = Y_1^{13} / X_1^{13} \\
 (e_1^1)^{\text{MMSE}} &= \min_{g_1^1} e_1^1 = Y_1^1 / X_1^1
 \end{aligned} \tag{3.17}$$

Using Eq.3.17, the SINR for different streams of user-1 (Eq.3.7-3.10) can be re-written as:

$$\begin{aligned}\eta_1^{123} &= \frac{1}{(e_1^{123})^{\text{MMSE}}} - 1 \\ \eta_1^{12} &= \frac{1}{(e_1^{12})^{\text{MMSE}}} - 1 \\ \eta_1^{13} &= \frac{1}{(e_1^{13})^{\text{MMSE}}} - 1 \\ \eta_1^1 &= \frac{1}{(e_1^1)^{\text{MMSE}}} - 1\end{aligned}\tag{3.18}$$

The corresponding rates in Eq.3.11 can be then re-written as:

$$\begin{aligned}R_1^{123} &= -\log_2((e_1^{123})^{\text{MMSE}}) \\ R_1^{12} &= -\log_2((e_1^{12})^{\text{MMSE}}) \\ R_1^{13} &= -\log_2((e_1^{13})^{\text{MMSE}}) \\ R_1^1 &= -\log_2((e_1^1)^{\text{MMSE}})\end{aligned}\tag{3.19}$$

Now we need to transform the WSR problem into WMMSE problem. In order to do so, we introduce a new term, namely augmented WMSE, which is defined as:

$$\begin{aligned}\phi_1^{123} &= u_1^{123}e_1^{123} - \log_2(u_1^{123}) \\ \phi_1^{12} &= u_1^{12}e_1^{12} - \log_2(u_1^{12}) \\ \phi_1^{13} &= u_1^{13}e_1^{13} - \log_2(u_1^{13}) \\ \phi_1^1 &= u_1^1e_1^1 - \log_2(u_1^1)\end{aligned}\tag{3.20}$$

where $u_1^{123}, u_1^{12}, u_1^{13}, u_1^1$ are weights associated with MSE. By substituting Eq.3.15 into Eq.3.20 and taking the partial derivative of Eq.3.20 with respect to the combiner g , we solve $\frac{\partial\phi_1^{123}}{\partial g_1^{123}} = 0$, $\frac{\partial\phi_1^{12}}{\partial g_1^{12}} = 0$, $\frac{\partial\phi_1^{13}}{\partial g_1^{13}} = 0$, $\frac{\partial\phi_1^1}{\partial g_1^1} = 0$ and can get the optimum combiners: $(g_1^{123})^* = (g_1^{123})^{\text{MMSE}}$, $(g_1^{12})^* = (g_1^{12})^{\text{MMSE}}$, $(g_1^{13})^* = (g_1^{13})^{\text{MMSE}}$, $(g_1^1)^* = (g_1^1)^{\text{MMSE}}$. Then the augmented WMMSE can be obtained by substituting the optimum combiners into Eq.3.20 :

$$\begin{aligned}\phi_1^{123}((g_1^{123})^{\text{MMSE}}) &= u_1^{123}(e_1^{123})^{\text{MMSE}} - \log_2(u_1^{123}) \\ \phi_1^{12}((g_1^{12})^{\text{MMSE}}) &= u_1^{12}(e_1^{12})^{\text{MMSE}} - \log_2(u_1^{12}) \\ \phi_1^{13}((g_1^{13})^{\text{MMSE}}) &= u_1^{13}(e_1^{13})^{\text{MMSE}} - \log_2(u_1^{13}) \\ \phi_1^1((g_1^1)^{\text{MMSE}}) &= u_1^1(e_1^1)^{\text{MMSE}} - \log_2(u_1^1)\end{aligned}\tag{3.21}$$

By further taking partial derivative of Eq.3.21 with respect to MSE weights and solving $\frac{\partial \phi_1^{123}((g_1^{123})^{\text{MMSE}})}{\partial u_1^{123}} = 0$, $\frac{\partial \phi_1^{12}((g_1^{12})^{\text{MMSE}})}{\partial u_1^{12}} = 0$, $\frac{\partial \phi_1^{13}((g_1^{13})^{\text{MMSE}})}{\partial u_1^{13}} = 0$, $\frac{\partial \phi_1^1((g_1^1)^{\text{MMSE}})}{\partial u_1^1} = 0$, the optimum MMSE weights can be obtained:

$$\begin{aligned}(u_1^{123})^* &= (u_1^{123})^{\text{MMSE}} = 1/(e_1^{123})^{\text{MMSE}} \\ (u_1^{12})^* &= (u_1^{12})^{\text{MMSE}} = 1/(e_1^{12})^{\text{MMSE}} \\ (u_1^{13})^* &= (u_1^{13})^{\text{MMSE}} = 1/(e_1^{13})^{\text{MMSE}} \\ (u_1^1)^* &= (u_1^1)^{\text{MMSE}} = 1/(e_1^1)^{\text{MMSE}}\end{aligned}\tag{3.22}$$

Finally, substitute Eq.3.22 into Eq.3.21, we can get the WMMSE-Rate relationship:

$$\begin{aligned}(\phi_1^{123})^{\text{MMSE}} &= \min_{u_1^{123}, g_1^{123}} \phi_1^{123} = 1 - R_1^{123} \\ (\phi_1^{12})^{\text{MMSE}} &= \min_{u_1^{12}, g_1^{12}} \phi_1^{12} = 1 - R_1^{12} \\ (\phi_1^{13})^{\text{MMSE}} &= \min_{u_1^{13}, g_1^{13}} \phi_1^{13} = 1 - R_1^{13} \\ (\phi_1^1)^{\text{MMSE}} &= \min_{u_1^1, g_1^1} \phi_1^1 = 1 - R_1^1\end{aligned}\tag{3.23}$$

Having obtained the WMMSE-Rate relationship, by considering weight vector $\mathbf{u} = [u_1, u_2, u_3]$, we can re-write the WSR optimization problem in Eq.3.12 as follows:

$$\begin{aligned}\text{WMMSE} &= \min_{\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u}} \sum_{k=1}^3 u_k \phi_{total,k} \\ \text{s.t.} \quad \text{tr}(\mathbf{P} \mathbf{P}^H) &\leq P_t \\ \phi_{total,k} &\leq 1 - R_k^{th}, k \in \{1, 2, 3\} \\ T_1^{123} + T_2^{123} + T_3^{123} + 1 &\geq \phi_{123} \\ T_1^{12} + T_2^{12} + 1 &\geq \phi_{12} \\ T_1^{13} + T_3^{13} + 1 &\geq \phi_{13} \\ T_2^{23} + T_3^{23} + 1 &\geq \phi_{23} \\ \mathbf{t} &\leq \mathbf{0}\end{aligned}\tag{3.24}$$

where $\mathbf{t} = [T_1^{123}, T_2^{123}, T_3^{123}, T_1^{12}, T_2^{12}, T_3^{12}, T_1^{13}, T_2^{13}, T_3^{13}]$ is the common WMSE vector, $\mathbf{u} = [u_1^{123}, u_2^{123}, u_3^{123}, u_1^{12}, u_2^{12}, u_3^{12}, u_1^{13}, u_2^{13}, u_3^{13}, u_2^{23}, u_3^{23}]$, $\mathbf{g} = [g_1^{123}, g_2^{123}, g_3^{123}, g_1^{12}, g_2^{12}, g_3^{12}, g_1^{13}, g_2^{13}, g_3^{13}, g_2^{23}, g_3^{23}]$, $\mathbf{P} = [\mathbf{p}_{123}, \mathbf{p}_{12}, \mathbf{p}_{13}, \mathbf{p}_{23}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]$. The individual WMSE for each user is $\phi_{total,1} = T_1^{123} + T_1^{12} + T_1^{13} + \phi_1^1$, $\phi_{total,2} = T_2^{123} + T_2^{12} + T_2^{23} + \phi_2^2$ and $\phi_{total,3} = T_3^{123} + T_3^{13} + T_3^{23} + \phi_3^3$. The achievable WMSE of common stream and partial common streams are: $\phi_{123} = \max\{\phi_1^{123}, \phi_2^{123}, \phi_3^{123}\}$, $\phi_{12} = \max\{\phi_1^{12}, \phi_2^{12}\}$, $\phi_{13} = \max\{\phi_1^{13}, \phi_3^{13}\}$ and $\phi_{23} = \max\{\phi_2^{23}, \phi_3^{23}\}$.

By minimizing the objective function in Eq.3.24 with respect to \mathbf{g} and \mathbf{u} we can get the MMSE solution \mathbf{g}^{MMSE} and \mathbf{u}^{MMSE} , which satisfy the KKT conditions for \mathbf{P} in Eq.3.24. Together with the transformation of $\mathbf{z} = -\mathbf{t}$ and the WMMSE-Rate relationship, we can conclude that the optimization of (\mathbf{P}, \mathbf{z}) in Eq.3.12 is equivalent to the optimization of $(\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u})$ in Eq.3.24. Therefore, WSR problem can be properly transformed into WMMSE problem.

Nevertheless, Eq.3.24 is still non-convex for the optimization of $(\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u})$. From Eq.3.16 and 3.22 we know that the optimal combiner is \mathbf{g}^{MMSE} when $(\mathbf{P}, \mathbf{t}, \mathbf{u})$ are fixed; the optimal weight is \mathbf{u}^{MMSE} when $(\mathbf{P}, \mathbf{t}, \mathbf{g})$ are fixed. However, (\mathbf{P}, \mathbf{t}) are coupled when (\mathbf{g}, \mathbf{u}) are fixed, which indicates we will not get closed-form solutions. If we consider (\mathbf{P}, \mathbf{t}) , (\mathbf{g}) , (\mathbf{u}) as three blocks, the WMMSE problem is block-wise convex in each of the blocks when all other blocks are fixed. Eq.3.24 is now a convex Quadratically Constrained Quadratic Program (QCQP) which can be solved using interior-point methods. The above-mentioned features will allow us using AO algorithm below to solve the optimization problem. In the m^{th} iteration, \mathbf{g} and \mathbf{u} are updated using \mathbf{P} from last iteration (\mathbf{g} and \mathbf{u} are functions of \mathbf{P}). Then (\mathbf{P}, \mathbf{t}) are updated using \mathbf{g} and \mathbf{u} from the current iteration. Such process is continued until WMMSE converges and we obtain the optimal parameters. Since the optimization problem is bounded by the power constraint $\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$ and WMMSE is decreasing iteratively, the AO algorithm will converge to a local minimum². The optimization problem in Eq.3.24 is solved using CVX toolbox [34].

Algorithm 1: Alternating Optimization Algorithm (For RSMA)

```

Initialize:  $m \leftarrow 0$ ,  $\mathbf{P}^{[m]}$ , WMMSE $^{[m]}$ 
repeat
   $m \leftarrow m + 1$ ,  $\mathbf{P}^{[m-1]} \leftarrow \mathbf{P}$ ;
   $\mathbf{g} \leftarrow \mathbf{g}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$ ,  $\mathbf{u} \leftarrow \mathbf{u}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$ ;
  update  $(\mathbf{P}, \mathbf{t})$  by solving Eq.3.24 using updated  $\mathbf{g}$  and  $\mathbf{u}$ ;
until  $|\text{WMMSE}^{[m]} - \text{WMMSE}^{[m-1]}| \leq \epsilon$ ;

```

²It should be noted that the WSR problem (Eq.3.12) does not necessarily monotonically converges (i.e. the local minimum is not necessarily the global minimum), although numerous simulation results indicate this is usually the case [33].

3.5 1-layer RSMA

In Section 3.3 we discussed RS for a 3-user system, and the complexity is introduced by the fact that we need to consider all decoding orders. Such problem can be simplified by using 1-layer RS, where we only have common (K -order) streams and private (1-order) streams. Except there's no partial common streams, other properties remains the same as the general RS; such configuration is similar to the 2-user case. The message intended for each user is split into two parts: $W_{c,k}$ and $W_k, \forall k \in \mathcal{K}$. The common message W_c consists all single-user common messages ($W_{c,k}$) and is encoded into common stream s_c , which will be decoded by all users. Private message W_k is encoded into private stream s_k , and will be decoded only by user- k . After applying linear precoding, the transmitted signal can be written as:

$$\mathbf{x} = \mathbf{Ps} = [\mathbf{p}_c, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K][s_c, s_1, s_2, \dots, s_K]^T \quad (3.25)$$

At the receiver end, the common stream s_c will firstly be decoded by viewing all private streams as noise. The SINR of the common (K -order) stream at user- k is:

$$\eta_{c,k} = \frac{|\mathbf{h}_k \mathbf{p}_c|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (3.26)$$

After the K -order stream is decoded, the contribution of it will be subtracted from the received signal. Then user- k will decode its corresponding private stream by treating the private streams of the other users as noise. At user- k , the SINR of the private stream is:

$$\eta_k = \frac{|\mathbf{h}_k \mathbf{p}_k|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (3.27)$$

The achievable rates for common stream s_c and private stream s_k of user- k are $R_{c,k} = \log_2 (1 + \eta_{c,k})$ and $R_k = \log_2 (1 + \eta_k)$ respectively. In order for the common stream to be decoded successfully by all users, the common achievable rate should not be greater than $R_c = \min\{R_{c,1}, R_{c,2}, \dots, R_{c,K}\}$. We assume that the common achievable rate is shared by all users such that $R_c = \sum_{k \in \mathcal{K}} Z_{c,k}$. We can then write the total achievable rate of user- k as: $R_{total,k} = Z_{c,k} + R_k$. Allocating a weight vector

$\mathbf{u} = [u_1, u_2, \dots, u_K]$ to all users, the WSR of 1-layer RS is:

$$\begin{aligned}
 \text{WSR}(\mathbf{u}) = & \max_{\mathbf{P}, \mathbf{z}} \sum_{k \in \mathcal{K}} u_k R_{total,k} \\
 \text{s.t. } & \text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\
 & R_{total,k} \geq R_k^{th}, \forall k \in \mathcal{K} \\
 & \sum_{k \in \mathcal{K}} Z_{c,k} \leq R_c \\
 & \mathbf{z} \geq \mathbf{0}
 \end{aligned} \tag{3.28}$$

where R_k^{th} is the rate threshold and $\mathbf{z} = [Z_{c,1}, Z_{c,2}, \dots, Z_{c,K}]$ is the common achievable rate vector. Eq. 3.28 can be solved by using WMMSE-AO algorithm in Section 3.4.

3.6 Results and Discussions

3.6.1 Two-User Network with Random Channel Realizations

We will first examine the rate region of the 2-user case. The rate region is obtained by firstly assigning the common message to one user the whole time and get the first region; then the common message is allocated to the other user the entire time to get the second region. The rate region is the convex hull enclosing the two aforementioned regions [2]. The boundary of the rate region is obtained by varying the weight vector $\mathbf{u} = [u_1, u_2]$. Following the similar parameter settings in [1], u_1 is fixed at 1 for all simulations, $u_2 = 10^{[-3, -1-0.95, -0.9, \dots, 0.9, 0.95, 1, 3]}$. The lower limit (R^{th}) of the achievable rate is set as 0 for both users in order to achieve the largest rate region.

The initialization of precoder \mathbf{P} is important as the WSR problem is non-convex. Using Singular Value Decomposition (SVD) for common-stream precoder initialization and Maximum Ratio Transmission (MRT) for private-stream precoder initialization can lead to a satisfactory overall performance [2]. The common stream precoder is initialized as $\mathbf{p}_{12} = \gamma P_t \cdot \mathbf{v}_{12}$ where $\gamma \in [0, 1]$ and $\mathbf{v}_{12} = \mathbf{V}(:, 1)$ which is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]^T$. The private stream precoders are initialized as $\mathbf{p}_k = \frac{(1-\gamma)P_t}{2} \cdot \frac{\mathbf{h}_k^H}{\|\mathbf{h}_k\|}$, $\forall k \in \{1, 2\}$. For MU-LP and SC-SIC, the initialization follows the same manner in order to conduct a fair comparison. In MU-LP, there's no common message, the precoders are initialized using MRT (as for private messages). In SC-SIC, the precoder of the

last-decoded user is initialized using MRT and precoder of the first-decoded user is initialized using SVD. In this subsection we will investigate rate region performance where the user channels are drawn from Independent and Identically Distributed (i.i.d.) complex Gaussian variables with zero mean and unity variance σ_k^2 , i.e. $\mathbf{h}_k \sim \mathcal{CN}(0, \sigma_k^2)$, $\sigma_k^2 = 1, \forall k \in \{1, 2\}$.

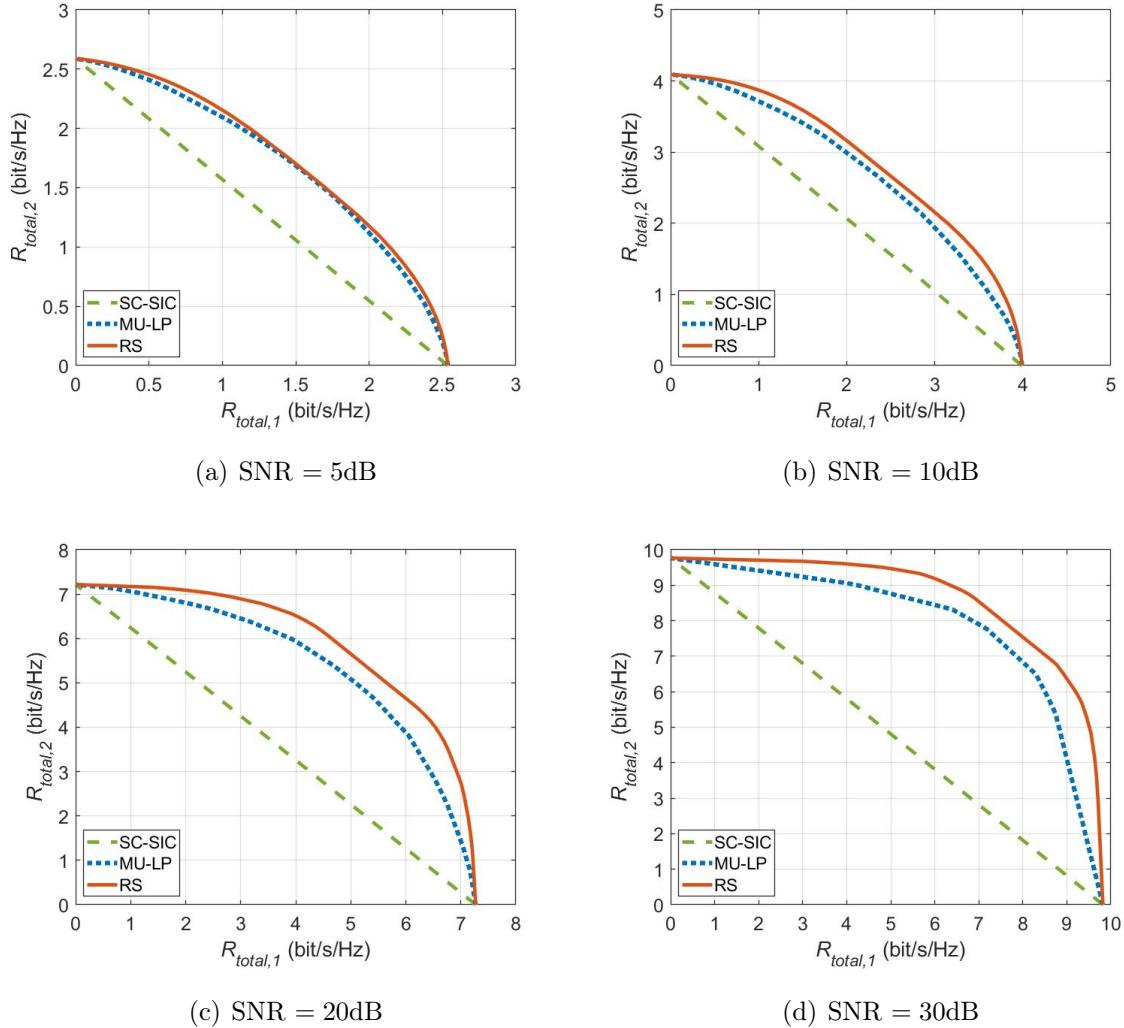


Figure 3.3: Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT averaged over 100 random channel realizations, $N_t = 2, K = 2$.

Fig.3.3 demonstrates the rate region of different schemes which is averaged over 100 random channel realizations³. As we can observe from the figure, as SNR increases,

³The rate which is averaged over various random channel realizations is actually ergodic rate, which is a long-term performance measure. A more detailed analysis can be found in [2].

the rate region grows as expected, since the rate-SNR function is monotonically increasing. In all subfigures, RS has the best performance, then it's MU-LP, SC-SIC has the smallest rate region. This is caused by that SC-SIC is not suitable for equal channel strength scenario (in this simulation $\sigma_1^2 = \sigma_2^2 = 1$). More analysis on the benefits of RS over MU-LP and SC-SIC will be presented in the following subsections.

3.6.2 Two-User Network with Specific Channel Realizations

As we have discussed in Chapter 2, using MU-LP will have a relatively good performance in an underloaded deployment if the user channels are orthogonal and the channel strengths are similar; SC-SIC will work effectively in an overloaded scenario if the user channels are aligned and the channel strengths are different. Therefore, in order to gain a better understanding of the advantages of RSMA over SDMA and NOMA, in this sub-section we will investigate the effect of the channel strength, the channel angle and the number of transmit antennas on the achievable rate by implementing specific channel realizations.

The number of transmit antennas set as 2 or 4: when $N_t = 2$, the channel is defined as: $\mathbf{h}_1 = [1, 1]$, $\mathbf{h}_2 = \lambda_2 \times [1, e^{j\theta}]$; when $N_t = 4$, the channel is defined as: $\mathbf{h}_1 = [1, 1, 1, 1]$, $\mathbf{h}_2 = \lambda_2 \times [1, e^{j\theta}, e^{j2\theta}, e^{j3\theta}]$, where λ_2 is the channel strength of user-2 and θ is the channel angle (channel strength of user-1 is set as default: 1). The channel strength of user-2 is chosen from 1 and 0.3, where $\lambda_2 = 0.3$ corresponds to a 5dB more of path loss compared with channel-1. Intuitively, when the channel angle is smaller than $\frac{\pi}{9}$, we can say the channel of user-1 and channel of user-2 are sufficiently aligned; on the other hand, when the channel angle is larger than $\frac{4\pi}{9}$, we can say the channel of user-1 and channel of user-2 are sufficiently orthogonal. Therefore, the channel angle is chosen from $\theta = [\frac{\pi}{9}, \frac{2\pi}{9}, \frac{3\pi}{9}, \frac{4\pi}{9}]$.

Fig.3.4 demonstrates the rate region of NOMA, SDMA and RSMA for $N_t = 4, K = 2, \lambda_2 = 1$. In all subfigures, RS rate region is larger or equal than the other two. When $\theta = \frac{\pi}{9}$, SC-SIC has relatively small rate region as the channel strengths of two channels are the same; MU-LP also has poor performance as the channel of user-1 and channel of user-2 are aligned. When θ increases, the gap between the SC-SIC rate region and RS rate region remains large, again due to the same channel strengths. For MU-LP, the rate region gap with RS is smaller as the two channels are more orthogonal as θ increases. As a result, in Fig.3.4(d), RS reduces to SC-SIC and two rate regions coincide.

