Imperial College London
Department of Electrical and Electronic Engineering

Rate-Splitting Multiple Access for Non-Terrestrial Communication and Sensing Networks

Longfei Yin

Supervised by Prof. Bruno Clerckx

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London and
the Diploma of Imperial College London, 2023

Declaration of Originality

I hereby declare that the content of this thesis is my own research work. The main parts of this thesis have been published in related conferences and journals. Where other sources of information have been used, they have been acknowledged.

Copyright Declaration

Abstract

Rate-splitting multiple access (RSMA) has emerged as a powerful and flexible non-orthogonal transmission, multiple access (MA) and interference management scheme for future wireless networks. This thesis is concerned with the application of RSMA to non-terrestrial communication and sensing networks. Various scenarios and algorithms are presented and evaluated.

First, we investigate a novel multigroup/multibeam multicast beamforming strategy based on RSMA in both terrestrial multigroup multicast and multibeam satellite systems with imperfect channel state information at the transmitter (CSIT). The max-min fairness (MMF)-degree of freedom (DoF) of RSMA is derived and shown to provide gains compared with the conventional strategy. The MMF beamforming optimization problem is formulated and solved using the weighted minimum mean square error (WMMSE) algorithm. Physical layer design and link-level simulations are also investigated. RSMA is demonstrated to be very promising for multigroup multicast and multibeam satellite systems taking into account CSIT uncertainty and practical challenges in multibeam satellite systems.

Next, we extend the scope of research from multibeam satellite systems to satellite-terrestrial integrated networks (STINs). Two RSMA-based STIN schemes are investigated, namely the coordinated scheme relying on CSI sharing and the co-operative scheme relying on CSI and data sharing. Joint beamforming algorithms are proposed based on the successive convex approximation (SCA) approach to optimize the beamforming to achieve MMF amongst all users. The effectiveness and robustness of the proposed RSMA schemes for STINs are demonstrated.

Finally, we consider RSMA for a multi-antenna integrated sensing and communications (ISAC) system, which simultaneously serves multiple communication users and estimates the parameters of a moving target. Simulation results demonstrate that RSMA is beneficial to both terrestrial and multibeam satellite ISAC systems by evaluating the trade-off between communication MMF rate and sensing Cramér-Rao bound (CRB).

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my supervisor, Prof. Bruno Clerckx for his constant support, guidance and continuous encouragement during my PhD study. I am profoundly grateful for his insightful comments and inspiring suggestions which set me on the right path from the very beginning and have made this research journey wonderful and fruitful. His great enthusiasm, rigorous attitude, professional skills, wide knowledge and kind personality will always inspire me in the future.

I am also truly thankful to my colleagues and friends in the Communications and Signal Processing group at the Department of Electrical and Electronic Engineering for the precious memories and wonderful friendships they have provided me during all this time and have enriched my life.

Finally, I am profoundly grateful to my parents for their boundless love that gives me the confidence to tackle all difficulties. They have always provided me with constant support and encouragement and supported my decisions. An infinite *thank you* for your endless love and endless support.