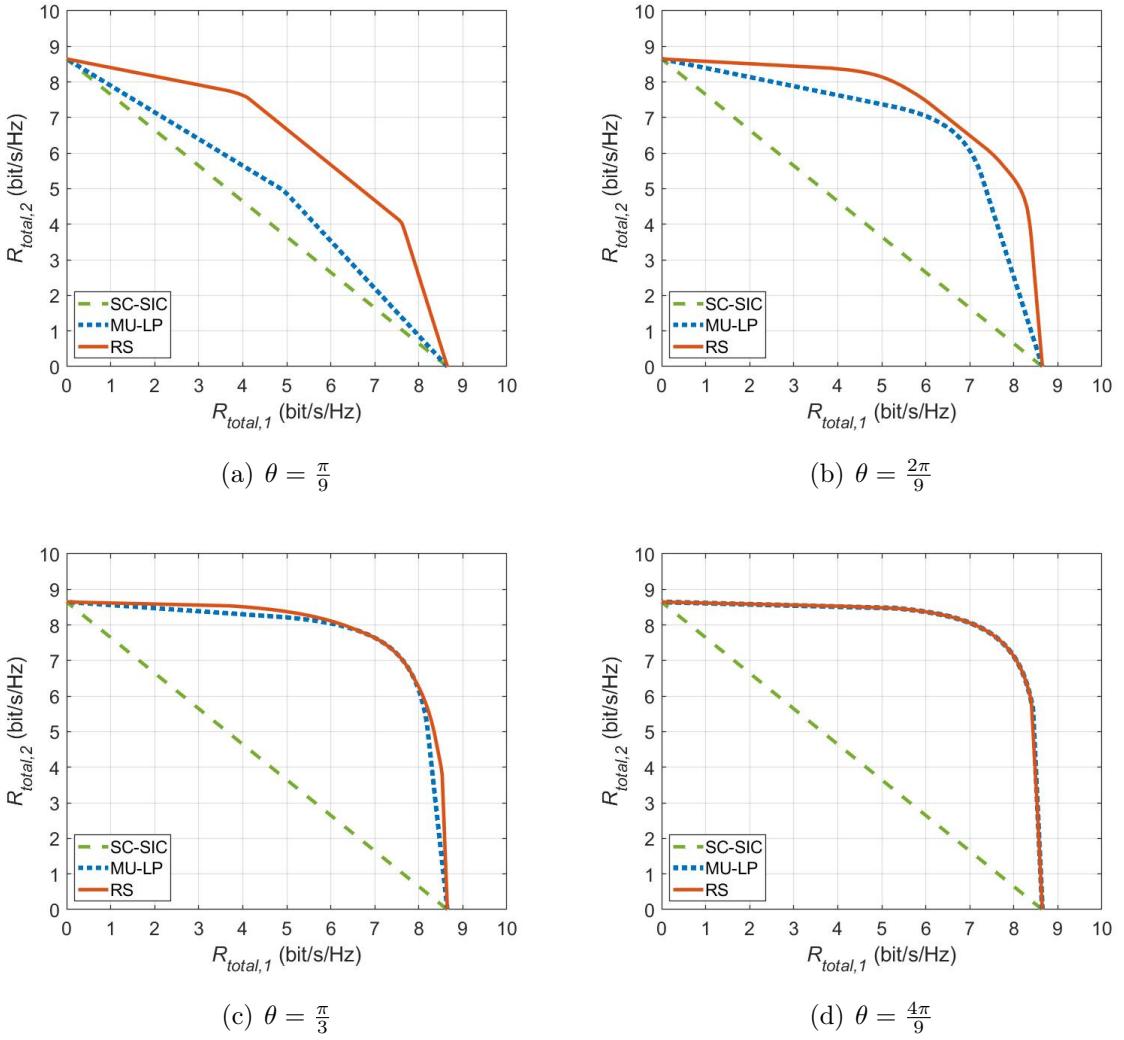


Figure 3.4: Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 4, K = 2, \lambda_2 = 1$, SNR = 20dB.

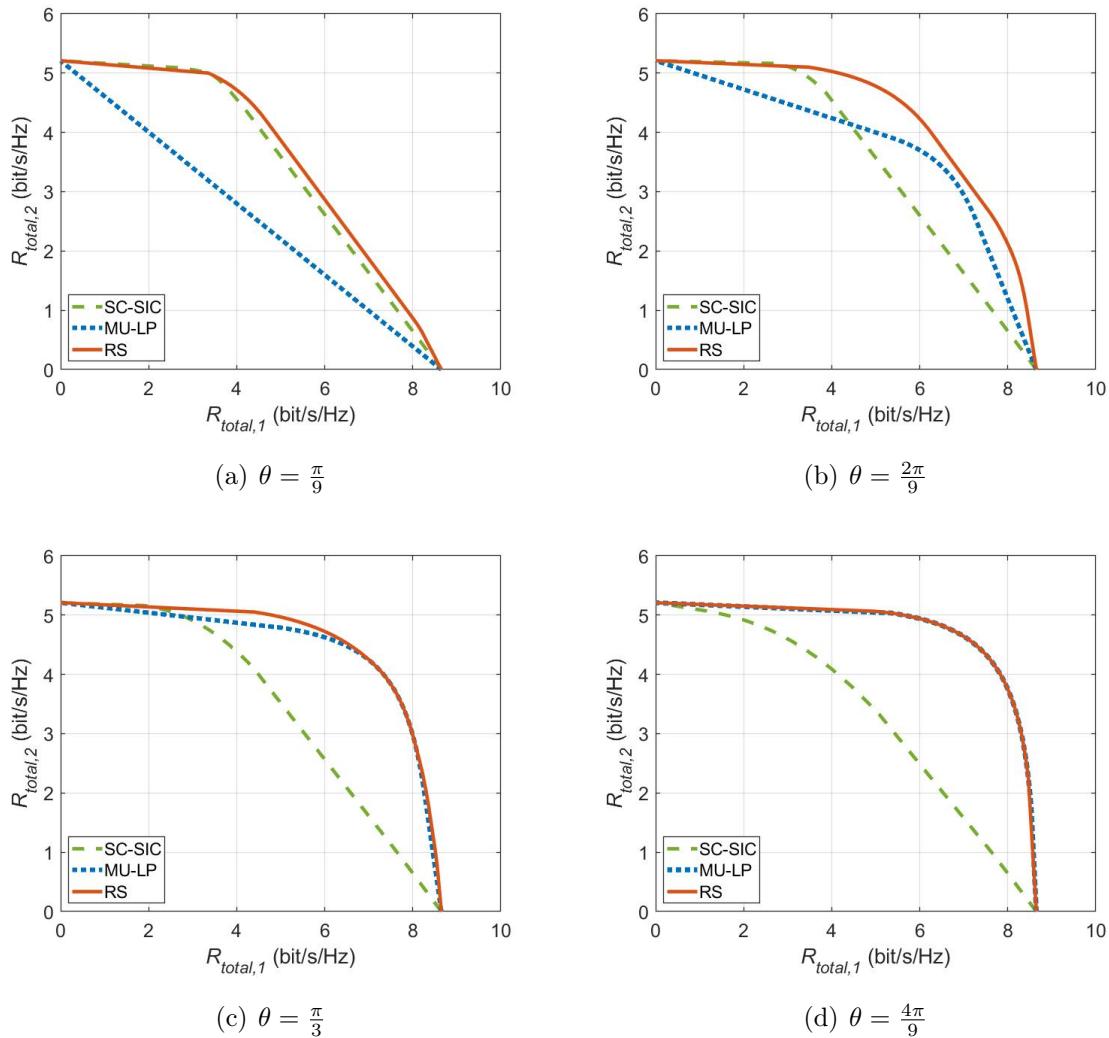


Figure 3.5: Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 4$, $K = 2$, $\lambda_2 = 0.3$, SNR = 20dB.

Fig.3.5 demonstrates the rate region of different strategies for $N_t = 4, K = 2, \lambda_2 = 0.3$. In all subfigures, RS rate region is larger or equal than the other two. The channel strengths have a 5dB difference now, so when $\theta = \frac{\pi}{9}$, SC-SIC has similar rate region with RS and MU-LP has the worst performance since two channels are sufficiently aligned. When $\theta = \frac{4\pi}{9}$, two channels are sufficiently orthogonal, SC-SIC has the worst performance and RS effectively becomes MU-LP. As θ increases, the MU-LP rate region expands from smaller than SC-SIC rate region to eventually larger than it. Therefore, unlike in Fig.3.4 where MU-LP always has better performance than SC-SIC and as θ increases the rate region gap between the two becomes larger, in Fig.3.5(b)(c), there is an intersection point between MU-LP and SC-SIC rate regions, and two schemes both have better performance than the other one in

certain regions.

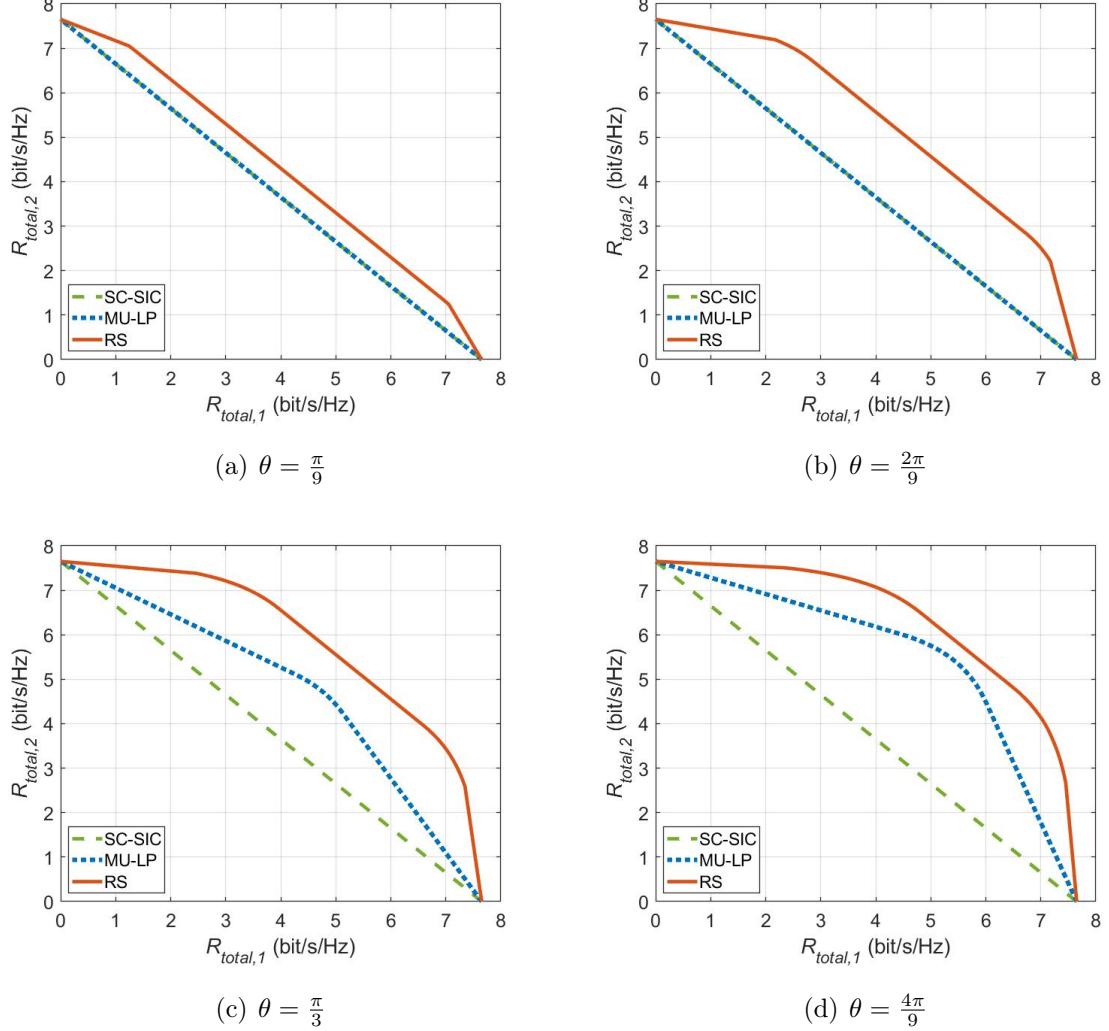


Figure 3.6: Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA with perfect CSIT, $N_t = 2, K = 2, \lambda_2 = 1$, SNR = 20dB.

Fig.3.6 demonstrates the rate region of different strategies for $N_t = 2, K = 2, \lambda_2 = 1$. In all subfigures, RS rate region is larger or equal than the other two. SC-SIC has the worst performance as the channel strengths are the same. The rate region gaps between RS and MU-LP get larger compared with the corresponding subfigures in Fig.3.4. This is because when there are less transmit antennas, it's more difficult for MU-LP to design orthogonal precoders.

As mentioned earlier, using DPC can achieve the capacity region of MISO BC. Compared with NOMA and SDMA, [1] shows that the rate region of RSMA is closer to the capacity region. To summarize, NOMA works well only in an over-

loaded network when user channels are aligned and channel strengths are different; SDMA performs better only for orthogonal channels with similar channel strengths in underloaded scenarios. RSMA outperforms NOMA and SDMA in a wide range of different parameter settings including the channel strength, the channel angle and the number of transmit antennas.

3.6.3 Three-User Network with Random Channel Realizations

We now move onto the 3-user case. The rate region is three-dimensional when $K = 3$, and is difficult to analyze visually. Therefore, we will use Sum-Rate/Weighted Sum-Rate (SR/WSR) as a measure to compare different schemes.

The initialization of the precoder is very similar to that of the 2-user example. For RS, SVD is used for (partial) common-stream precoder and MRT is used for initializing private-stream precoder. The common stream (3-order stream) is initialized using $\mathbf{P}_{123} = \gamma_1 P_t \cdot \mathbf{v}_{123}$ where $\mathbf{v}_{123} = \mathbf{V}(:, 1)$ is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]^T$. The partial common (2-order) streams are also initialized using SVD. Take precoder of stream s_{12} for example, $\mathbf{p}_{12} = \frac{\gamma_2 P_t}{3} \cdot \mathbf{v}_{12}$ where $\mathbf{v}_{12} = \mathbf{V}(:, 1)$ which is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]^T$; \mathbf{p}_{13} and \mathbf{p}_{23} are initialized the same way using channel $[\mathbf{h}_1, \mathbf{h}_3]^T$ and $[\mathbf{h}_2, \mathbf{h}_3]^T$ respectively. The private stream precoders are initialized using MRT: $\mathbf{p}_k = \frac{\gamma_3 P_t}{3} \cdot \frac{\mathbf{h}_k^H}{\|\mathbf{h}_k\|}, \forall k \in \{1, 2, 3\}$. The power fraction factor γ_k satisfies $\gamma_k \in [0, 1], \forall k \in \{1, 2, 3\}, \sum_{k=1}^3 \gamma_k = 1$. Regarding 1-layer RS, as there's no partial common streams, we can initialize the precoder by simply setting $\gamma_2 = 0$.

For MU-LP and SC-SIC, the initialization follows the same manner in order to conduct a fair comparison. In MU-LP, there's no common message, the precoders are initialized using MRT (as for private messages). For SC-SIC, assuming the decoder order is from user- $i \rightarrow$ user- $j \rightarrow$ user- k . The precoder of the firstly-decoded user- i is initialized as $\mathbf{P}_i = \gamma_1 P_t \cdot \mathbf{v}_i$ where $\mathbf{v}_i = \mathbf{V}(:, 1)$ which is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]^T$. The precoder of the secondly-decoded user- j is initialized as $\mathbf{p}_j = \gamma_2 P_t \cdot \mathbf{v}_j$ where $\mathbf{v}_j = \mathbf{V}(:, 1)$ which is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_i, \mathbf{h}_j]^T$. The precoder of the thirdly-decoded user- k is

initialized using MRT: $\mathbf{p}_k = \gamma_3 P_t \cdot \frac{\mathbf{h}_k^H}{\|\mathbf{h}_k\|}$.

Fig.3.7 shows the SR vs SNR plot for different strategies averaged over 100 random channel realizations when using different R_{th} . As we can observe, in all subfigures, RSMA outperforms NOMA and SDMA and the SR of 1-layer RSMA is very close to that of the RSMA. 1-layer RS approach RS regardless different QoS constraints, from subfigure (a) to (d), the SR gap between RS and 1 layer RS at 30dB is 0.6, 0.56, 0.54, 0.54 bit/s/Hz respectively. Comparing MU-LP and RS in Fig3.7(a), at low SNR, they have the same or very similar SR as the only the strongest user(s) is served for both schemes. In this case, RSMA reduces to SDMA since no power is allocate to the common message. As SNR increases, MU transmission starts and RS outperforms MU-LP as a result of the common message being turned on. When some QoS is considered, even at low SNR, RSMA starts to departing from SDMA thanks to the common message. At high SNR, such influence is more explicit: from (a) to (d), the SR gap between 1 layer RS and MU-LP at 30dB is: 1.61, 1.85, 1.92, 2.15 bit/s/Hz respectively. Such results show the benefits of the common message of RS, and RSMA outperforms SDMA even more when some QoS is considered. From the DoF perspective, although the slope of the MU-LP curve is slightly smaller than the slope of the RS, this doesn't mean RS have higher DoF since SNR is finite (DoF is defined as SNR tends to infinity). In fact, with perfect CSIT, RSMA and SDMA have the same DoF as discussed in Section 2.4. The DoF advantage of RS will unveil if CSIT is imperfect, and we will discuss more on DoF in Chapter 5 and 6.

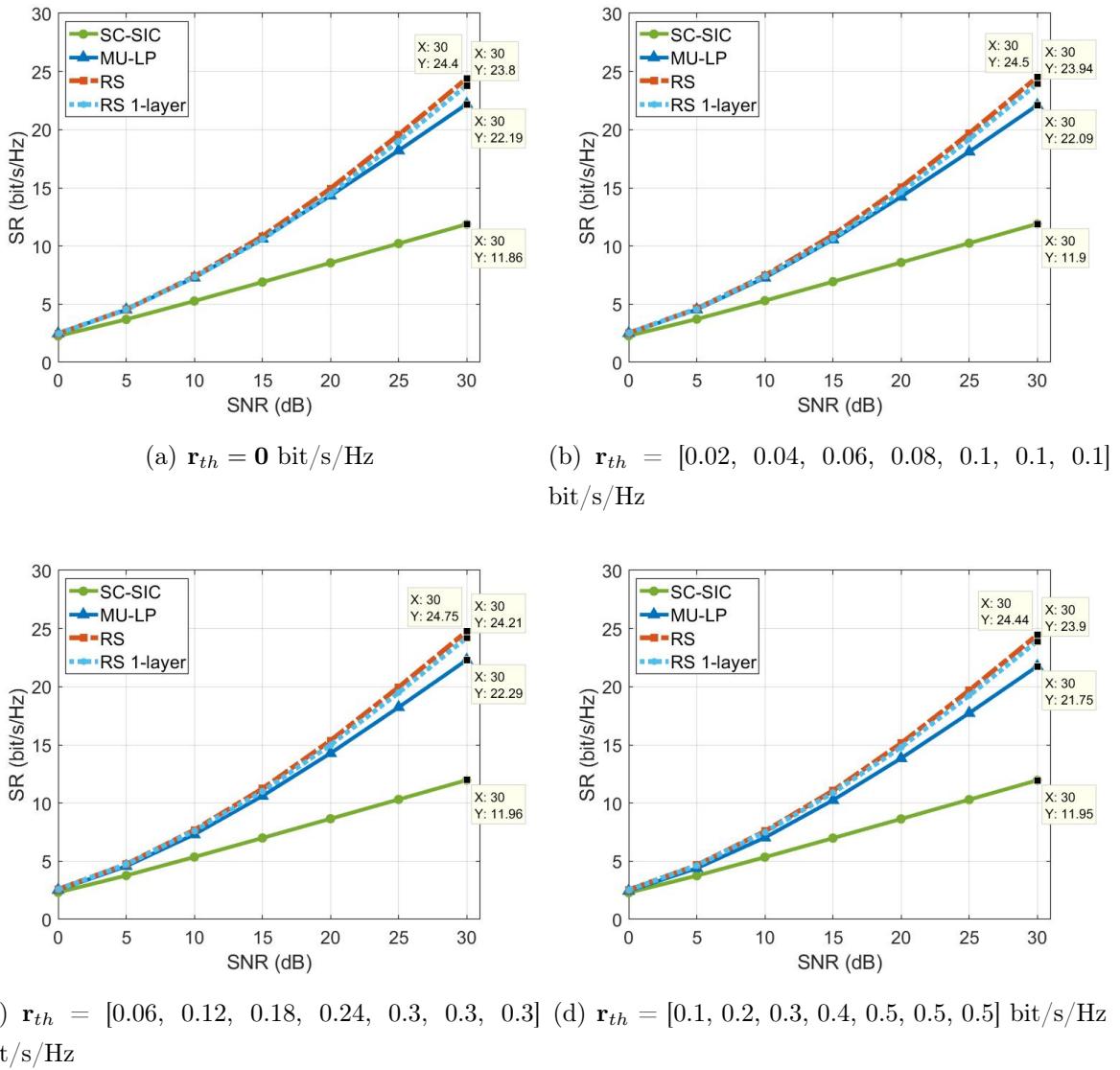


Figure 3.7: SR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA averaged over 100 channel realizations with perfect CSIT, $N_t = 3, K = 3$.

3.6.4 Three-User Network with Specific Channel Realizations

We now use specific channel realizations to further analyse the advantages of RS over MU-LP and SC-SIC. We will change several parameters including channel angle, user weight and number of transmit antennas and investigate their influence on different transmission schemes.

We first consider $N_t = 4, K = 3$, an underloaded case. The channel is defined as:

$$\begin{aligned}\mathbf{h}_1 &= [1, 1, 1, 1] \\ \mathbf{h}_2 &= \lambda_2 \times [1, e^{j\theta_2}, e^{2j\theta_2}, e^{3j\theta_2}] \\ \mathbf{h}_3 &= \lambda_3 \times [1, e^{j\theta_3}, e^{2j\theta_3}, e^{3j\theta_3}]\end{aligned}\tag{3.29}$$

where λ is channel strength and θ is channel angle. The channel angles are set as $\theta_2 = [\frac{\pi}{9}, \frac{2\pi}{9}, \frac{3\pi}{9}, \frac{4\pi}{9}]$ and $\theta_3 = 2\theta_2$; the channel strength is chosen as $\lambda_2 = 1, \lambda_3 = 0.3$.

Fig.3.8 and Fig.3.9 demonstrate the WSR vs SNR plot for $\mathbf{u} = [0.4, 0.3, 0.3]$ and $\mathbf{u} = [0.2, 0.3, 0.5]$ respectively. In almost all subfigures, RS and 1-layer RS have the same or better performance over MU-LP and SC-SIC, and 1-layer RS has the same or slightly smaller WSR compared with RS. Fig.3.8(c)(d) and Fig.3.9(c)(d) are similar in the sense that RS, 1-layer RS and MU-LP have very similar and better performance than SC-SIC. This is due to SC-SIC is not suitable for the scenario where user channels are sufficiently orthogonal, MU-LP performs well in such case and RS reduces to MP-LP eventually. Besides orthogonality, another reason for poor performance of SC-SIC is that one of the users need to decode messages of all users, which causes the sum DoF decreasing to 1 and eventually leads to lower SR.

In Fig.3.8(b)3.9(b), SC-SIC still has the worst performance, whereas RS/1-layer RS achieve higher WSR than MU-LP since the channels are not aligned enough. Comparing Fig.3.8(a)3.9(a), we can find some interesting features: in Fig.3.9(a) MU-LP has the worst performance whereas in Fig.3.8(a) (as well as other subfigures) the poorest is SC-SIC. Other than channels are aligned, another main factor affecting the performance of MU-LP in Fig.3.9(a) is the weight vector \mathbf{u} . When no individual rate constraints are set, MU-LP tends to serve users with larger channel strengths and higher weights by turning off the users with small weights and weak channel strengths. When a high weight is assigned to a weak user to assure some fairness, MU-LP is not able to deal with it effectively.

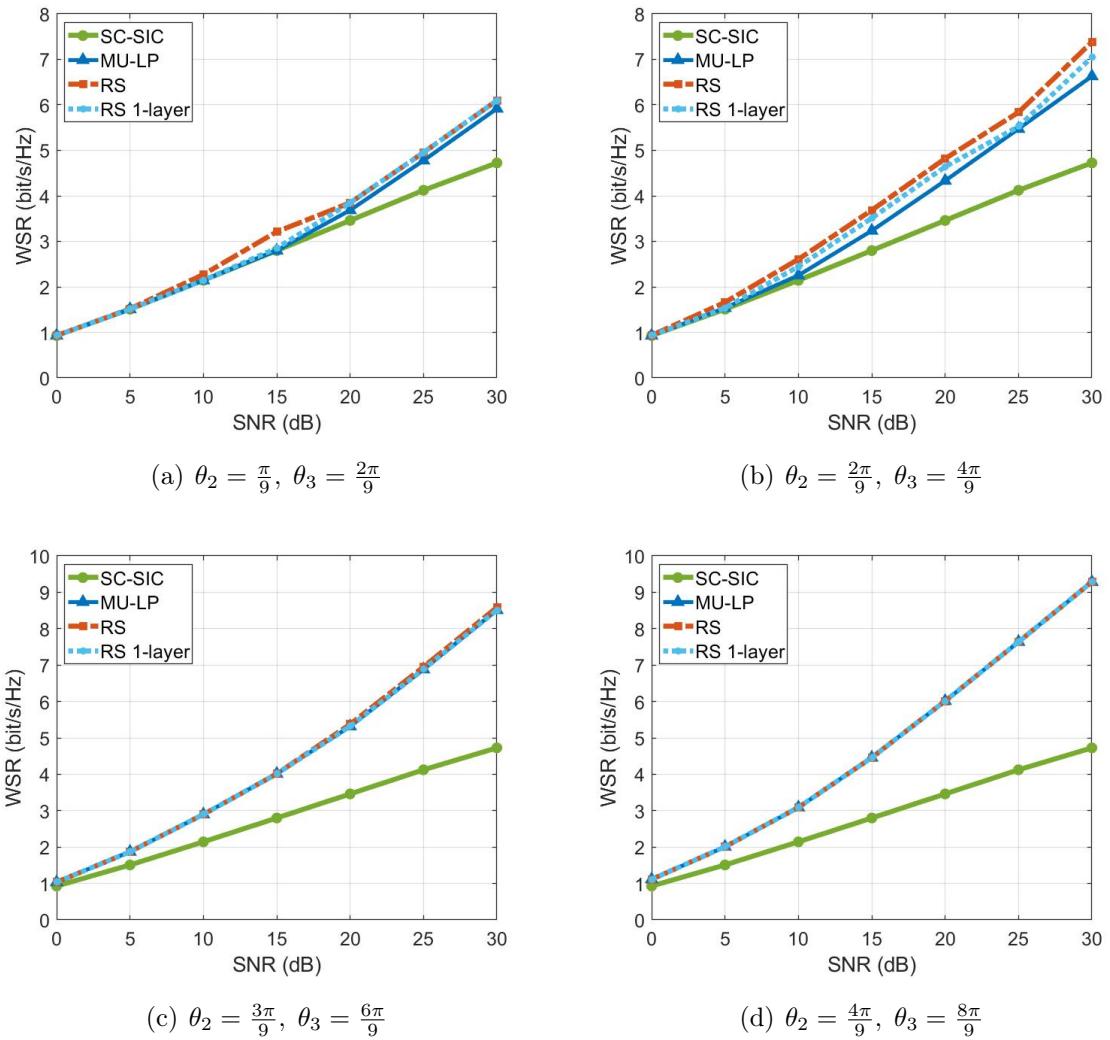


Figure 3.8: WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 4, K = 3, \mathbf{u} = [0.4, 0.3, 0.3]$.

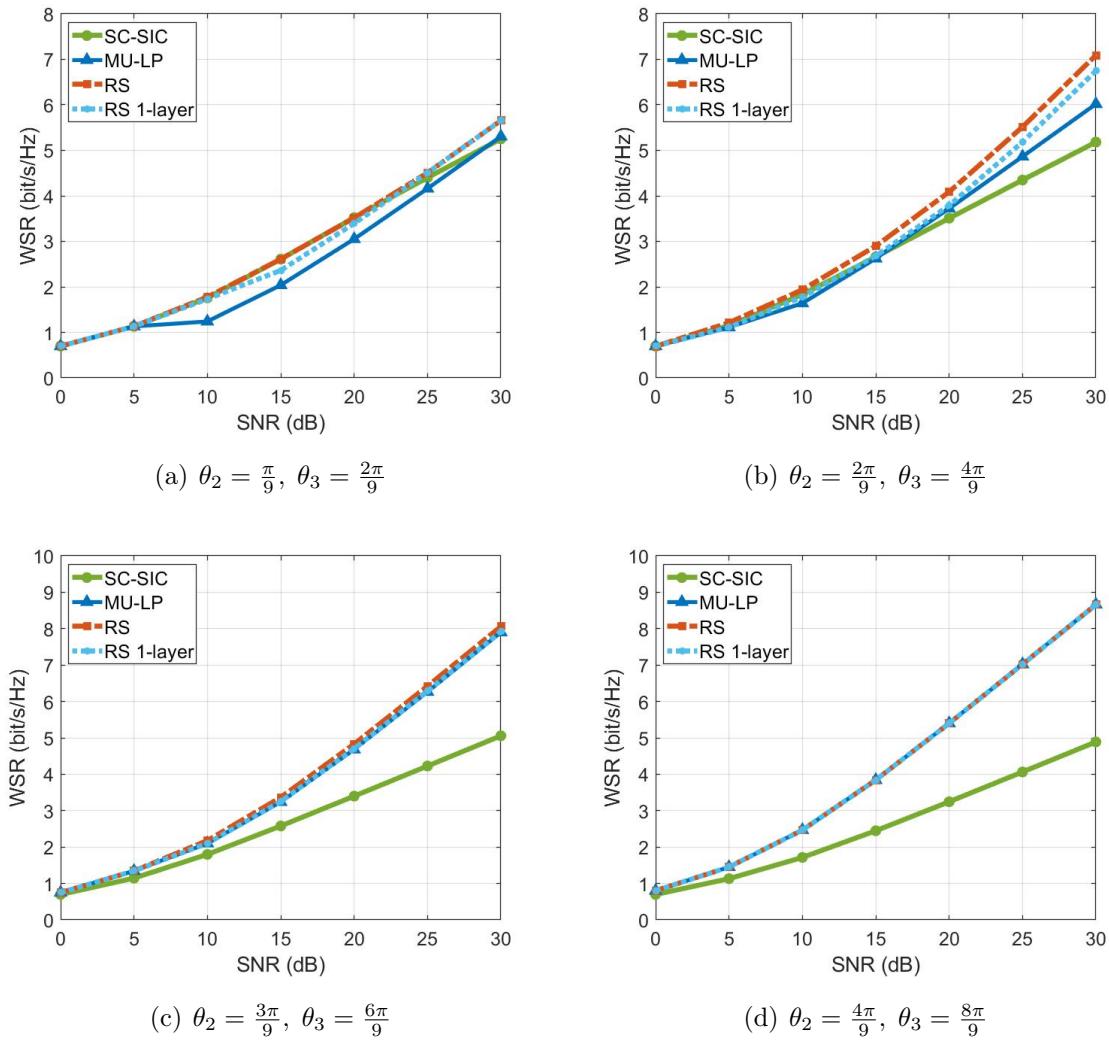


Figure 3.9: WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 4$, $K = 3$, $\mathbf{u} = [0.2, 0.3, 0.5]$.

We now turn to an overloaded scenario ($N_t = 2, K = 3$). The channel is defined as:

$$\begin{aligned}\mathbf{h}_1 &= [1, 1] \\ \mathbf{h}_2 &= \lambda_2 \times [1, e^{j\theta_2}] \\ \mathbf{h}_3 &= \lambda_3 \times [1, e^{j\theta_3}]\end{aligned}\tag{3.30}$$

In order to have some degree of QoS, the rate thresholds are set to non-zero. As SNR increases, the BS is capable of serving users with higher rate thresholds, thus the threshold is set such that it increases with SNR. Assuming the rate threshold for each user are the same, for $\text{SNR} = [0, 5, 10, 15, 20, 25, 30]$ dB we set $\mathbf{r}_{th} = [0.02, 0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bit/s/Hz.

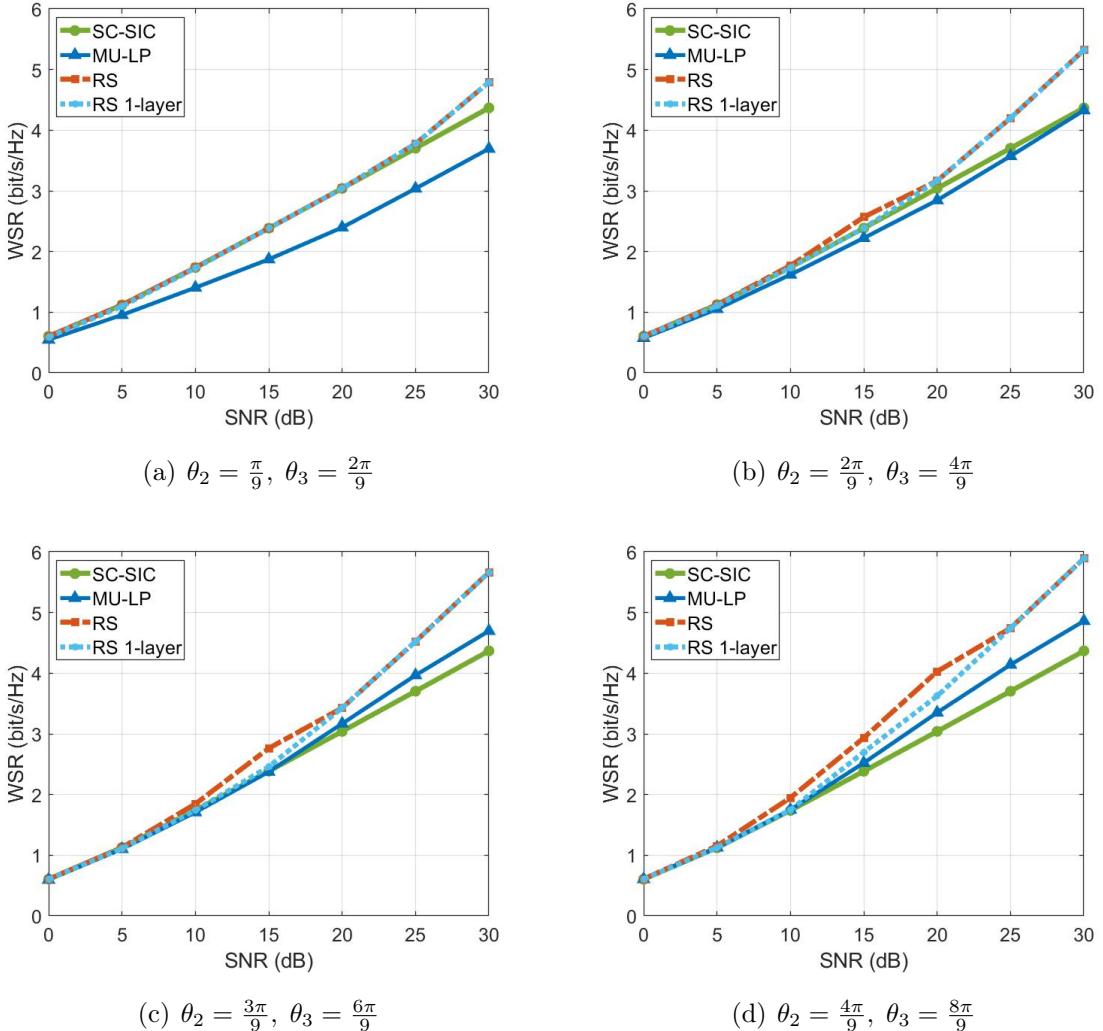


Figure 3.10: WSR vs SNR plot of SC-SIC based NOMA, MU-LP based SDMA, RSMA and 1-layer RSMA with perfect CSIT, $N_t = 2, K = 3, \mathbf{u} = [0.4, 0.3, 0.3]$, $\mathbf{r}_{th} = [0.02, 0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bit/s/Hz.

Fig.3.10 shows the rate performance for $\mathbf{u} = [0.4, 0.3, 0.3]$ with R_{th} for different schemes. Compared with previous results, in all subfigures of Fig.3.10, MU-LP has relatively worse performance. In (a)(b) SC-SIC outperforms MU-LP (channels are aligned), in (c)(d) SDMA has higher SR than NOMA (channels are orthogonal). This is the result of that MU-LP is not able to deal with multi-user interference effectively in such an overloaded network.

By examining the simulation results under different sorts of parameter setting, we further analyze the limitations of NOMA and SDMA; and in general, NOMA has the worst performance. RSMA is more robust to general network loads and user deployments, and is able to provide rate enhancement with and without QoS. In various simulations, the rate performance of 1-layer RS is almost as good as RS; in addition, 1-layer RS is easier to implement as only one SC-SIC layer is required for each user, we conclude that 1-layer is an efficient alternative of RS. Therefore, we will mainly compare 1-layer RS and SDMA in the following chapters.

Chapter 4

RSMA for CoMP-JT

In this chapter¹ we will discuss RSMA for CoMP-JT. We have examined single-cell setup where the interference from nearby cells are ignored, now we turn to the multi-cell scenario and the inter-cell interference is taken into account. According to [4], one of the best methods to handle inter-cell interference is letting the BSs in different cells work in a cooperative way such that the BSs are aware of the inter-cell interference. In 3rd Generation Partnership Project (3GPP), this is known as Coordinated Multi-Point Joint Transmission (CoMP-JT) [5]. Chapter 3 is used as the building block of RSMA for CoMP-JT, and additional features of CoMP-JT will be defined and investigated. I will first introduce the corresponding WMMSE-AO algorithm, and then examine the performance of 2-cell and 3-cell networks.

4.1 System Model

Let's consider a multi-cell SISO configuration where there are L cells, and in each cell- l ($l \in \mathcal{L} = \{1, 2, \dots, L\}$), the BS (each with $N_t = 1$ transmit antennas) serves K_l users (each with $N_r = 1$ receive antenna). The total number of users in L cells is $K = \sum_{l \in \mathcal{L}} K_l$. As mentioned earlier, the BSs work cooperatively, hence we can view the network as a 'super BS' (contains L 'transmit antennas' where each BS can be viewed as a 'transmit antenna') serves K users. It should be noted that other than the total power constraint which the 'super BS' has to meet, each individual base station- l is also subject to its own power constraint, which we will see later, is the major difference from the single-cell MISO network.

¹The main reference of this chapter is [6] unless otherwise specified.

The transmitted signal vector from the 'super BS' for a given user is $\mathbf{x} \in \mathbb{C}^{L \times 1}$, which is subject to the total as well as the individual BS transmit power constraint: $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq P_t, \mathbb{E}\{\|x_l\|^2\} \leq P_{t,l}, \forall l \in \mathcal{L}$. The channel between the 'super BS' and user- k is given by $\mathbf{h}_k \in \mathbb{C}^{1 \times L}, k \in \mathcal{K} = \{1, 2, \dots, K\}$. The received signal at user- k can be expressed as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \quad (4.1)$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is AWGN. CSIT and CSIR are also assumed to be perfect. As now there are multiple cells, the channel strength difference between each cell and each user needs to be considered. Therefore, we will implement the Wyner model and detailed explanations will be presented in Section 4.3.

4.2 SR Maximization and WMMSE-AO Algorithm

We use 1-layer RS in this chapter as such scheme introduces low complexity and maintain high performance at the same time. The mechanism for RS remains the same as in Chapter 3. The major difference is that now the individual BS power constraint needs to be considered. In addition, in baseline RS we have one BS with multiple transmit antennas whereas in RS for CoMP-JT we have one 'super BS' with multiple single-antenna BSs, thus each BS of the 'super BS' can be viewed as each transmit antenna of the BS in single-cell MISO BC.

The message intended for each user is split into two parts: $W_{c,k}$ and $W_k, \forall k \in \mathcal{K}$. The common message W_c consists all single-user common messages ($W_{c,k}$) and is encoded into common stream s_c , which will be decoded by all users. Private message W_k is encoded into private stream s_k , and will be decoded by user- k . After implementing linear precoding, the transmitted signal can be written as:

$$\mathbf{x} = \mathbf{Ps} = [\mathbf{p}_c, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K][s_c, s_1, s_2, \dots, s_K]^T \quad (4.2)$$

The common stream will firstly be decoded by viewing all private streams as noise. The SINR of the common stream at user- k is:

$$\eta_{c,k} = \frac{|\mathbf{h}_k \mathbf{p}_c|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (4.3)$$

After the common stream is decoded, the contribution of it will be subtracted from the received signal. Then user- k will decode its corresponding private stream by

treating the private streams of the other users as noise. At user- k , the SINR of the private stream is:

$$\eta_k = \frac{|\mathbf{h}_k \mathbf{p}_k|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (4.4)$$

The achievable rates for common stream s_c and private stream s_k of user- k are $R_{c,k} = \log_2 (1 + \eta_{c,k})$ and $R_k = \log_2 (1 + \eta_k)$ respectively. In order for the common message W_c to be decoded successfully by all users, the common achievable rate R_c should not exceed $R_c = \min\{R_{c,j}\}_{j=1}^K$. Assuming that the common achievable rate is shared by users: $R_c = \sum_{k \in \mathcal{K}} Z_{c,k}$, then the total achievable rate for user- k can be expressed as $R_{total,k} = Z_{c,k} + R_k$.

Following Eq.3.28, the WSR of 1-layer RS for CoMP-JT is:

$$\begin{aligned} \text{WSR}(\mathbf{u}) &= \max_{\mathbf{P}, \mathbf{z}} \sum_{k \in \mathcal{K}} u_k R_{total,k} \\ \text{s.t. } &\text{diag}(\mathbf{P} \mathbf{P}^H)_l \leq P_{t,l}, \forall l \in \mathcal{L} \\ &R_{total,k} \geq R_k^{th}, \forall k \in \mathcal{K} \\ &\sum_{k \in \mathcal{K}} Z_{c,k} \leq R_c \\ &\mathbf{z} \geq \mathbf{0} \end{aligned} \quad (4.5)$$

where R_k^{th} is the rate threshold and $\mathbf{z} = [Z_{c,1}, Z_{c,2}, \dots, Z_{c,K}]$ is the common achievable rate vector.

We now need to convert the WSR problem to the WMMSE problem. Following the discussions in Section 3.4, the received signal power X of user- k is given by:

$$\begin{aligned} X_{c,k} &= |\mathbf{h}_k \mathbf{p}_c|^2 + \sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1 \\ X_k &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1 \end{aligned} \quad (4.6)$$

And the noise power Y of user- k is:

$$\begin{aligned} Y_{c,k} &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1 \\ Y_k &= \sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k \mathbf{p}_j|^2 + 1 \end{aligned} \quad (4.7)$$

Refer to Eq.3.15, the MSE can be written as:

$$\begin{aligned} e_{c,k} &= |g_{c,k}|^2 X_{c,k} - 2\Re\{g_{c,k} \mathbf{h}_k \mathbf{p}_c\} + 1 \\ e_k &= |g_k|^2 X_k - 2\Re\{g_k \mathbf{h}_k \mathbf{p}_k\} + 1 \end{aligned} \quad (4.8)$$

From Eq.3.16, we can express the optimal MMSE combiner g as:

$$\begin{aligned}(g_{c,k})^{\text{MMSE}} &= \mathbf{p}_c^H \mathbf{h}_k^H / X_{c,k} \\ (g_k)^{\text{MMSE}} &= \mathbf{p}_k^H \mathbf{h}_k^H / X_k\end{aligned}\quad (4.9)$$

Refer to Eq.3.17 we can write the MMSE as:

$$\begin{aligned}(e_{c,k})^{\text{MMSE}} &= Y_{c,k} / X_{c,k} \\ (e_k)^{\text{MMSE}} &= Y_k / X_k\end{aligned}\quad (4.10)$$

From Eq.3.22, the optimum MMSE weights can be written as:

$$\begin{aligned}u_{c,k} &= 1/(e_{c,k})^{\text{MMSE}} \\ u_k &= 1/(e_k)^{\text{MMSE}}\end{aligned}\quad (4.11)$$

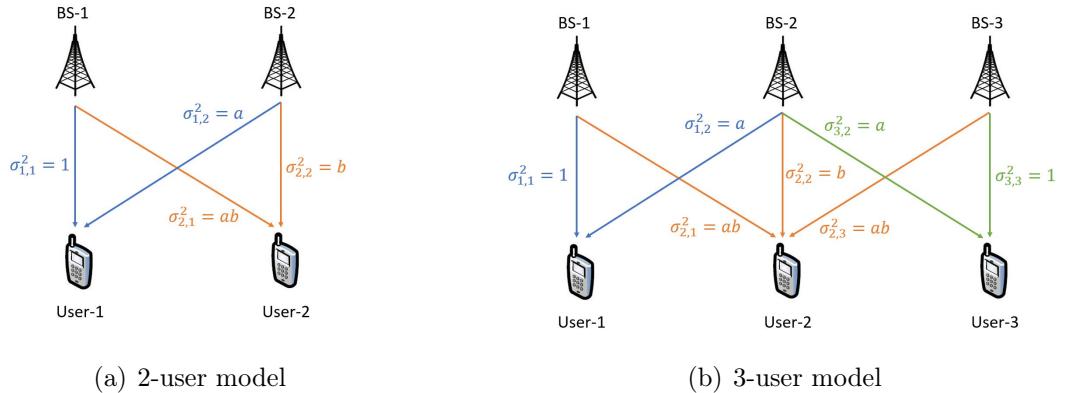
We can then convert the SR problem into WMMSE problem and solve it using AO Algorithm (Algorithm 1) shown in Section 3.4. The WMMSE problem is expressed as:

$$\begin{aligned}\text{WMMSE} &= \min_{\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u}} \sum_{k \in \mathcal{K}} u_k \phi_{total,k} \\ \text{s.t.} \quad &\text{diag}(\mathbf{P} \mathbf{P}^H)_l \leq P_{t,l}, \forall l \in \mathcal{L} \\ &\phi_{total,k} \leq 1 - R_k^{th}, \forall k \in \mathcal{K} \\ &\sum_{j \in \mathcal{K}} T_{c,j} \geq \phi_{c,k} - 1, \forall k \in \mathcal{K} \\ &\mathbf{t} \leq \mathbf{0}\end{aligned}\quad (4.12)$$

where \mathbf{t} is the common WMSE vector and $\phi_{total,k}$ is the individual WMSE for each user: $\phi_{total,k} = T_{c,k} + \phi_k, \forall k \in \mathcal{K}$.