Abbreviations

4G fourth generation

5G fifth generation

6G sixth generation

ADMM alternating direction method of multipliers

AMC adaptive modulation and coding

AWGN additive white Gaussian noise

B5G beyond fifth generation

BC broadcast channel

BS base station

CCP convex-concave procedure

CDMA code division multiple access

CRB Cramér-Rao bound

CSIR channel state information at the receiver

CSIT channel state information at the transmitter

CUs cellular users

DoA direction of arrival

DoD direction of departure

DoF degree of freedom

EE energy efficiency

eURLLC extremely ultra reliable and low-latency communication

FDMA frequency division multiple access

FeMBB further-enhanced mobile broadband

FIM Fisher information matrix

GEO geostationary orbit

GW gateway

IoT Internet-of-Things

ISAC integrated sensing and communications

LEO low Earth orbit

LLS link-level simulation

LMI linear matrix inequality

LOS line-of-sight

MA multiple access

MFR minimum fairness rate

MIMO multiple-input multiple-output

MISO multiple-input single-output

MMF max-min fairness

MMSE minimum mean square error

MSE mean square error

MU multiuser

MU-LP multiuser linear precoding

NOMA non-orthogonal multiple access

NTN non-terrestrial network

OFDMA orthogonal frequency division multiple access

OMA orthogonal multiple access

PHY physical layer

QAM quadrature amplitude modulation

QoS quality-of-service

QPSK quadrature-phase-shift-keying

RAN radio access network

RCS radar cross-section

RSMA rate-splitting multiple access

RF radio frequency

RMSE root mean square error

SAA sample average approximation

SAGIN space-air-ground integrated network

SC superposition coding

SCA successive convex approximation

SDMA space-division multiple access

SDP semi-definite programming

SDR semi-definite relaxation

SIC successive interference cancellation

SINR signal-to-interference-noise ratio

SISO single-input single-output

SNR signal-to-noise ratio

SOCP second-order cone program

STIN satellite-terrestrial integrated network

SUs satellite users

SWIPT simultaneous wireless information and power transfer

TDMA time division multiple access

UAV unmanned aerial vehicle

ULA uniform linear array

umMTC ultra massive machine type communication

UPA uniform planar array

V2X vehicle-to-everything

VoD video on demand

WIPT wireless information and power transfer

WMMSE weighted minimum mean square error

WSR weighted sum-rate

ZFBF zero-forcing beamforming

Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the rapid development of wireless communications over the past few decades, the next-generation wireless networks, e.g., beyond fifth generation (B5G) and sixth generation (6G) have attracted widespread attention from both academia and industry. It is envisioned that B5G/6G will enable Internet to Everything, and will cope with the increasing demands for high throughput, reliability, heterogeneity of quality-of-service (QoS), and massive connectivity to satisfy the requirements of further-enhanced mobile broadband (FeMBB), extremely ultra reliable and low-latency communication (eURLLC), ultra massive machine type communication (umMTC) and new services such as integrated sensing and communications, integrated satellite-terrestrial, and extended reality [1]. To accommodate these requirements of next-generation wireless networks, multiple access (MA) techniques have become increasingly imperative to make better use of wireless resources and manage interference more efficiently.

1.1 Toward Rate-Splitting Multiple Access

The past decades have witnessed the evolution of MA schemes. The previous generations of wireless networks rely on orthogonal multiple access (OMA) which allocates orthogonal radio resources to users to alleviate multi-user interference, such as using frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA) or orthogonal frequency division multiple access (OFDMA). The choice of orthogonal radio resource allocation is motivated by avoiding multiuser interference and high transceiver complexity [2]. However, such an approach leads to inefficient use of radio resources. In fourth generation (4G) and fifth generation (5G), multiple-input multiple-output (MIMO) processing plays a pivotal role in wireless systems, and MA techniques are adopted in conjunction with multiuser (MU)-MIMO to achieve higher throughput by exploiting the spatial dimension resources.

The utilization of spatial domain and multi-antenna processing opens the door for space-division multiple access (SDMA), a well-established MA technique based on multiuser linear precoding (MU-LP). MU-LP is an efficient precoding strategy[1] for the multi-antenna broadcast channel (BC), which relies on linear precoding (also called beamforming) at the transmitter, and treats multiuser interference as noise at the receivers. It is able to achieve near-capacity performance when perfect channel state information at the transmitter (CSIT) is assumed and the user channels are nearly orthogonal with similar channel strengths or similar long-term signal-to-noise ratio (SNR) [2]. Through SDMA, multiple users are served in a non-orthogonal manner in the same time-frequency domain and the interference can be significantly mitigated by spatial beamforming. An alternative interpretation is that SDMA relies

[1] In this thesis, only channel-level precoding strategies are considered. These strategies exploit the knowledge of CSIT to design precoders to be applied to multiple data streams, thus suppressing interference. Note that symbol-level precoding uses the knowledge of both symbols of users and CSIT to exploit, rather than suppress constructive interference [3, 4].

on a transmit-side interference cancellation strategy. It has received considerable attention in the past decade and has become the basic principle behind numerous 4G and 5G techniques. However, its limitations are as follows. First, when the system is overloaded, i.e., when the number of served users is larger than the number of transmit antennas, the multiuser interference cannot be successfully suppressed, thus leading to significant multiplexing gain and rate loss. Second, SDMA is sensitive to the channel strength and orthogonality. Schedulers are expected to pair users with nearly orthogonal channels and relatively similar channel strengths. Therefore, the complexity of scheduling and user pairing according to the user channel conditions is not practical when conducting an exhaustive search process [5]. Scheduling algorithms with low complexity are required. Third, SDMA highly relies on the availability of accurate CSIT. The beamforming and interference nulling ability heavily depend on the CSIT accuracy. Applying SDMA designed for perfect CSIT in the presence of imperfect CSIT can result in residual multiuser interference caused by the imprecise beamforming at the transmitter. However, the CSIT is always imperfect in practice due to channel estimation errors, pilot contamination, limited and quantized feedback accuracy, delay, mobility, inaccurate calibration of radio frequency (RF) chains, etc [6].