4.3 Results and Discussions

The channel between cell- l and user- k is drawn from i.i.d. complex Gaussian variables with zero mean and variance $\sigma_{k,l}^2$: $h_{k,l} \sim \mathcal{CN}(0, \sigma_{k,l}^2)$. For each simulation, the result is based on the average of 100 random channel realizations. We use the linear Wyner model which is based on the symmetric Gaussian interference channels, meaning that the cells are positioned in a linear array and the interference is considered to be only from adjacent cells [4]. The channel strength disparity between BSs is $a \in (0, 1]$, the channel strength disparity between the users is $b \in (0, 1]$. Fig.4.1 demonstrate the Wyner model for 2-user and 3-user networks. Take 2-user network for example, for a given user, a is the ratio between the channel strength from the neighbour BS and the channel strength from the BS in the same cell with the user, b is the ratio between the channel strength of user-2 and the channel strength of user-1. We will implement such model in the following simulations.



(a) 2-user model

(b) 3-user model

 Figure 4.1: Wyner model, $L = 3, K_l = 1, l \in \{1, 2, 3\}$.

4.3.1 Two-Cell Network

We first consider the 2-cell network. The channel for user-1 is $\mathbf{h}_1 = [h_{1,1}, h_{1,2}]$ and the channel for user-2 is $\mathbf{h}_2 = [h_{2,1}, h_{2,2}]$. The common stream precoder is initialized as $\mathbf{p}_c = \gamma P_t \cdot \mathbf{v}_{12}$ where $\gamma \in [0, 1]$ and $\mathbf{v}_{12} = \mathbf{V}(:, 1)$ is the right singular vector corresponding to the maximum singular value of channel $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]^T$. The private stream precoders are initialized as $\mathbf{p}_k = \frac{(1-\gamma)P_t}{2} \cdot \frac{\mathbf{h}_k^H}{\|\mathbf{h}_k\|}, \forall k \in \{1, 2\}$. The power constraint is set as $P_{t,l} \leq \frac{P_{total}}{L}, \forall l \in \{1, 2\}$ where $L = 2$ and $P_{total} = 100$ Watt (baseline, SNR = 20dB).

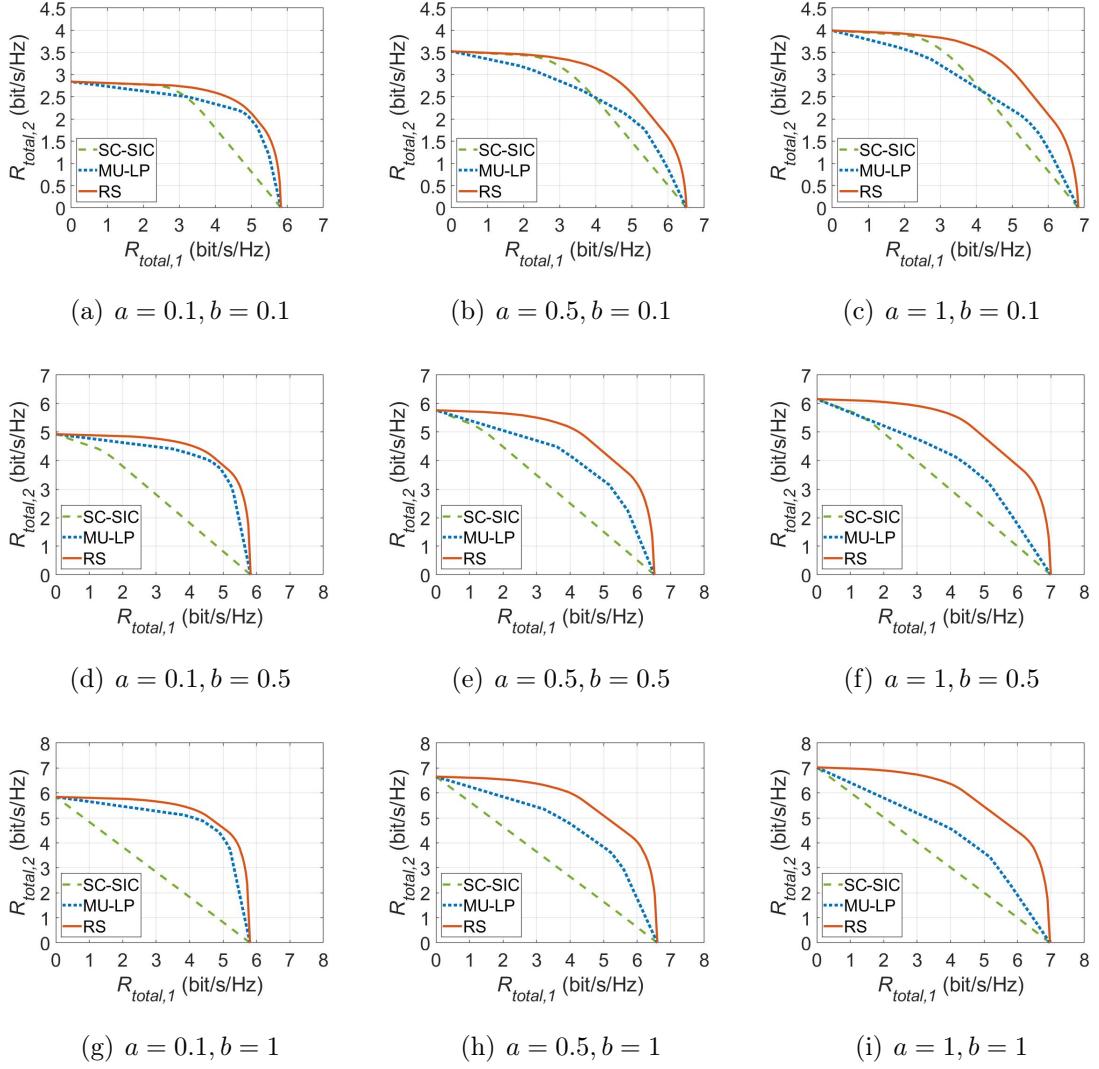


Figure 4.2: Achievable rate region of SC-SIC based NOMA, MU-LP based SDMA and RSMA for CoMP-JT with perfect CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}$, SNR = 20dB.

Fig.4.2 shows the rate region of different strategies under different inter-cell channel strength disparity a and inter-user channel strength disparity b . As we can see from the all subfigures, RS outperforms SC-SIC and MU-LP in different channel disparity settings. Through the cooperation between BSs, the network effectively becomes MISO BC (with individual power constraints). As a and b increases, the user channel strength increases, leading to larger rate region. As discussed before, NOMA works well when channel strengths are different and channels are aligned, while SDMA performs better for orthogonal channel with similar channel strengths. The same conclusion can be drawn here. Within each column of subfigures, the rate region difference between RSMA and NOMA gets bigger as b increases, this is

because NOMA is inefficient when channel strengths disparity is small (i.e. large b). Comparing subfigures within each row, the rate region gap between RSMA and SDMA grows as a increases, this is due to the fact that SDMA is more suitable for orthogonal channels, while big a indicates sufficient alignment [6]. In summary, NOMA works better at big a and small b (e.g. Fig4.2(c)), SDMA is efficient when a is small and b is large (e.g. Fig4.2(g)). RS is more robust and outperforms the other two in different network settings and user deployments.

4.3.2 Three-Cell Network

Fig.4.3 demonstrates the SR of SDMA and 1-layer RSMA under rate threshold constraints $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, 0.3]$ bits/s/Hz for SNR = [0, 5, 10, 15, 20, 25, 30] dB. The QoS is considered here as the channel strength is sometimes weak under certain a and b values, and possibly some user will never be served if no QoS is concerned. RSMA outperforms SDMA in all subfigures, and the SR increases with a and b . The rate improvement of RS is not huge in Fig.4.3, but as we will see later, using RS is more beneficial in CoMP-LT networks when CSIT is not perfect. We can again conclude that RS is able to deliver higher rates and is more robust to general networks.

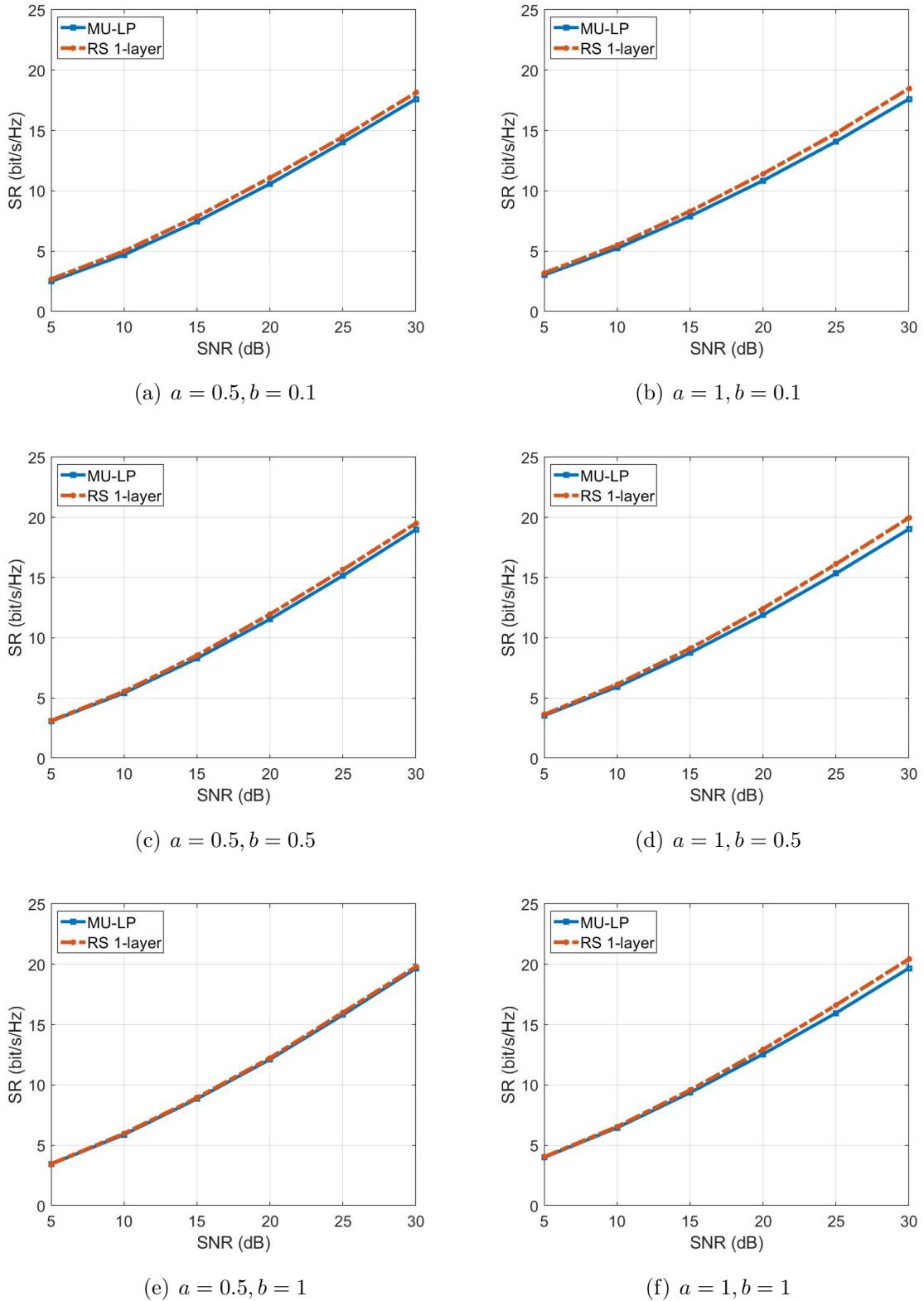


Figure 4.3: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with perfect CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, 0.3]$ bits/s/Hz.

Chapter 5

RSMA with Partial CSIT

Having examined the case where perfect CSIT is available, we now move onto the imperfect CSIT scenario for MISO BC. The WMMSE-AO algorithm will still be used for rate maximization, but before implementing such algorithm, the stochastic Average Sum-Rate (ASR) problem will be firstly converted into a deterministic problem using Sample Average Approximation (SAA) which is based on the statistical knowledge of CSIT error. In this chapter¹, WMMSE-AO with SAA algorithm will be explained, and relevant result will be presented.

5.1 System Model

Let's consider a MISO configuration where a BS (with N_t transmit antennas) serves K users (each with $N_r = 1$ receive antenna) in the cell. The interference from other nearby cells are ignored. The transmitted signal vector for a given user channel is $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$, which is subject to the total transmit power constraint: $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq P_t$. The channel between BS and user- k is given by $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}, k \in \mathcal{K} = \{1, 2, \dots, K\}$. The received signal at user- k can be expressed as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \quad (5.1)$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is AWGN. CSIR is assumed to be perfect.

5.2 ASR Maximization

In this chapter, we will use 1-layer RS due to its low complexity. The message intended for each user is split into two parts: $W_{c,k}$ and $W_k, \forall k \in \mathcal{K}$. The common

¹The main reference of this chapter is [2] unless otherwise specified.

message W_c consists all single-user common messages ($W_{c,k}$) and is encoded into common stream s_c . Private message W_k is encoded into private stream $s_k, k \in \mathcal{K}$. After implementing linear precoding, the transmitted signal can be written as:

$$\mathbf{x} = \mathbf{Ps} = [\mathbf{p}_c, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K][s_c, s_1, s_2, \dots, s_K]^T \quad (5.2)$$

The common stream will firstly be decoded by all users by viewing all private streams as noise. The SINR of the common stream at user- k is:

$$\eta_{c,k} = \frac{|\mathbf{h}_k \mathbf{p}_c|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (5.3)$$

After the common stream is decoded, the contribution of it will be subtracted from the received signal. Then user- k will decode its corresponding private stream by treating the private streams of the other users as noise. At user- k , the SINR of the private stream is:

$$\eta_k = \frac{|\mathbf{h}_k \mathbf{p}_k|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (5.4)$$

The achievable rates for common stream s_c and private stream s_k of user- k are $R_{c,k} = \log_2 (1 + \eta_{c,k})$ and $R_k = \log_2 (1 + \eta_k)$ respectively. In order for the common message W_c to be decoded successfully by all users, the common achievable rate R_c should not exceed $R_c = \min\{R_{c,j}\}_{j=1}^K$. The SR is then given by $R_s = R_c + \sum_{k \in \mathcal{K}} R_k$. Assuming that the common achievable rate is shared by users: $R_c = \sum_{k \in \mathcal{K}} Z_{c,k}$, then the total achievable rate for user- k can then be expressed as $R_{total,k} = Z_{c,k} + R_k$, and the SR can be also expressed as $R_s = \sum_{k \in \mathcal{K}} R_{total,k}$.

Now let's have a more detailed look at CSIT. We can write the fading channel as $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]^T$, which is available at the receiver (perfect CSIR). At the BS end, such knowledge of channel is not available, and the BS only have an estimate of instantaneous channel state $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_K]^T$. In addition, for a given channel estimate, the channel estimation error matrix is given by $\mathbf{H}_e = [\mathbf{h}_{e,1}, \mathbf{h}_{e,2}, \dots, \mathbf{h}_{e,K}]^T$. Therefore, we have $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_e$, and the error introduced by the estimation can be characterized by the conditional probability density function $p_{H|\hat{H}}(\mathbf{H}|\hat{\mathbf{H}})$. For each user, the marginal probability density function can be written as $p_{h_k|\hat{h}_k}(\mathbf{h}_k|\hat{\mathbf{h}}_k)$. The mean of the probability distribution is the estimate, meaning $\mathbb{E}_{h_k|\hat{h}_k}\{\mathbf{h}_k|\hat{\mathbf{h}}_k\} = \hat{\mathbf{h}}_k$. Denote $\mathbf{R}_{e,k}$ as the CSIT error covariance matrix of user- k (assume $\mathbf{R}_{e,k}$ and $\hat{\mathbf{h}}_k$ are independent), we have $\mathbb{E}_{h_k|\hat{h}_k}\{\mathbf{h}_k \mathbf{h}_k^H|\hat{\mathbf{h}}_k\} = \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{R}_{e,k}$. The CSIT error power at user- k is given by: $\sigma_{e,k}^2 = \text{tr}(\mathbf{R}_{e,k})$, with $\sigma_{e,k}^2 = O(\text{SNR}^{-\alpha})$ where α is a non-negative constant. The constant α is quality scaling exponent and is defined as:

$\alpha = -\lim_{\text{SNR} \rightarrow \infty} \frac{\log(\sigma_{e,k}^2)}{\log(\text{SNR})}$, which indicates the quality of CSIT as SNR increases. Take two extreme cases for example: $\alpha \rightarrow \infty$ makes CSIT error power $\sigma_{e,k}^2 \rightarrow 0$ which indicates CSIT is perfect; $\alpha = 0$ means the CSIT error power is at a fixed level. When $\alpha \neq 0$, CSIT error power decreases as SNR increases.

Let's now consider SR maximization for RS when CSIT is imperfect. Again, we need to find the optimal precoder to maximize the SR. A naive method is to design a precoder using $\hat{\mathbf{H}}$ directly assuming it's perfect. However, this would not work as it will not be able to handle MU interference. More importantly, the BS is unaware of the imperfection of CSIT. The BS may overestimate the achievable rate and which will eventually lead to a transmission with undecodable rates. Therefore, we need to design a robust method which is able to cope with partial CSIT and also delivers rate decodability.

Although perfect CSIT is not available, given the channel estimate and channel error distribution, the average rates is available to the BS. Determined by the channel estimate, the average rate is defined as : $\bar{R}_k = \mathbb{E}_{H|\hat{H}}\{R_k|\hat{\mathbf{H}}\}, \forall k \in \mathcal{K}$. It's important to note that the average rate or Average Sum-Rate (ASR) here is a short-term measure, i.e. for a given channel estimate, the expectation of the rates over the CSIT error distribution. By averaging the ASRs over the variation of channel estimate, the achievable rate we get is called the Ergodic Sum-Rate (ESR)², which will be produced through simulations and be analyzed in Section 5.4. Therefore, the ASR optimization problem can be written as:

$$\begin{aligned} \text{ASR} &= \max_{\mathbf{P}, \bar{R}_c} \left\{ \bar{R}_c + \sum_{k \in \mathcal{K}} \bar{R}_k \right\} \\ \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\ &\bar{R}_{c,k} \geq \bar{R}_c, \forall k \in \mathcal{K} \end{aligned} \tag{5.5}$$

If we want to have some degree of fairness between different users, Eq.5.5 can be reformulated based on the methodology of forming the SR optimization problem in

²A more detailed explanation of ASR and ESR can be found in [2].

[1], and can be expressed as:

$$\begin{aligned}
 \text{ASR} &= \max_{\mathbf{P}, \mathbf{z}} \sum_{k \in \mathcal{K}} \bar{R}_{total,k} \\
 \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\
 &\bar{R}_{total,k} \geq R_k^{th}, \forall k \in \mathcal{K} \\
 &\sum_{k \in \mathcal{K}} \bar{Z}_{c,k} \leq \bar{R}_{c,k}, \forall k \in \mathcal{K} \\
 &\mathbf{z} \geq \mathbf{0}
 \end{aligned} \tag{5.6}$$

where $\mathbf{z} = [\bar{Z}_{c,1}, \bar{Z}_{c,2}, \dots, \bar{Z}_{c,K}]$ is the average common rate vector and $\bar{R}_{total,k} = \bar{Z}_{c,k} + \bar{R}_k, \forall k \in \mathcal{K}$. The ASR maximization problem can be solved by using WMMSE-AO with SAA algorithm in the next section.

5.3 WMMSE-AO with SAA algorithm

The logic of the algorithm follows Section 3.4 but with SAA incorporated in it. As Eq.5.5 is stochastic, we need to convert it to a deterministic problem using SAA, and then transform the deterministic SR problem to a WMMSE problem.

Let's first start with SAA. For a given channel estimate $\hat{\mathbf{H}}$, define a sample of N i.i.d realizations which is drawn from the conditional distribution with probability density function $p_{H|\hat{H}}(\mathbf{H}|\hat{\mathbf{H}})$:

$$\mathbb{H}^{(N)} = \{\mathbf{H}^{(n)} = \hat{\mathbf{H}} + \mathbf{H}_e^{(n)}, n \in \mathcal{N}\} \tag{5.7}$$

where $\mathcal{N} = \{1, 2, \dots, N\}$. Such sample is introduced so that we can approximate the average rate using Sample Average Function (SAF). SAFs of user- k are defined as: $\bar{R}_{c,k}^{(N)} = \frac{1}{N} \sum_{n \in \mathcal{N}} R_{c,k}^{(n)}$ and $\bar{R}_k^{(N)} = \frac{1}{N} \sum_{n \in \mathcal{N}} R_k^{(n)}$, where $R_{c,k}^{(n)}$ and $R_k^{(n)}$ are the rates for the n th realization. Therefore, SAA of Eq.5.5 can be written as:

$$\begin{aligned}
 \text{ASR}^{(N)} &= \max_{\mathbf{P}, \bar{R}_c} \left\{ \bar{R}_c + \sum_{k \in \mathcal{K}} \bar{R}_k^{(N)} \right\} \\
 \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\
 &\bar{R}_{c,k}^{(N)} \geq \bar{R}_c, \forall k \in \mathcal{K}
 \end{aligned} \tag{5.8}$$

From the strong Law of Large Numbers, we have:

$$\begin{aligned}\lim_{N \rightarrow \infty} \bar{R}_{c,k}^{(N)}(\mathbf{P}) &= \bar{R}_{c,k}(\mathbf{P}), \text{ almost surely } \forall \mathbf{P} \in \mathbb{P} \\ \lim_{N \rightarrow \infty} \bar{R}_k^{(N)}(\mathbf{P}) &= \bar{R}_k(\mathbf{P}), \text{ almost surely } \forall \mathbf{P} \in \mathbb{P}\end{aligned}\tag{5.9}$$

where \mathbb{P} is a set of possible precoders³. Therefore, by using SAA, we can say that the stochastic ASR problem (Eq.5.5) is converted into a deterministic problem (Eq.5.8) when $N \rightarrow \infty$ ⁴.

We now need to convert the deterministic ASR problem into the WMMSE problem. Following the discussions in Section 3.4, the received signal power X of user- k ($k \in \mathcal{K} = \{1, 2, \dots, K\}$) is given by:

$$\begin{aligned}X_{c,k}^{(n)} &= |\mathbf{h}_k^{(n)} \mathbf{p}_c|^2 + \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1 \\ X_k^{(n)} &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1\end{aligned}\tag{5.10}$$

And the noise power Y of user- k is:

$$\begin{aligned}Y_{c,k}^{(n)} &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1 \\ Y_k^{(n)} &= \sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1\end{aligned}\tag{5.11}$$

Refer to Eq.3.15, the MSE can be written as:

$$\begin{aligned}e_{c,k}^{(n)} &= |g_{c,k}^{(n)}|^2 X_{c,k}^{(n)} - 2\Re\{g_{c,k}^{(n)} \mathbf{h}_k^{(n)} \mathbf{p}_c\} + 1 \\ e_k^{(n)} &= |g_k^{(n)}|^2 X_k^{(n)} - 2\Re\{g_k^{(n)} \mathbf{h}_k^{(n)} \mathbf{p}_k\} + 1\end{aligned}\tag{5.12}$$

From Eq.3.16, we can express the optimal MMSE combiner g as:

$$\begin{aligned}(g_{c,k}^{(n)})^{\text{MMSE}} &= \mathbf{p}_c^H \mathbf{h}_k^{(n)H} / X_{c,k}^{(n)} \\ (g_k^{(n)})^{\text{MMSE}} &= \mathbf{p}_k^H \mathbf{h}_k^{(n)H} / X_k^{(n)}\end{aligned}\tag{5.13}$$

Refer to Eq.3.17 we can write the MMSE as:

$$\begin{aligned}(e_{c,k}^{(n)})^{\text{MMSE}} &= Y_{c,k}^{(n)} / X_{c,k}^{(n)} \\ (e_k^{(n)})^{\text{MMSE}} &= Y_k^{(n)} / X_k^{(n)}\end{aligned}\tag{5.14}$$

³The precoder is fixed over N channel error realizations for a given channel estimate.

⁴Detailed proof can be found in [2].

From Eq.3.22, the optimum MMSE weights can be written as:

$$\begin{aligned} u_{c,k}^{(n)} &= 1/(e_{c,k}^{(n)})^{\text{MMSE}} \\ u_k^{(n)} &= 1/(e_k^{(n)})^{\text{MMSE}} \end{aligned} \quad (5.15)$$

So far we have calculated different parameters for one channel realization given a fixed channel estimate, with the channel error drawn from the error distribution. In order to implement SAA, we need to define a few new SAFs, which will be used to update the precoder using AO algorithm. The SAFs is defined as: $\bar{u}_{c,k}, \bar{u}_k, \bar{r}_{c,k}, \bar{r}_k, \bar{\mathbf{f}}_{c,k}, \bar{\mathbf{f}}_k, \bar{\mathbf{v}}_{c,k}, \bar{\mathbf{v}}_k, \bar{w}_{c,k}, \bar{w}_k$. Take $\bar{u}_{c,k}$ and \bar{u}_k for example, such SAFs can be obtained by calculating the ensemble average of Eq.5.15 over N realizations. Other SAFs are obtained in the same way by calculating the ensemble average of the following equations:

$$\begin{aligned} r_{c,k}^{(n)} &= u_{c,k}^{(n)} |g_{c,k}^{(n)}|^2 \\ r_k^{(n)} &= u_k^{(n)} |g_k^{(n)}|^2 \end{aligned} \quad (5.16)$$

$$\begin{aligned} \mathbf{f}_{c,k}^{(n)} &= r_{c,k}^{(n)} \mathbf{h}_k^{(n)H} \mathbf{h}_k^{(n)} \\ \mathbf{f}_k^{(n)} &= r_k^{(n)} \mathbf{h}_k^{(n)H} \mathbf{h}_k^{(n)} \end{aligned} \quad (5.17)$$

$$\begin{aligned} \mathbf{v}_{c,k}^{(n)} &= u_{c,k}^{(n)} \mathbf{h}_k^{(n)H} g_k^{(n)H} \\ \mathbf{v}_k^{(n)} &= u_k^{(n)} \mathbf{h}_k^{(n)H} g_k^{(n)H} \end{aligned} \quad (5.18)$$

$$\begin{aligned} w_{c,k}^{(n)} &= \log_2(u_{c,k}^{(n)}) \\ w_k^{(n)} &= \log_2(u_k^{(n)}) \end{aligned} \quad (5.19)$$

By using above SAFs, we can re-write the augmented WMSE in Eq.3.20 into average WMSE:

$$\begin{aligned} \bar{\phi}_{c,k}^{(N)} &= \mathbf{p}_c^H \bar{\mathbf{f}}_{c,k} \mathbf{p}_c + \sum_{j=1}^K \mathbf{p}_j^H \bar{\mathbf{f}}_{c,k} \mathbf{p}_j + \sigma_n^2 \bar{r}_{c,k} - 2\Re\{\bar{\mathbf{v}}_{c,k}^H \mathbf{p}_c\} + \bar{u}_{c,k} - \bar{w}_{c,k} \\ \bar{\phi}_k^{(N)} &= \sum_{j=1}^K \mathbf{p}_j^H \bar{\mathbf{f}}_k \mathbf{p}_j + \sigma_n^2 \bar{r}_k - 2\Re\{\bar{\mathbf{v}}_k^H \mathbf{p}_k\} + \bar{u}_k - \bar{w}_k \end{aligned} \quad (5.20)$$

Similar to Eq.3.23, the WMMSE-Rate relationship can be written as $\bar{\phi}_{c,k}^{(N)} = 1 - \bar{R}_{c,k}^{(N)}$, $\bar{\phi}_k^{(N)} = 1 - \bar{R}_k^{(N)}$. Then the ASR problem in Eq.5.8 can be transformed into

WMMSE optimization which is written as:

$$\begin{aligned} \text{WMMSE} &= \min_{\mathbf{P}, \bar{\phi}_c, \mathbf{g}, \mathbf{u}} \left\{ \bar{\phi}_c + \sum_{k \in \mathcal{K}} \bar{\phi}_k^{(N)} \right\} \\ \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\ &\bar{\phi}_{c,k}^{(N)} \leq \bar{\phi}_c, \forall k \in \mathcal{K} \end{aligned} \quad (5.21)$$

If we want to ensure some user fairness, using the methodology in [1], we can re-write the WMMSE problem as:

$$\begin{aligned} \text{WMMSE} &= \min_{\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u}} \sum_{k \in \mathcal{K}} \bar{\phi}_{total,k} \\ \text{s.t. } &\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \\ &\bar{\phi}_{total,k} \leq 1 - R_k^{th}, \forall k \in \mathcal{K} \\ &\sum_{j \in \mathcal{K}} \bar{T}_{c,j} \geq \bar{\phi}_{c,k}^{(N)} - 1, \forall k \in \mathcal{K} \\ &\mathbf{t} \leq \mathbf{0} \end{aligned} \quad (5.22)$$

where \mathbf{t} is the common WMSE vector and $\bar{\phi}_{total,k}$ is the individual WMSE for each user: $\bar{\phi}_{total,k} = \bar{T}_{c,k} + \bar{\phi}_k^{(N)}, \forall k \in \mathcal{K}$. Similar to Algorithm 1, we can use the AO algorithm below to solve Eq.5.22.

Algorithm 2: Alternating Optimization Algorithm (For RSMA with partial CSIT)

Initialize: $m \leftarrow 0$, $\mathbf{P}^{[m]}$, $\text{WMMSE}^{[m]}$

repeat

$m \leftarrow m + 1$, $\mathbf{P}^{[m-1]} \leftarrow \mathbf{P}$;

$\mathbf{g} \leftarrow \mathbf{g}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$, $\mathbf{u} \leftarrow \mathbf{u}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$;

update SAFs: $\bar{u}_{c,k}, \bar{u}_k, \bar{r}_{c,k}, \bar{r}_k, \bar{\mathbf{f}}_{c,k}, \bar{\mathbf{f}}_k, \bar{\mathbf{v}}_{c,k}, \bar{\mathbf{v}}_k, \bar{w}_{c,k}, \bar{w}_k$ using updated \mathbf{g} and \mathbf{u} ;

update (\mathbf{P}, \mathbf{t}) by solving Eq.5.22 using updated SAFs;

until $|\text{WMMSE}^{[m]} - \text{WMMSE}^{[m-1]}| \leq \epsilon$;

5.4 Results and Discussions

The initialization process follows Section 3.6. One difference is that the proportion of the total power allocated to the common message is $P_t - P_t^\alpha$ and the power allocated to each private message is $\frac{P_t^\alpha}{K}$. We choose the SAA sample size⁵ as $N = 100$. The simulation result is averaged over 100 random channel estimate realizations with $\hat{\mathbf{h}}_k \sim \mathcal{CN}(0, \sigma_k^2)$, $\sigma_k^2 = 1, \forall k \in \mathcal{K}$. For a given realization, we define the channel as: $\mathbf{H}^{(n)} = \sqrt{1 - \sigma_e^2} \hat{\mathbf{H}} + \sqrt{\sigma_e^2} \mathbf{H}_e^{(n)}$ where σ_e^2 is the CSIT error power which is assumed to be the same for every user. We will vary the quality scaling exponent α and evaluate the performance of RSMA and SDMA⁶.

5.4.1 Two-User Network

Fig.5.1 shows the rate region of SDMA and RSMA for $\alpha = 0.6$ and $\alpha = 0.9$. The RS rate region is larger than the MU-LP rate region, and both rate regions grow as α increases (i.e. better CSIT quality). Comparing with Fig.3.3(c), the corresponding rate regions in Fig.5.1 are smaller due to imperfect CSIT. We will investigate more on the quality scaling exponent α by analyzing the 3-user case in the next subsection.

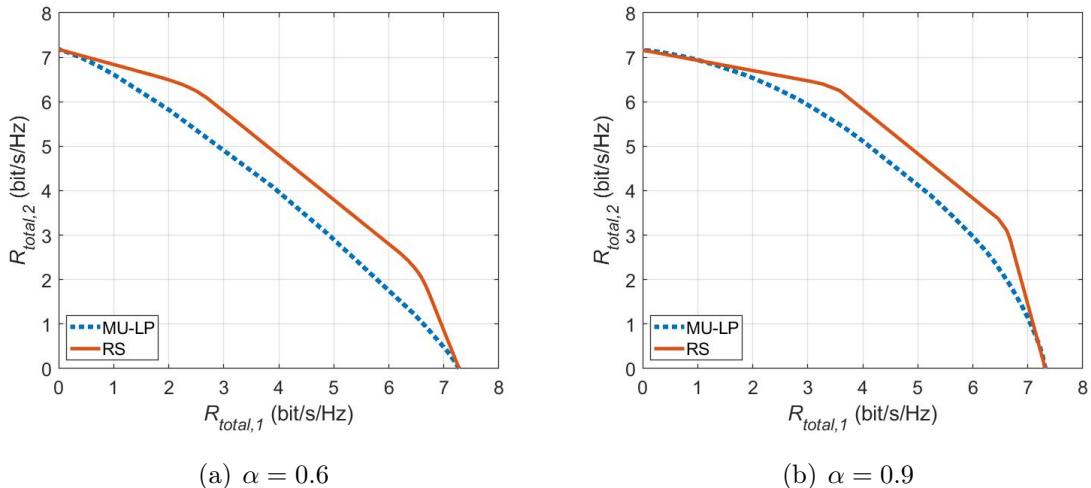


Figure 5.1: Achievable rate region of MU-LP based SDMA and RSMA with partial CSIT, $N_t = 2, K = 2, \sigma_e^2 = P_t^{-\alpha}$, SNR = 20dB.

⁵In [2], the author discussed the selection of N , and $N = 100$ can be viewed as a large enough sample size which doesn't lead to unsatisfactory results.

⁶In order to have a fair comparison, SAA is also combined with SDMA. Such scheme is implemented by assigning zero power to the common message of RS (RSMA then boils down to SDMA).

5.4.2 Three-User Network

In this subsection we will focus on the effects of both quality scaling exponent and QoS requirements. Fig.5.2 demonstrates the SR of SDMA and RSMA with partial CSIT when no QoS is considered. As we discussed in Section 2.4, DoF of MU-LP is $\max\{1, K\alpha\}$ and DoF of RS is $1 + (K - 1)\alpha$, thus for $\alpha = [0.1, 0.3, 0.6, 0.9]$, the DoF difference is 0.2, 0.6, 0.4, 0.1 respectively. This also agrees with the results in Fig.5.2: in (b), RS outperforms MU-LP the most; in (a)(d), RS and MU-LP have the similar performance⁷. We take a further look at Fig.5.2(c): at low SNR, the weakest user(s) is turned off and both schemes have very similar SR as they both reduce to single-user transmission. As SNR grows, MU transmission starts; however, the additional power may be inaccurately distributed due to partial CSIT, which may enhance the interference (in SDMA interference is considered as noise). Therefore, at relatively high SNR, RS starts to have an higher rate than MU-LP as the common message of RS starts to contribute by decoding the interference.

Fig.5.3 demonstrates the SR of RS and MU-LP for different α given some QoS. Comparing Fig.5.3 with Fig.5.2, it can be seen that the gap between RSMA and SDMA is increased when certain QoS is considered. As mentioned in Section 3.6, the advantages of the common message of RS is more explicit when QoS is taken into account. When the CSIT quality is poor (e.g. Fig.5.3(a)(b)), using RSMA is even more beneficial than SDMA. This means that using RS can relax the requirements on CSIT quality for achieving certain performance⁸ under QoS requirements.

⁷Note that as SNR is finite, the slope difference in the plot at high SNR region is not the same as theoretical DoF value but quite close; we will discuss DoF more in the next subsection.

⁸It's intuitive to say that if RSMA outperforms SDMA under the same CSIT quality, then in order to achieve the same performance, SDMA needs better CSIT quality than RSMA. One way to validate this is by solving $K\alpha_{\text{MU-LP}} = 1 + (K - 1)\alpha_{\text{RS}}$, we can set $\alpha_{\text{MU-LP}} = \frac{1+(K-1)\alpha_{\text{RS}}}{K}$ (assume $\alpha_{\text{MU-LP}} > \frac{1}{K}$). The SR plots of SDMA and RSMA are expected to coincide or very close to each other under such α settings, relevant results in [2] show this is indeed the case (without R_{th}).

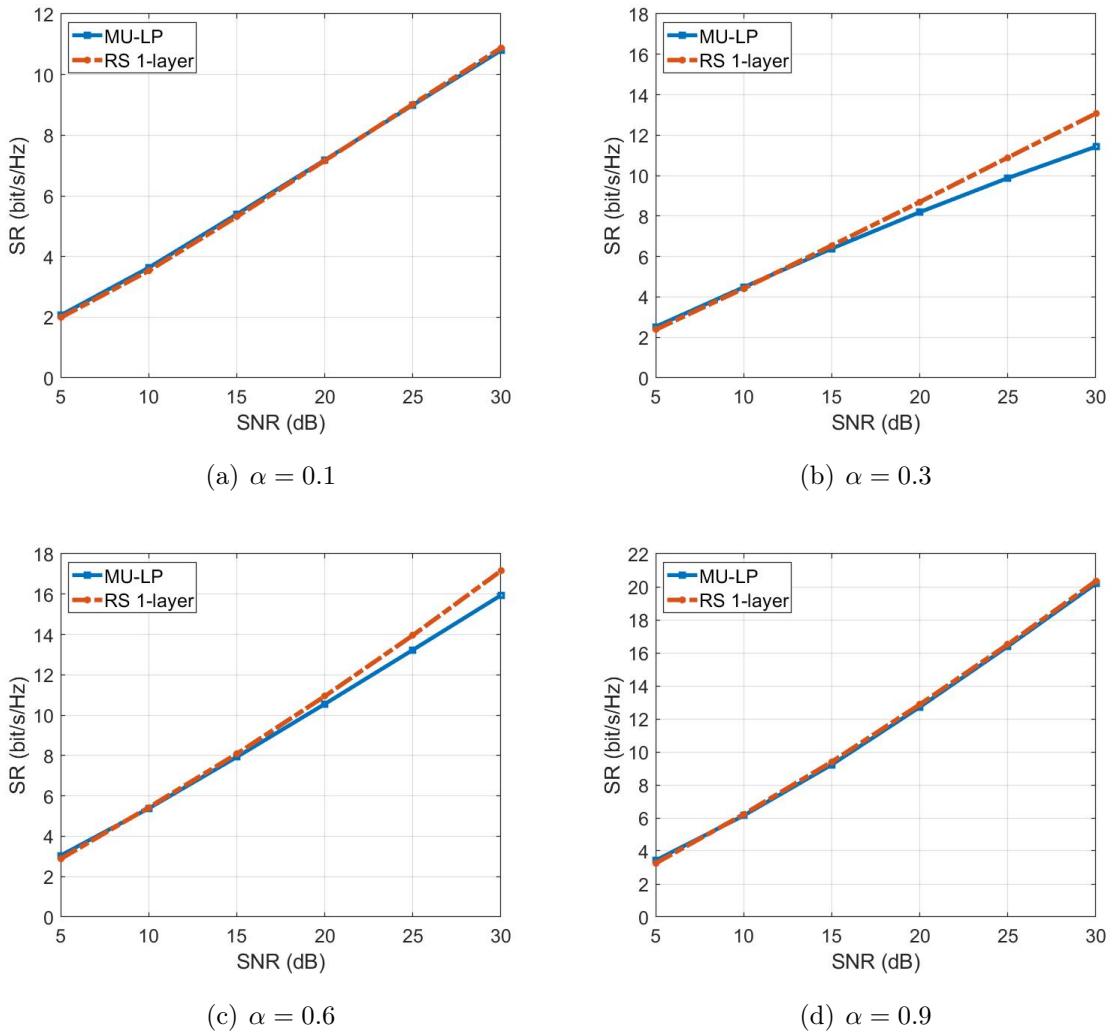


Figure 5.2: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 3, K = 3, \sigma_e^2 = P_t^{-\alpha}, \mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz.

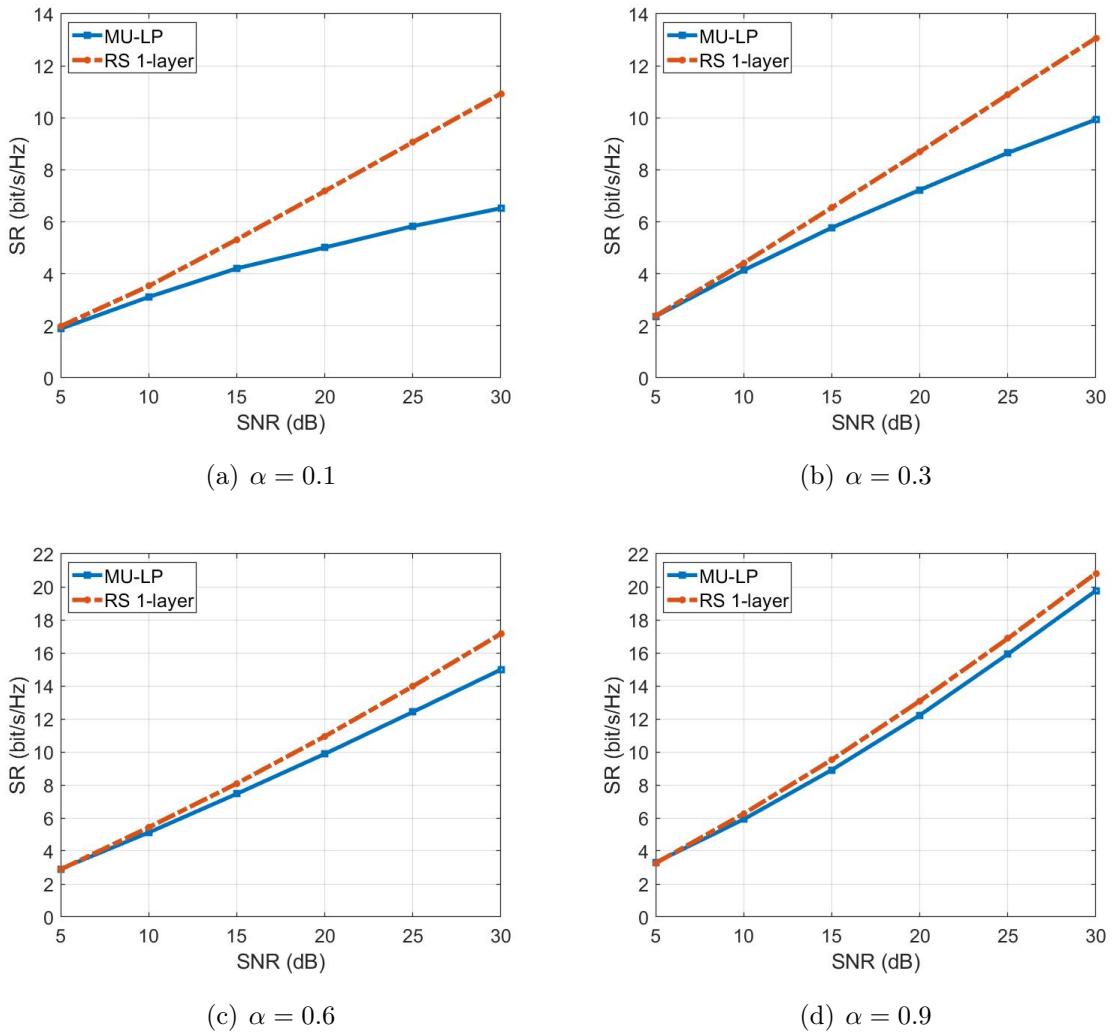


Figure 5.3: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 3, K = 3, \sigma_e^2 = P_t^{-\alpha}, \mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, 0.3]$ bit/s/Hz.

5.4.3 DoF Analysis

In the previous section we discussed the DoF but didn't verify the simulations results with the theoretical values. As mentioned earlier, because the SNR is finite, the slope obtained using simulations for up to 30dB is able to capture the features of DoF, but not precise enough to represent the accurate DoF value. Here the SR maximization problem is simulated for a 2-user network using higher SNRs (up to 45dB). It should be noted that DoF is a theoretical value when SNR tends to infinity; in the real-world deployment, SNR mainly lies in the range between -5 and 25dB [35].

Fig.5.4 demonstrates the SR versus SNR plot for 2-user case for different α with and without R_{th} . When $\alpha = 0.3/0.6/0.9$, the DoF of SDMA for 2-user network is $\max\{1, K\alpha\} = 1/1.2/1.8$, and the DoF of RSMA for 2-user network is $1 + (k - 1)\alpha = 1.3/1.6/1.9$, which can be also obtained⁹ from the subfigure (a)(c)(e) (without R_{th}). If certain QoS is considered, when CSIT quality is relatively good ($\alpha = 0.6/0.9$), in Fig.5.4(d)(f) we can see that the DoF is the same as that under no R_{th} . This is because when SNR is as high as 45dB, there are enough power allocated to each user; therefore, the rate with value of R_{th} can be achieved anyway with or without QoS requirements. However in (b), with $\alpha = 0.3$ and R_{th} , even at high SNR the slope of the SDMA curve is only half of the DoF value, while the slope of the RS plot is the same as DoF (thanks to the common message). Furthermore, comparing Fig.5.4(c) and (d), although the slope of corresponding SDMA/RSMA plots are the same, the SR under R_{th} is a little lower than the SR without R_{th} . This is caused by the fact that QoS is a limiting factor of SR at low SNR, hence there is this 'offset' between the with- R_{th} and without- R_{th} plots for both RS and MU-LP. In summary, under partial CSIT, RS outperforms MU-LP not only in terms of SR and QoS enhancement, but also from DoF perspective.

⁹In the result sections of Chapter 5 and 6, the slopes obtained from the data points shown in the figures may not be exactly the same as DoF (but very close). Such error may be caused by three reasons: SNR is finite, SAA sample size is finite and the number of random channel realizations is finite.

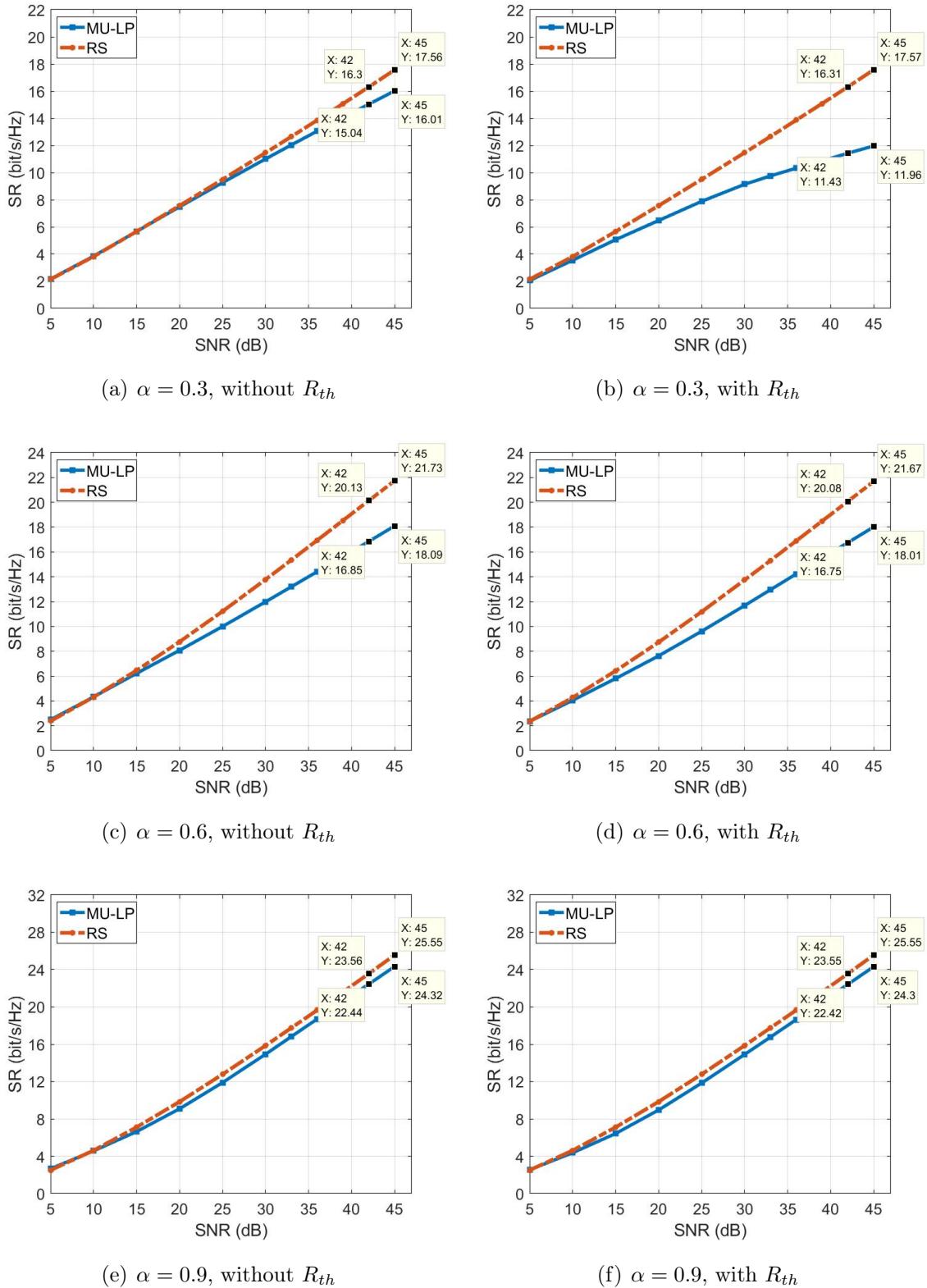


Figure 5.4: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 2, K = 2, \sigma_e^2 = P_t^{-\alpha}$. (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, \dots, 0.3]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.

We now turn to the overloaded network. Fig.5.5 demonstrates the SR of SDMA and RSMA with two transmit antennas and four single-antenna users under partial CSIT. From the DoF equation of underloaded NoRS schemes in Section 2.4, we can further express the DoF of NoRS in the overloaded networks as: $\max\{1, \alpha N_t\}$. Hence in general network loads, NoRS DoF can be written as: $\max\{1, \min\{N_t, K\} \cdot \alpha\}$ [2][35]. The theoretical proof of DoF for RS in the overloaded MISO networks is still unknown; from the figure it seems that the overloaded RS DoF is slightly smaller than its counterpart in the underloaded network. Nevertheless, through simulation results we show that RSMA outperforms SDMA regarding both SR and DoF in overloaded networks with partial CSIT. One major difference between Fig.5.4 and Fig.5.5 is that when certain QoS is considered, the performance of SDMA in the overloaded case is poor as shown in Fig.5.5(b)(d)(f). This is the direct result of the incapability of MU-LP in overloaded networks as we discussed before.

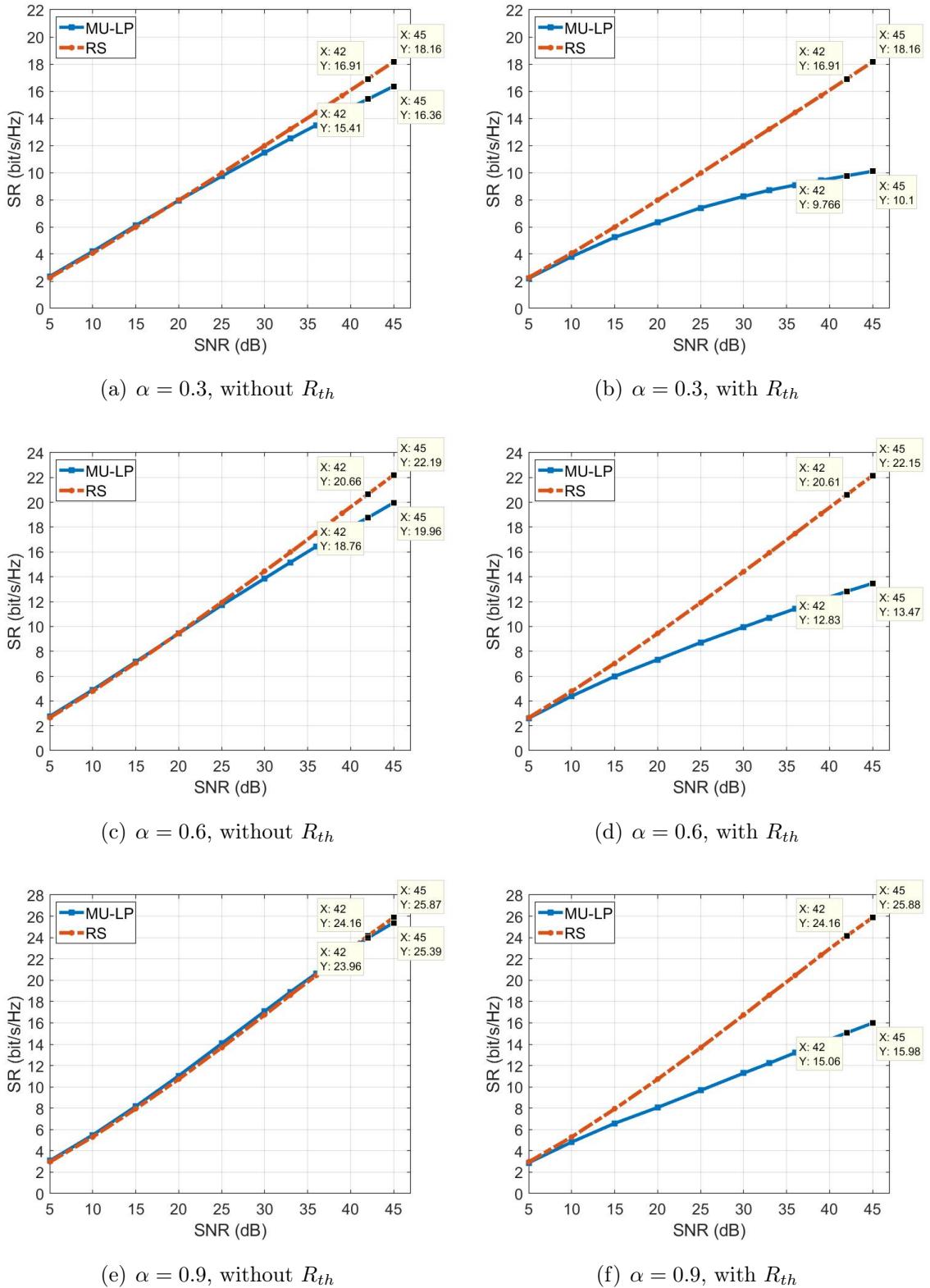


Figure 5.5: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA with partial CSIT, $N_t = 2, K = 4, \sigma_e^2 = P_t^{-\alpha}$. (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.12, 0.18, 0.24, 0.3, 0.3, \dots, 0.3]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.

Chapter 6

RSMA for CoMP-JT with Partial CSIT

Having investigated RS with partial CSIT for single-cell MISO BC, we now move onto RS with partial CSIT for CoMP-JT network, which hasn't been proposed in the literature. Chapter 4 and 5 are used as the building blocks of RSMA for CoMP-JT with partial CSIT; similar to Chapter 3/4, the major difference between Chapter 5 and Chapter 6 is that each base station is subject to its own power constraint, and the inter-cell and inter-user channel strength disparities are also need to be considered. I will first introduce the corresponding WMMSE-AO with SAA algorithm, and then analyze the performance of 2-cell and 3-cell networks.

6.1 System Model

Let's consider a multi-cell SISO configuration where there are L cells, and in each cell- l ($l \in \mathcal{L} = \{1, 2, \dots, L\}$), the BS (each with $N_t = 1$ transmit antennas) serves K_l users (each with $N_r = 1$ receive antenna). The total number of users in L cells is $K = \sum_{l \in \mathcal{L}} K_l$. As mentioned earlier, the BSs work cooperatively, hence we can view the network as a 'super BS' serves K users. Other than the total power constraint that the 'super BS' has to meet, each individual base station- l is also subject to its own power constraint.

The transmitted signal vector from the 'super BS' for a given user is $\mathbf{x} \in \mathbb{C}^{L \times 1}$, which is subject to the total as well as the individual BS transmit power constraint: $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq P_t, \mathbb{E}\{\|x_l\|^2\} \leq P_{t,l}, \forall l \in \mathcal{L}$. The channel between the 'super BS' and user- k is given by $\mathbf{h}_k \in \mathbb{C}^{1 \times L}, k \in \mathcal{K} = \{1, 2, \dots, K\}$. The received signal at user- k

can be expressed as:

$$y_k = \mathbf{h}_k \mathbf{x} + n_k \quad (6.1)$$

where $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is AWGN. CSIR is assumed to be perfect.

6.2 ASR Maximization and WMMSE-AO with SAA Algorithm

Following Chapter 5, we can write the SINR of the common stream at user- k as:

$$\eta_{c,k} = \frac{|\mathbf{h}_k \mathbf{p}_c|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (6.2)$$

After the common stream is decoded, the contribution of it will be subtracted using SIC. Then user- k will decode its corresponding private stream by treating the private streams of the other users as noise. At user- k , the SINR of the private stream is:

$$\eta_k = \frac{|\mathbf{h}_k \mathbf{p}_k|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k \mathbf{p}_j|^2 + 1} \quad (6.3)$$

The achievable rates for common stream s_c and private stream s_k are $R_{c,k} = \log_2 (1 + \eta_{c,k})$ and $R_k = \log_2 (1 + \eta_k)$ respectively. In order for the common message W_c to be decoded successfully by all users, the common achievable rate R_c should not exceed $R_c = \min\{R_{c,j}\}_{j=1}^K$. Assuming that the common achievable rate is shared by users: $R_c = \sum_{k \in \mathcal{K}} Z_{c,k}$, then the total achievable rate for user- k can then be expressed as $R_{total,k} = Z_{c,k} + R_k$, and the SR is given by $R_s = \sum_{k \in \mathcal{K}} R_{total,k}$. Then the Average Sum-Rate optimization problem can be written as:

$$\begin{aligned} \text{ASR} &= \max_{\mathbf{P}, \mathbf{z}} \sum_{k \in \mathcal{K}} \bar{R}_{total,k} \\ \text{s.t. } &\text{diag}(\mathbf{P} \mathbf{P}^H)_l \leq P_{t,l}, \forall l \in \mathcal{L} \\ &\bar{R}_{total,k} \geq R_k^{th}, \forall k \in \mathcal{K} \\ &\sum_{k \in \mathcal{K}} \bar{Z}_{c,k} \leq \bar{R}_{c,k}, \forall k \in \mathcal{K} \\ &\mathbf{z} \geq \mathbf{0} \end{aligned} \quad (6.4)$$

where $\mathbf{z} = [\bar{Z}_{c,1}, \bar{Z}_{c,2}, \dots, \bar{Z}_{c,K}]$ is the average common rate vector and $\bar{R}_{total,k} = \bar{Z}_{c,k} + \bar{R}_k, \forall k \in \mathcal{K}$.

By utilizing SAA, we can transform Eq.6.4, a stochastic problem into a deterministic problem which is given by:

$$\begin{aligned}
 \text{ASR}^{(N)} &= \max_{\mathbf{P}, \mathbf{z}} \sum_{k \in \mathcal{K}} \bar{R}_{total,k} \\
 \text{s.t. } &\text{diag}(\mathbf{P}\mathbf{P}^H)_l \leq P_{t,l}, \forall l \in \mathcal{L} \\
 &\bar{R}_{total,k} \geq R_k^{th}, \forall k \in \mathcal{K} \\
 &\sum_{k \in \mathcal{K}} \bar{Z}_{c,k} \leq \bar{R}_{c,k}^{(N)}, \forall k \in \mathcal{K} \\
 &\mathbf{z} \geq \mathbf{0}
 \end{aligned} \tag{6.5}$$

where $\bar{R}_{total,k} = \bar{Z}_{c,k} + \bar{R}_k^{(N)}$, $\forall k \in \mathcal{K}$.

We now transform the deterministic ASR problem into the WMMSE problem. The received signal power X of user- k ($k \in \mathcal{K} = \{1, 2, \dots, K\}$) is given by:

$$\begin{aligned}
 X_{c,k}^{(n)} &= |\mathbf{h}_k^{(n)} \mathbf{p}_c|^2 + \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1 \\
 X_k^{(n)} &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1
 \end{aligned} \tag{6.6}$$

And the noise power Y of user- k is:

$$\begin{aligned}
 Y_{c,k}^{(n)} &= \sum_{j \in \mathcal{K}} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1 \\
 Y_k^{(n)} &= \sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_k^{(n)} \mathbf{p}_j|^2 + 1
 \end{aligned} \tag{6.7}$$

The MSE can be written as:

$$\begin{aligned}
 e_{c,k}^{(n)} &= |g_{c,k}^{(n)}|^2 X_{c,k}^{(n)} - 2\Re\{g_{c,k}^{(n)} \mathbf{h}_k^{(n)} \mathbf{p}_c\} + 1 \\
 e_k^{(n)} &= |g_k^{(n)}|^2 X_k^{(n)} - 2\Re\{g_k^{(n)} \mathbf{h}_k^{(n)} \mathbf{p}_k\} + 1
 \end{aligned} \tag{6.8}$$

The optimal MMSE combiner g can be expressed as:

$$\begin{aligned}
 (g_{c,k}^{(n)})^{\text{MMSE}} &= \mathbf{p}_c^H \mathbf{h}_k^{(n)H} / X_{c,k}^{(n)} \\
 (g_k^{(n)})^{\text{MMSE}} &= \mathbf{p}_k^H \mathbf{h}_k^{(n)H} / X_k^{(n)}
 \end{aligned} \tag{6.9}$$

We can now write the MMSE as:

$$\begin{aligned}
 (e_{c,k}^{(n)})^{\text{MMSE}} &= Y_{c,k}^{(n)} / X_{c,k}^{(n)} \\
 (e_k^{(n)})^{\text{MMSE}} &= Y_k^{(n)} / X_k^{(n)}
 \end{aligned} \tag{6.10}$$

The optimum MMSE weights can be expressed using the MMSE:

$$\begin{aligned} u_{c,k}^{(n)} &= 1/(e_{c,k}^{(n)})^{\text{MMSE}} \\ u_k^{(n)} &= 1/(e_k^{(n)})^{\text{MMSE}} \end{aligned} \quad (6.11)$$

The SAFs are defined as: $\bar{u}_{c,k}, \bar{u}_k, \bar{r}_{c,k}, \bar{r}_k, \bar{\mathbf{f}}_{\mathbf{c},\mathbf{k}}, \bar{\mathbf{f}}_{\mathbf{k}}, \bar{\mathbf{v}}_{\mathbf{c},\mathbf{k}}, \bar{\mathbf{v}}_{\mathbf{k}}, \bar{w}_{c,k}, \bar{w}_k$, and can be obtained by calculating the ensemble average of the following equations:

$$\begin{aligned} r_{c,k}^{(n)} &= u_{c,k}^{(n)} |g_{c,k}^{(n)}|^2 \\ r_k^{(n)} &= u_k^{(n)} |g_k^{(n)}|^2 \end{aligned} \quad (6.12)$$

$$\begin{aligned} \mathbf{f}_{\mathbf{c},\mathbf{k}}^{(\mathbf{n})} &= r_{c,k}^{(n)} \mathbf{h}_k^{(n)H} \mathbf{h}_k^{(n)} \\ \mathbf{f}_{\mathbf{k}}^{(\mathbf{n})} &= r_k^{(n)} \mathbf{h}_k^{(n)H} \mathbf{h}_k^{(n)} \end{aligned} \quad (6.13)$$

$$\begin{aligned} \mathbf{v}_{\mathbf{c},\mathbf{k}}^{(\mathbf{n})} &= u_{c,k}^{(n)} \mathbf{h}_k^{(n)H} g_k^{(n)H} \\ \mathbf{v}_{\mathbf{k}}^{(\mathbf{n})} &= u_k^{(n)} \mathbf{h}_k^{(n)H} g_k^{(n)H} \end{aligned} \quad (6.14)$$

$$\begin{aligned} w_{c,k}^{(n)} &= \log_2 \left(u_{c,k}^{(n)} \right) \\ w_k^{(n)} &= \log_2 \left(u_k^{(n)} \right) \end{aligned} \quad (6.15)$$

By using the above SAFs, we can define average WMSE:

$$\begin{aligned} \bar{\phi}_{c,k}^{(N)} &= \mathbf{p}_c^H \bar{\mathbf{f}}_{c,k} \mathbf{p}_c + \sum_{j=1}^K \mathbf{p}_j^H \bar{\mathbf{f}}_{c,k} \mathbf{p}_j + \sigma_n^2 \bar{r}_{c,k} - 2\Re\{\bar{\mathbf{v}}_{c,k}^H \mathbf{p}_c\} + \bar{u}_{c,k} - \bar{w}_{c,k} \\ \bar{\phi}_k^{(N)} &= \sum_{j=1}^K \mathbf{p}_j^H \bar{\mathbf{f}}_k \mathbf{p}_j + \sigma_n^2 \bar{r}_k - 2\Re\{\bar{\mathbf{v}}_k^H \mathbf{p}_k\} + \bar{u}_k - \bar{w}_k \end{aligned} \quad (6.16)$$

The WMMSE-Rate relationship is $\bar{\phi}_{c,k}^{(N)} = 1 - \bar{R}_{c,k}^{(N)}$, $\bar{\phi}_k^{(N)} = 1 - \bar{R}_k^{(N)}$, and the ASR problem is now transformed into the WMMSE problem:

$$\begin{aligned} \text{WMMSE} &= \min_{\mathbf{P}, \mathbf{t}, \mathbf{g}, \mathbf{u}} \sum_{k \in \mathcal{K}} \bar{\phi}_{total,k} \\ \text{s.t.} \quad &\text{diag}(\mathbf{P}\mathbf{P}^H)_l \leq P_{t,l}, \forall l \in \mathcal{L} \\ &\bar{\phi}_{total,k} \leq 1 - R_k^{th}, \forall k \in \mathcal{K} \\ &\sum_{j \in \mathcal{K}} \bar{T}_{c,j} \geq \bar{\phi}_{c,k}^{(N)} - 1, \forall k \in \mathcal{K} \\ &\mathbf{t} \leq \mathbf{0} \end{aligned} \quad (6.17)$$

where \mathbf{t} is the common WMSE vector and $\bar{\phi}_{total,k}$ is the individual WMSE for each user: $\bar{\phi}_{total,k} = \bar{T}_{c,k} + \bar{\phi}_k^{(N)}, \forall k \in \mathcal{K}$. We can use the AO algorithm below to solve Eq.6.17.

Algorithm 3: Alternating Optimization Algorithm (RSMA for CoMP-JT with partial CSIT)

Initialize: $m \leftarrow 0$, $\mathbf{P}^{[m]}$, WMMSE $^{[m]}$

repeat

$m \leftarrow m + 1$, $\mathbf{P}^{[m-1]} \leftarrow \mathbf{P}$;

$\mathbf{g} \leftarrow \mathbf{g}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$, $\mathbf{u} \leftarrow \mathbf{u}^{\text{MMSE}}(\mathbf{P}^{[m-1]})$;

update SAFs: $\bar{u}_{c,k}, \bar{u}_k, \bar{r}_{c,k}, \bar{r}_k, \bar{\mathbf{f}}_{\mathbf{c},\mathbf{k}}, \bar{\mathbf{f}}_{\mathbf{k}}, \bar{\mathbf{v}}_{\mathbf{c},\mathbf{k}}, \bar{\mathbf{v}}_{\mathbf{k}}, \bar{w}_{c,k}, \bar{w}_k$ using updated \mathbf{g} and \mathbf{u} ;

update (\mathbf{P}, \mathbf{t}) by solving Eq.6.17 using updated SAFs;

until $|\text{WMMSE}^{[m]} - \text{WMMSE}^{[m-1]}| \leq \epsilon$;

6.3 Results and Discussions

The initialization process follows Section 5.4. The SAA sample size is $N = 100$. The simulation result is averaged over 100 random channel estimate realizations. We will vary the quality scaling exponent α , inter-cell channel disparity a and inter-cell channel disparity b to evaluate the performance of RSMA and SDMA.

6.3.1 Two-Cell Network

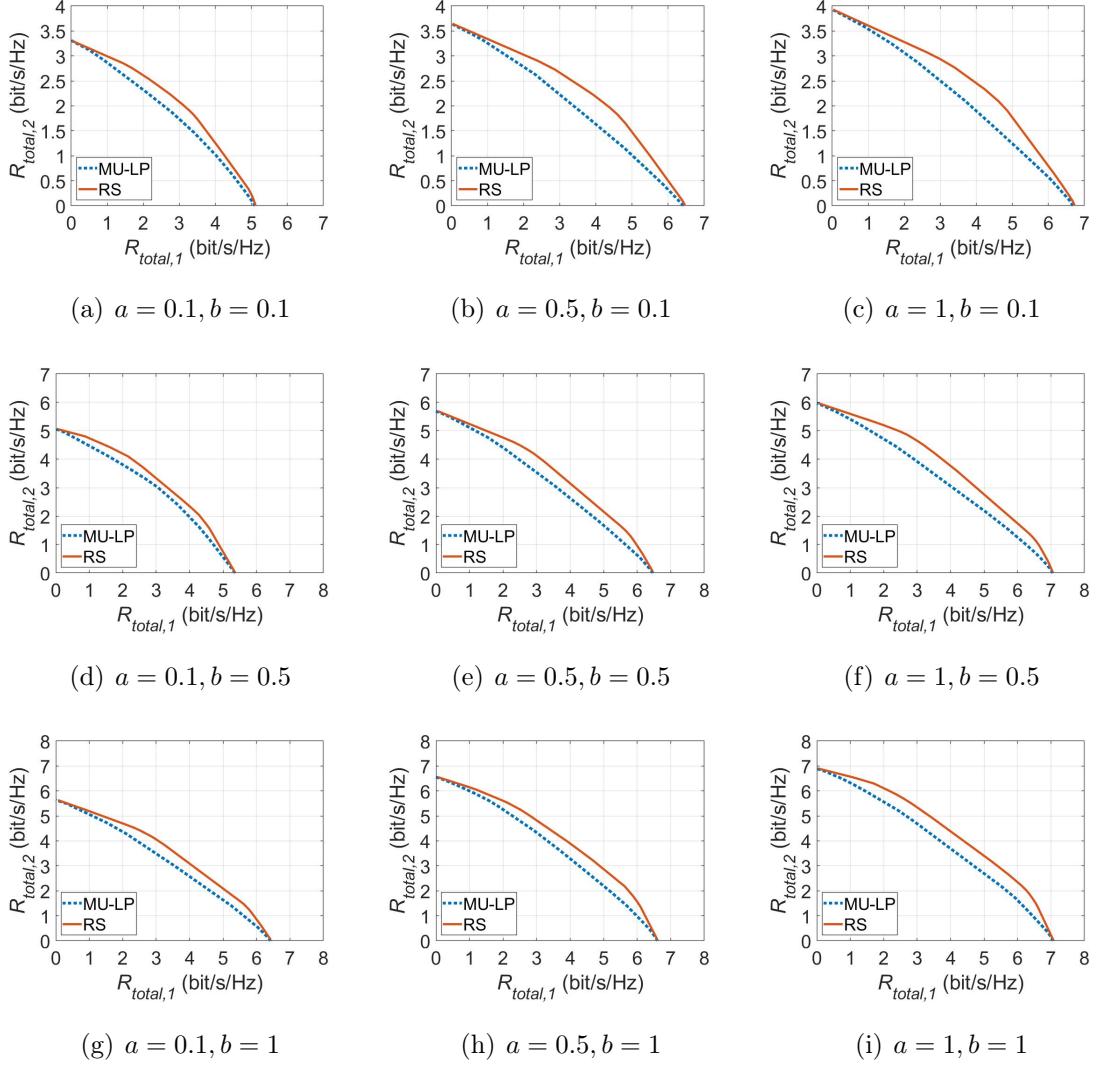


Figure 6.1: Achievable rate region of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-0.6}$, SNR = 20dB.

Fig.6.1 demonstrates the rate region of SDMA and RSMA in CoMP-JT networks with partial CSIT. In all subfigures which consider different values of channel dis-

parity, RSMA outperforms SDMA. As discussed in Section 4.3, SDMA is efficient when a is small and b is large, the same applies here: in Fig.6.1(g), the rate region gap is (one of the) smallest, and the rate region gap in Fig.6.1(c) is (one of the) largest. RS on the other hand, is more robust to random channel disparities than MU-LP.

Fig.6.3 demonstrates the SR versus SNR plot for 2-user case with and without R_{th} for CoMP-JT ($a, b = 0.5$) with partial CSIT. When $\alpha = 0.3/0.6/0.9$, the DoF of SDMA for 2-user network is $\max\{1, K\alpha\} = 1/1.2/1.8$, and the DoF of RSMA for 2-user network is $1 + (k - 1)\alpha = 1.3/1.6/1.9$, which can be also obtained from the subfigure (a)(c)-(f). In 6.3(b), the slope of the SDMA plot is not the same as DoF as a result of the very poor CSIT quality ($\alpha = 0.3$). Other features in Fig.6.3 are the same as those of Fig.5.4 in Section 5.4. Fig.6.2 demonstrates the SR with $a = 0.1, b = 0.1$ and $\alpha = 0.6$. When a and b are small, RS still have better SR and DoF than MU-LP. In summary, in CoMP-JT network with partial CSIT, RS outperforms MU-LP not only in terms of SR and QoS enhancement, but also from DoF perspective. We will demonstrate the advantage of RS in the next subsection through extensive results using various sets of a, b and α .

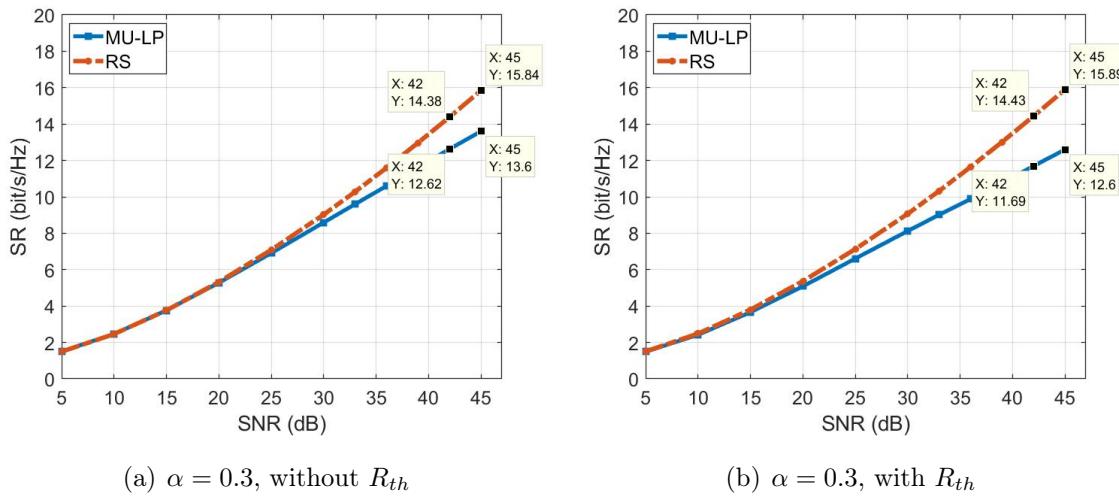


Figure 6.2: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-0.6}, a = 0.1, b = 0.1$. (a) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b) $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, \dots, 0.1]$ bit/s/Hz for SNR = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45] dB.

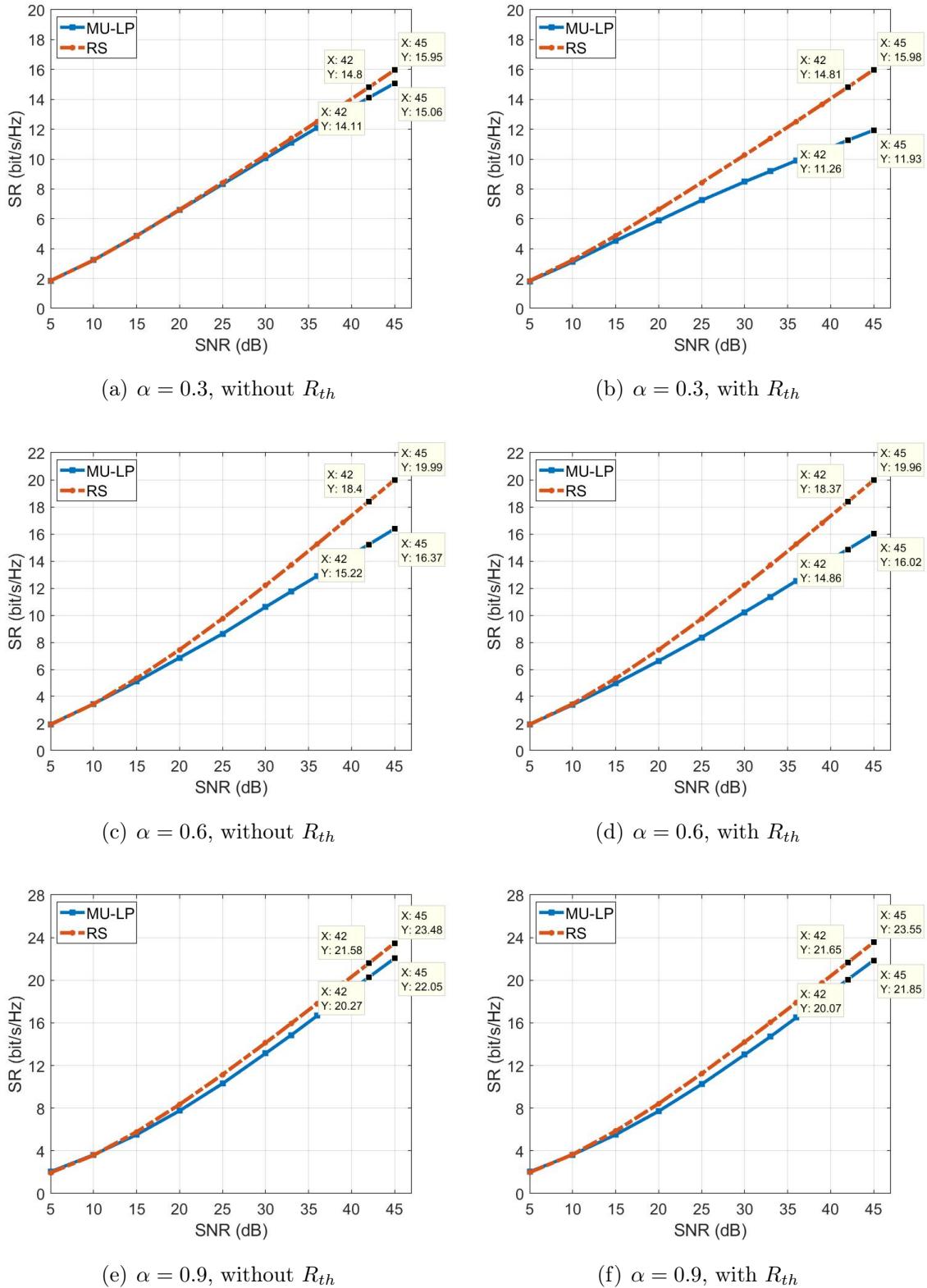


Figure 6.3: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 2, N_t = 1, K_l = 1, l \in \{1, 2\}, \sigma_e^2 = P_t^{-\alpha}, a = 0.5, b = 0.5$.
 (a/c/e) $\mathbf{r}_{th} = \mathbf{0}$ bit/s/Hz. (b/d/f) $\mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, \dots, 0.1]$ bit/s/Hz for $\text{SNR} = [5, 10, 15, 20, 25, 30, 33, 36, 39, 42, 45]$ dB.

6.3.2 Three-Cell Network

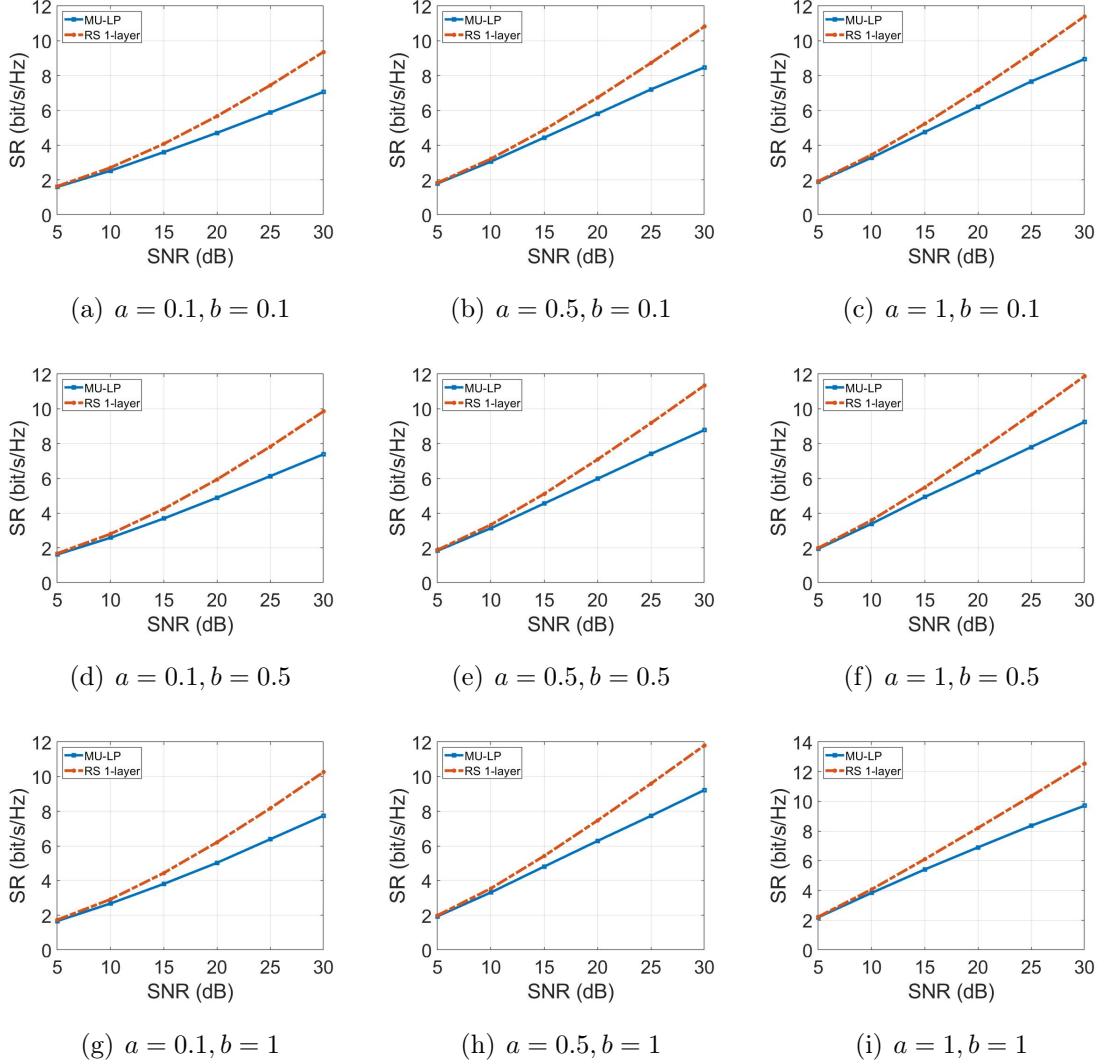


Figure 6.4: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.3}, \mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.

Fig.6.4-6.6 show the SR of a 3-cell CoMP-JT network with partial CSIT and QoS constraints, for different a, b and quality scaling exponent α . In all three figures, RS has higher SR and DoF than MU-LP. As mentioned earlier, the common message is turned on at high SNR, and contribute to the SR and DoF enhancement. The SR increases with a and b as the rate is increased with the channel strength. The advantage of RS is more explicit when α is small (i.e. poor CSIT quality); when CSIT is nearly perfect (e.g. $\alpha = 0.9$), RS is not as dominate as it is for $\alpha = 0.3$ and $\alpha = 0.6$. Again, such results show that using RS can relax CSIT requirements

under QoS constraints.

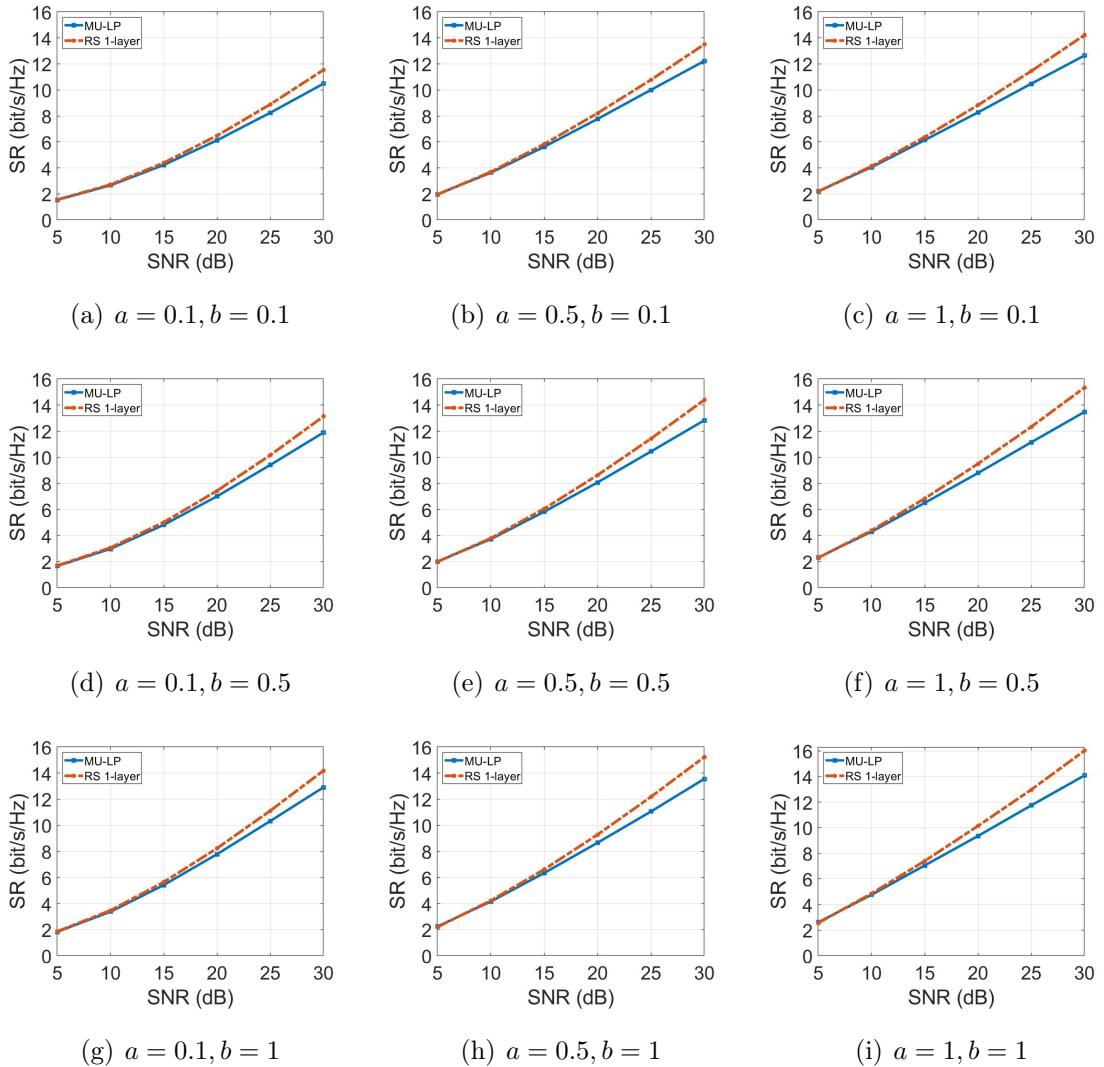


Figure 6.5: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.6}, \mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.

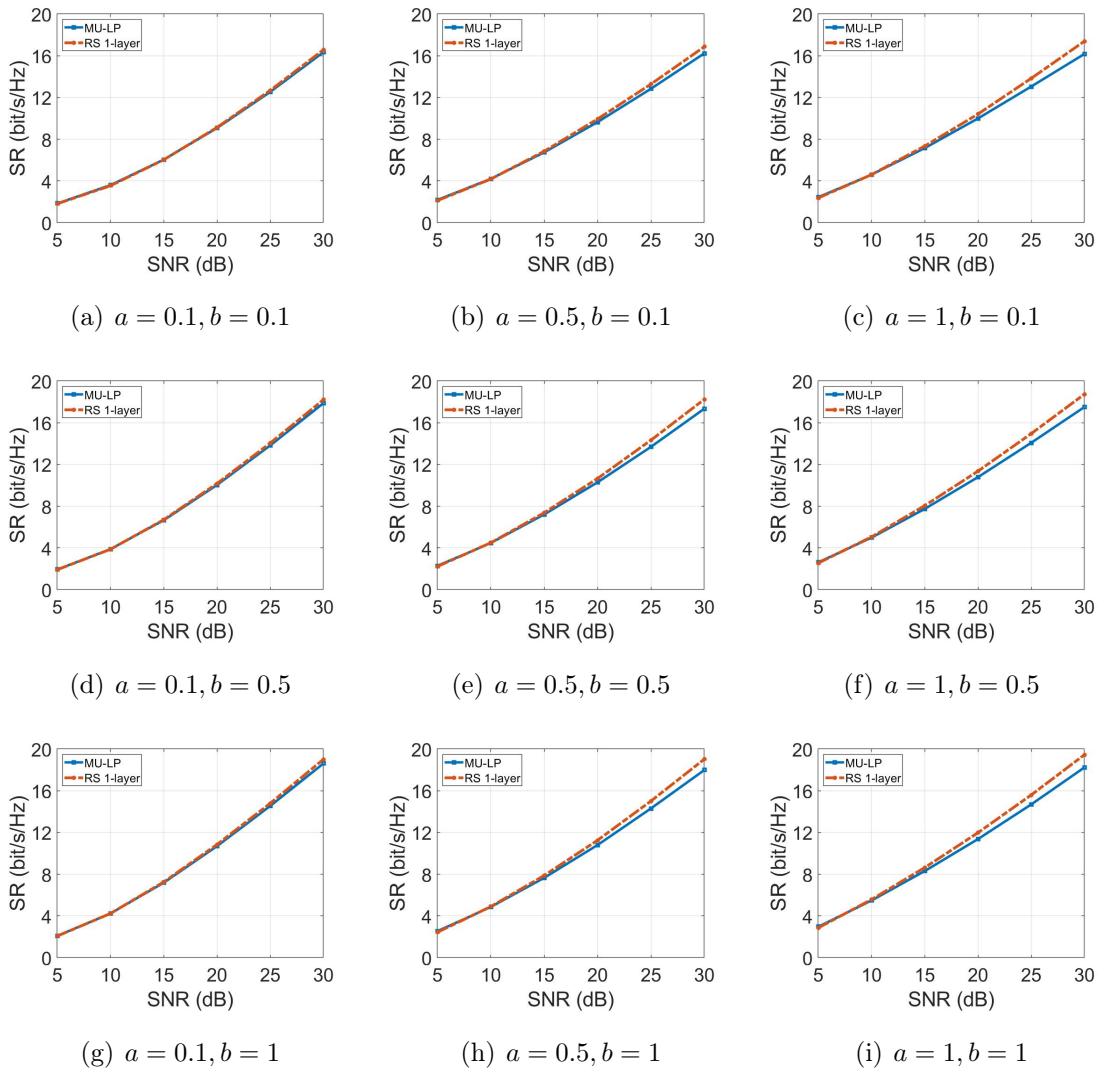


Figure 6.6: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.9}, \mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.

Fig.6.7-6.9 show the SR at 20dB with $R_{th} = 0.1$ bit/s/Hz for different a , b and α , in a 3-cell CoMP-JT network. The SR increases with a , b and α . By considering more sets of a and b , we show the benefits of RSMA over SDMA in various network settings. Therefore, we can conclude that RSMA has a better SR and DoF performance than SDMA, and is able to provide robustness and QoS enhancement in the CoMP-JT network¹ with partial CSIT.

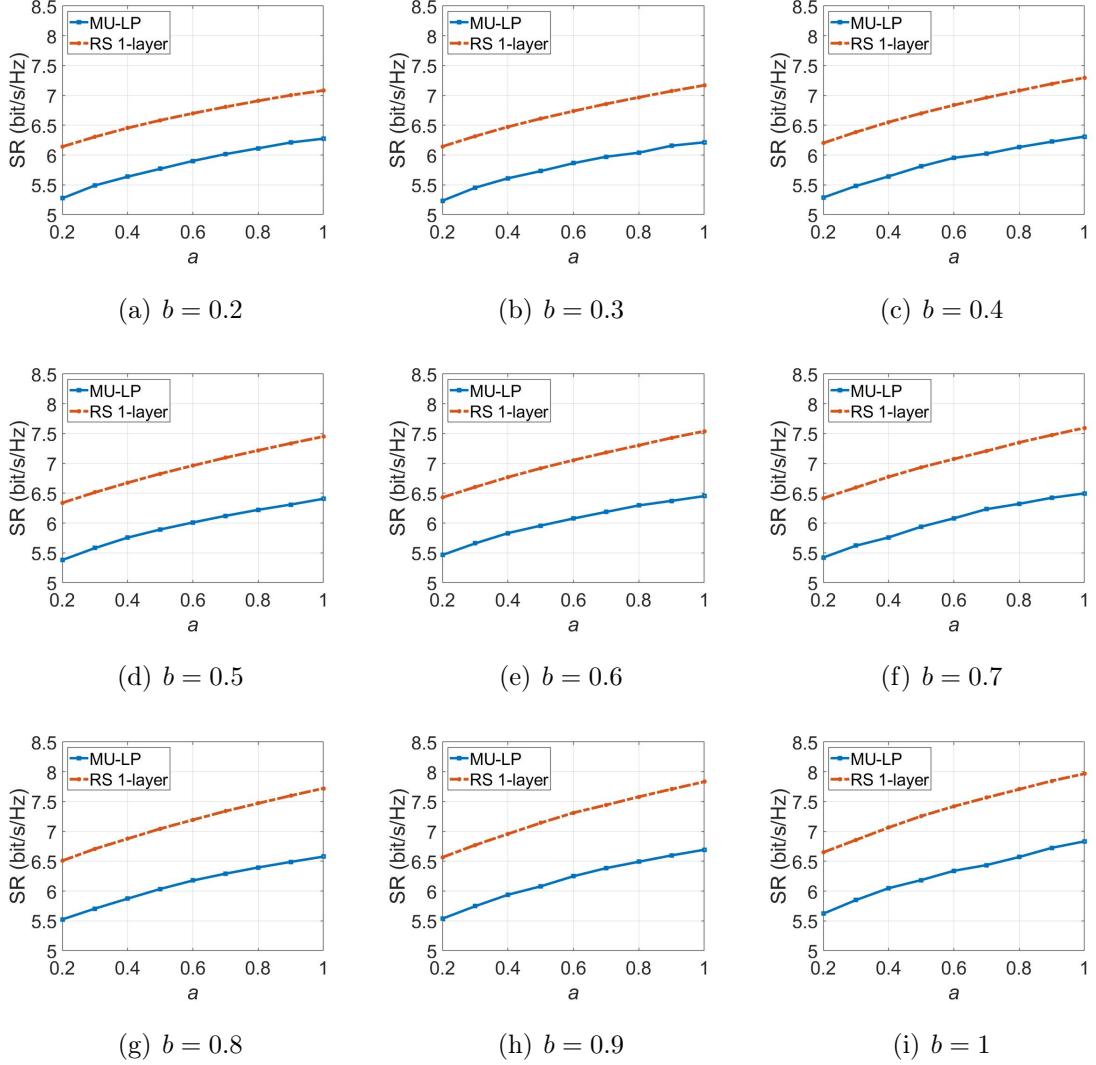


Figure 6.7: SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3$, $N_t = 1$, $K_l = 1$, $l \in \{1, 2, 3\}$, $\sigma_e^2 = P_t^{-0.3}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.

¹As mentioned earlier, the 2/3-cell networks are typical for CoMP-JT as the benefits gained from the cooperation between more cells is little. Nevertheless, the simulations of 5-cell network with partial CSIT is also conducted, which shows the rate enhancement using RS and also the inefficiency of cooperation between more than 3 cells. Such results can be found in Appendix B.

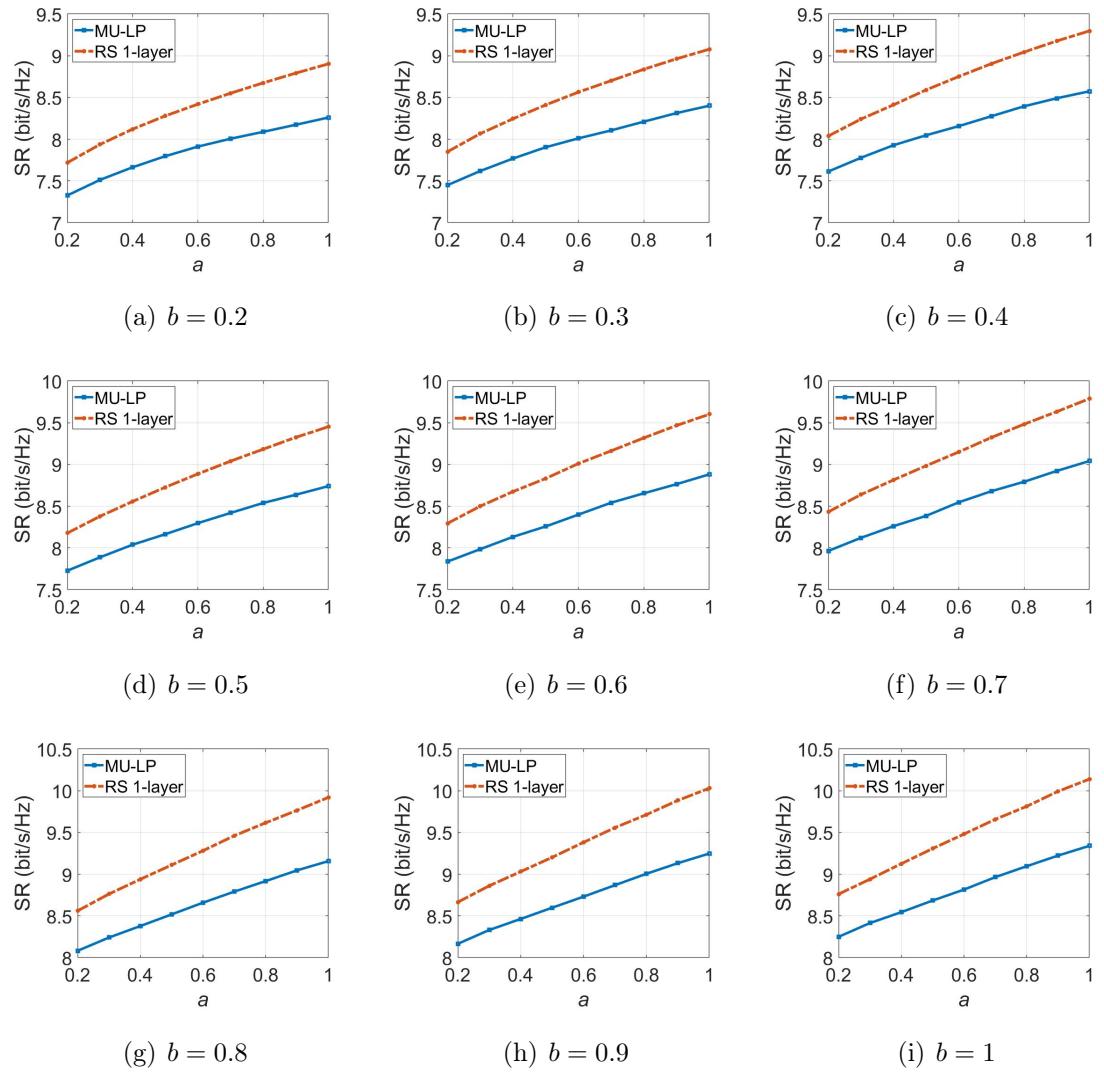


Figure 6.8: SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.6}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.

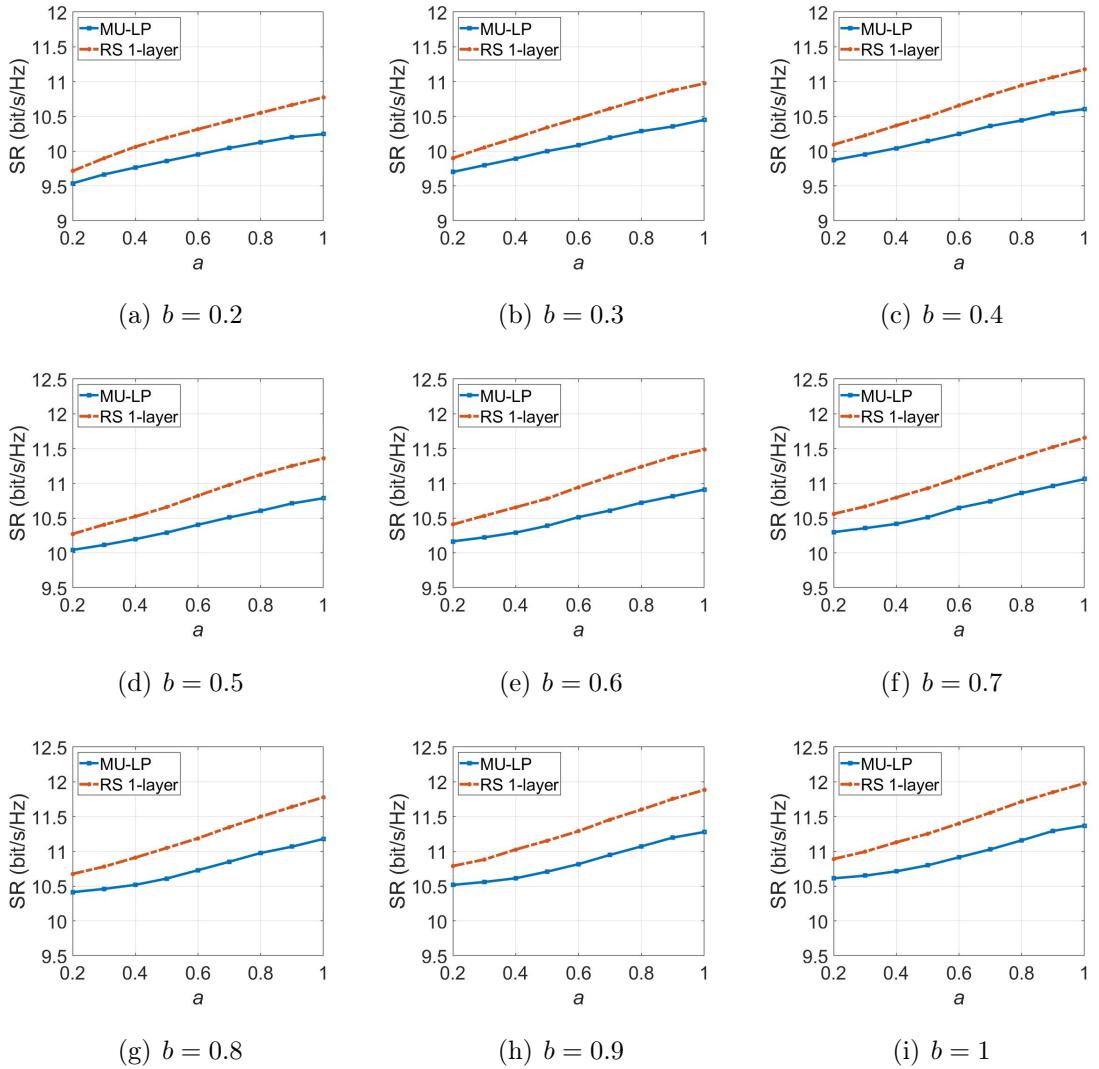


Figure 6.9: SR versus a for different b of MU-LP based SDMA and RSMA for CoMP-JT with partial CSIT, $L = 3, N_t = 1, K_l = 1, l \in \{1, 2, 3\}, \sigma_e^2 = P_t^{-0.9}$, SNR = 20dB, $R_{th} = 0.1$ bits/s/Hz.

Chapter 7

Conclusions and Future Work

To conclude, the use of RSMA in CoMP-JT with partial CSIT is proposed for the first time, and it has been shown through extensive simulation results that RSMA outperforms SDMA regarding both sum-rate maximization and DoF in a wide range of network settings. Such results have successfully reach the goal set at the start of the project. We discussed the limitations of NOMA and SDMA, and the benefits of RSMA by implementing linear precoding at the transmitter and SIC at the receiver. The ability of RSMA to partially decode interference and partially treat interference as noise bridges NOMA and SDMA, and takes advantages of their benefits while lifting the limitations. NOMA works well only in overloaded networks when channel strengths are different and channels are aligned, whereas SDMA has a good performance in the underloaded networks when user channels are orthogonal and have similar channel strengths. RSMA on the other hand, is robust to general transmission networks and user deployments. In addition, RSMA is able to provide rate gain with and without QoS.

We also investigated the use of RSMA in CoMP-JT, a cooperated multi-cell network, where the individual BS power constraint is considered. We show that RSMA is robust to any inter-cell and inter-user channel strength disparities. Using RSMA to tackle imperfect CSIT is also discussed, and RSMA is superior than SDMA in terms of both DoF and SR. Finally, we have brought different pieces together for SR maximization using RS in CoMP-JT with partial CSIT. Firstly, the stochastic ASR problem is transformed into a deterministic problem using sample average approximation method. Then by establishing the WMMSE-Rate relationship, we further convert the deterministic ASR problem to a WMMSE problem and solve it using the alternating optimization algorithm. We show that compared with SDMA,

RSMA is able to provide higher DoF, rate gain, QoS enhancement and robustness in CoMP-JT with partial CSIT.

Due to the limited time available for this project, only a small number of features in RSMA have been investigated. Other aspects such as max-min fairness, DoF region, optimal coding, etc., can be further studied under the CoMP-JT with partial CSIT setup in the future. In addition, other methods besides SAA can be explored to marry with RS for tackling the imperfect CSIT. Moreover, the RSMA approach in CoMP-JT with partial CSIT can be further modified and extended, and could play an important role in Massive MIMO and also other next generation of wireless networks.

In terms of theoretical limits, the DoF of RS in the overloaded network is still an open problem although simulation results have demonstrated DoF benefits of RS over NoRS in overloaded MISO BC. Another theoretical limit to be discovered is the capacity region under imperfect CSIT. It has been proved that DPC can achieve the capacity region under perfect CSIT, but the capacity region of MISO BC with partial CSIT is still unknown. Unlike DPC or SDMA or other conventional multiple access schemes whose design is under perfect CSIT, using RS with SAA is superior as the one of the fundamental design specifications is that partial CSIT assumed. In addition, the common message's ability to decode interference leads to performance enhancement. Therefore, having these three features in mind, namely, DPC's benefits under perfect CSIT, the advantage of SAA in partial CSIT, and the benefits of RS/common message for both perfect and imperfect CSIT, we can combine DPC, RS and SAA together and investigate the rate region. One possible philosophy could be encoding private messages using DPC and encoding common message using linear precoding techniques, and the receiver decodes the common message first then private message using SIC. SAA should also be incorporated into such scheme. As DPC introduces complexity, it's preferred to use less complicated 1-layer RS due to the fact that decoding orders don't need to be considered in 1-layer RS. The SR and DoF are expected to be enhanced by using such strategies, and could bring us closer to finding the capacity region of MISO BC under partial CSIT.

Appendix A

For full MATLAB programmes, please refer to the author's [Github](#) page¹.

¹ Available at: <https://github.com/LiangMa5/MScProject-2018-19> .

Appendix B

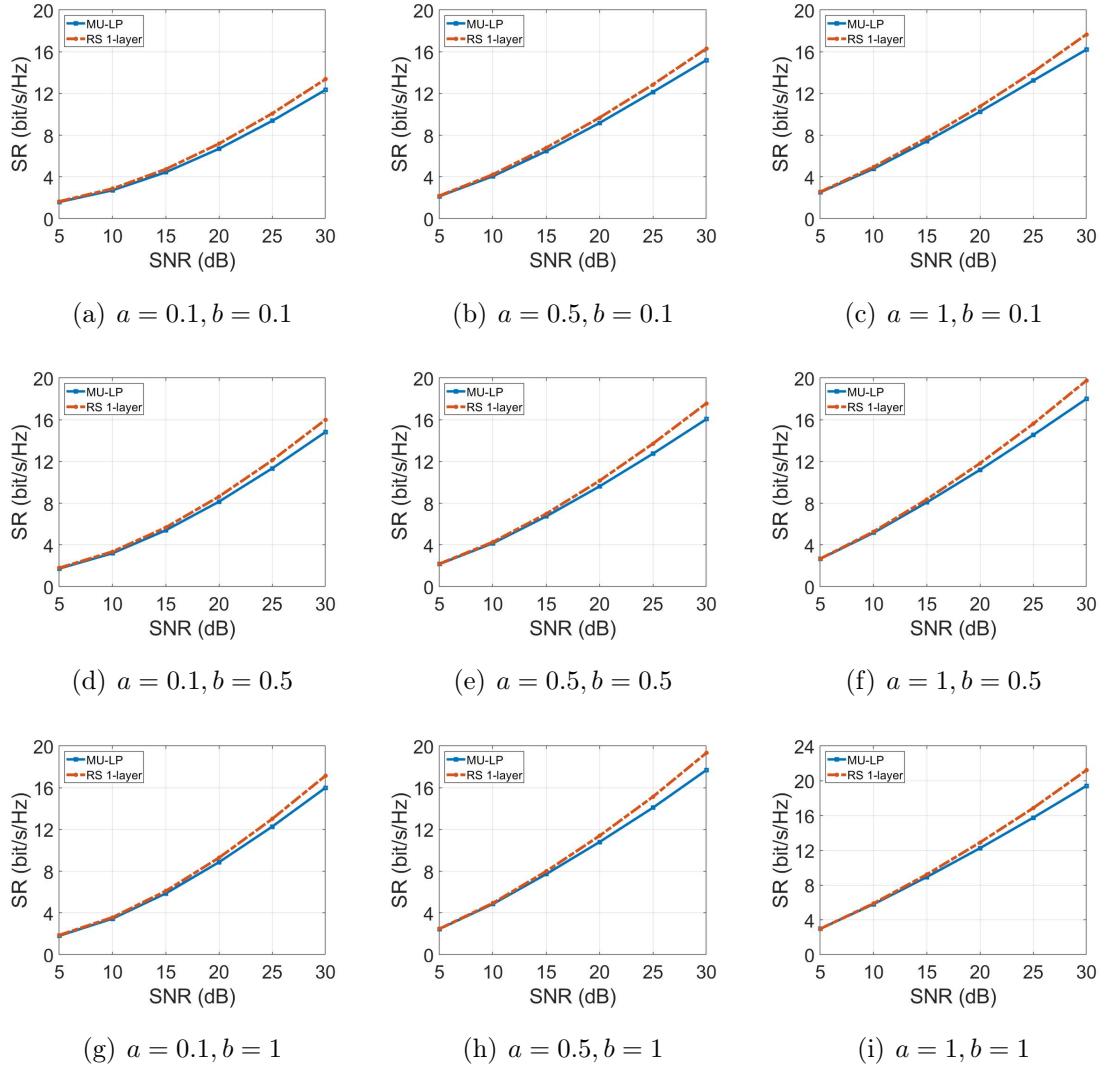


Figure B.1: SR vs SNR plot of MU-LP based SDMA and 1-layer RSMA for CoMP-JT with partial CSIT, $L = 5, N_t = 1, K_l = 1, l \in \{1, 2, 3, 4, 5\}, \sigma_e^2 = P_t^{-0.6}, \mathbf{r}_{th} = [0.04, 0.06, 0.08, 0.1, 0.1, 0.1]$ bits/s/Hz.

References

- [1] Y. Mao, B. Clerckx, and V. Li, “Rate-Splitting Multiple Access for Downlink Communication Systems: Bridging, Generalizing and Outperforming SDMA and NOMA,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 133–187, May 2018.
- [2] H. Joudeh and B. Clerckx, “Sum-Rate Maximization for Linearly Precoded Downlink Multiuser MISO Systems With Partial CSIT: A Rate-Splitting Approach,” *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.
- [3] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, “Rate Splitting for MIMO Wireless Networks: A Promising PHY-Layer Strategy for LTE Evolution,” *IEEE Communications Magazine*, vol. 54, no. 5, pp. 98–105, May 2016.
- [4] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [5] 3GPP TR 36.819 v11.0.0, “Coordinated Multi-Point Operation for LTE Physical Layer Aspects,” Sept. 2011.
- [6] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-Splitting Multiple Access for Co-ordinated Multi-Point Joint Transmission,” in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2019, pp. 1–6.
- [7] H. Joudeh and B. Clerckx, “Achieving Max-Min Fairness for MU-MISO with Partial CSIT: A Multicast Assisted Transmission,” in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 4739–4744.

-
- [8] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, “A Rate Splitting Strategy for Massive MIMO With Imperfect CSIT,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4611–4624, Jul. 2016.
 - [9] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, “Nonorthogonal Multiple Access for 5G and Beyond,” *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
 - [10] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access,” in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, Jun. 2013, pp. 1–5.
 - [11] H. Nikopour, E. Yi, A. Bayesteh, K. Au, M. Hawryluck, H. Baligh, and J. Ma, “SCMA for Downlink Multiple Access of 5G Wireless Networks ,” in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 3940–3945.
 - [12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2005.
 - [13] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, “A Minorization-Maximization Method for Optimizing Sum Rate in the Downlink of Non-Orthogonal Multiple Access Systems,” *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 76–88, Jan. 2016.
 - [14] Q. Zhang, Q. Li, and J. Qin, “Robust Beamforming for Nonorthogonal Multiple-Access Systems in MISO Channels,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10 231–10 236, Dec. 2016.
 - [15] H. Joudeh and B. Clerckx , “Rate-Splitting for Max-Min Fair Multigroup Multicast Beamforming in Overloaded Systems,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7276–7289, Nov. 2017.
 - [16] H. Weingarten, Y. Steinberg, and S. S. Shamai, “The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
 - [17] Taesang Yoo and A. Goldsmith, “On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

-
- [18] T. S. Han and K. Kobayashi, “A New Achievable Rate Region for the Interference Channel,” *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
 - [19] H. Joudeh and B. Clerckx, “A Rate-Splitting Strategy for Max-Min Fair Multi-group Multicasting,” in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2016, pp. 1–5.
 - [20] E. Piovano and B. Clerckx, “Optimal DoF Region of the K -User MISO BC With Partial CSIT,” *IEEE Communications Letters*, vol. 21, no. 11, pp. 2368–2371, Nov. 2017.
 - [21] H. Joudeh and B. Clerckx, “Robust Transmission in Downlink Multiuser MISO Systems: A Rate-Splitting Approach,” *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.
 - [22] C. Hao and B. Clerckx, “MISO Networks With Imperfect CSIT: A Topological Rate-Splitting Approach,” *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2164–2179, May 2017.
 - [23] C. Hao, B. Rassouli, and B. Clerckx, “Achievable DoF Regions of MIMO Networks With Imperfect CSIT,” *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6587–6606, Oct. 2017.
 - [24] M. Dai and B. Clerckx, “Multiuser Millimeter Wave Beamforming Strategies With Quantized and Statistical CSIT,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7025–7038, Nov. 2017.
 - [25] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, “Rate-Splitting to Mitigate Residual Transceiver Hardware Impairments in Massive MIMO Systems,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8196–8211, Sep. 2017.
 - [26] S. Lee, J. Kim, S. Moon, H. Kong, and I. Lee, “Zero-Forcing Beamforming in Multiuser MISO Downlink Systems Under Per-Antenna Power Constraint and Equal-Rate Metric,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 228–236, Jan. 2013.
 - [27] G. Caire and S. Shamai, “On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

-
- [28] N. Jindal, “MIMO Broadcast Channels With Finite-Rate Feedback,” *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
 - [29] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of Freedom of Time Correlated MISO Broadcast Channel With Delayed CSIT,” *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
 - [30] A. Gholami Davoodi and S. A. Jafar, “Aligned Image Sets Under Channel Uncertainty: Settling Conjectures on the Collapse of Degrees of Freedom Under Finite Precision CSIT,” *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5603–5618, Oct 2016.
 - [31] E. Piovano, H. Joudeh, and B. Clerckx, “Overloaded Multiuser MISO Transmission with Imperfect CSIT,” in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 34–38.
 - [32] H. Joudeh and B. Clerckx, “Sum Rate Maximization for MU-MISO with partial CSIT using Joint Multicasting and Broadcasting,” in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 4733–4738.
 - [33] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, “Weighted Sum-Rate Maximization Using Weighted MMSE for MIMO-BC Beamforming Design,” in *2009 IEEE International Conference on Communications*, Jun. 2009, pp. 1–6.
 - [34] M. Grant and S. Boyd, “CVX: MATLAB Software for Disciplined Convex Programming.” [Online]. Available: <http://cvxr.com/cvx>.
 - [35] B. Clerckx and C. Oestges, *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*, 2nd ed. Elsevier Science, 2013.