Another non-orthogonal MA scheme which superposes users in the same time-frequency resource is known as power-domain non-orthogonal multiple access (NOMA).[2] NOMA relies on superposition coding (SC) at the transmit side and successive interference cancellation (SIC) at the receive side. SC-SIC has been studied for decades and is well known to achieve the capacity region of the single-input single-output (SISO) BC [7]. Through NOMA, at least one user is forced to fully decode the messages of the other co-scheduled users, and then remove them from its observation before decoding its own message. Interference is therefore removed. An

[2]In this thesis, we focus only on power-domain NOMA and simply use NOMA to represent power-domain NOMA.

alternative interpretation is that NOMA relies on a receiver-side interference cancellation strategy. Due to the benefits of multi-antenna over single-antenna systems, multi-antenna NOMA has been studied in a great number of literature in the past few years. Similar to the benefit of using SC-SIC in SISO BC, multi-antenna NOMA is very effective to cope with an overloaded deployment especially when the channels are closely aligned, and channel disparity can further promote the advantage of NOMA. However, the limitations of multi-antenna NOMA are as follows. First, different from the degraded SISO BC where the users can be ordered based on their channel strengths, the multi-antenna BC is non-degraded and the users cannot be ordered based on channel strengths. The precoders and decoding orders need to be jointly optimized. As the number of users increases, the number of decoding orders increases exponentially. Second, a great number of SICs are conducted at the receivers, which results in high receiver complexity. The number of SIC layers increases as the user number grows. Accordingly, multi-antenna NOMA can impose a significant computational burden on both the transmitter and the receivers. Third, the spatial domains cannot be efficiently used, and there exists a multiplexing gain loss. As analyzed in [8], for a multiple-input single-output (MISO) BC, the sum multiplexing gain achieved by multi-antenna NOMA is reduced to unity, which is the same as the multiplexing gain of OMA/single user MISO transmission. Fourth, the performance of NOMA is sensitive to channel strength and orthogonality. When the user channels are not aligned or with similar channel strengths, the performance degrades.

Indeed, SDMA and NOMA can be seen as two extreme interference management strategies, namely fully treating interference as noise and fully decoding interference. To overcome the limitations of both strategies and take full advantage of their benefits, rate-splitting multiple access (RSMA) has emerged as a promising and powerful non-orthogonal transmission, interference management and MA scheme for future multi-antenna wireless networks owning to its capability to enhance

the system performance in a wide range of network loads, user deployments and CSIT qualities. In [9], RSMA has been analytically demonstrated to generalize several existing MA techniques, namely SDMA, NOMA, OMA and physical-layer multicasting. RSMA relies on linearly preceded rate-splitting at the transmitter, and SIC at the receivers. The key behind the flexibility and robust manner of RSMA is to split user messages into common and private parts such that each of these parts can be decoded flexibly at one or multiple receivers. Through SIC, users sequentially decode the intended common streams (and therefore decode part of the interference). The private streams are only decoded by their corresponding users. This framework enables the capability of RSMA to partially decode the interference and partially treat interference as noise. Alternatively, RSMA can be interpreted as a smart combination of transmit-side and receive-side interference cancellation strategy, where the contribution of the common parts and the power allocated to the common and private parts can be adjusted flexibly [1]. This departs from the transmit-side-only and receive-side-only interference management strategies, e.g., SDMA and NOMA respectively. As a consequence, RSMA has the flexibility to cope with various interference levels and user deployment scenarios. RSMA is very robust to channel disparity, channel orthogonality and network loads [5]. It is demonstrated to provide benefits in terms of multiplexing gain, system spectral and energy efficiency with both perfect CSIT and imperfect CSIT [1, 2].

1.2 Motivation and Organization

With the explosive growth of data traffic and high demand for wireless connectivity in B5G/6G, existing cellular infrastructures may no longer provide ubiquitous and high-capacity global coverage to rural and remote areas [10]. Thereby, non-terrestrial network (NTN) is envisioned to provide heterogeneous services and seamless network

coverage for everyone and everything by complementing and extending terrestrial networks. The roles of NTNs lie in enhancing the availability in unserved (e.g., deserts, oceans, forests) or underserved areas (e.g., rural areas), enabling service reliability by providing service continuity for Internet-of-Things (IoT) devices or passengers on board moving platforms, and offering an infrastructure resilient to natural disasters on the ground.

In this thesis, amongst the NTN platforms spanning from satellite-based and airborne-based platforms, we particularly focus on the multibeam satellite systems that have received considerable attention in recent years due to the full frequency reuse across multiple narrow spot beams towards higher throughput [11, 12]. The available spectrum is aggressively reused, and thus inter-beam interference increases. Moreover, by combining the advantages of both satellite and terrestrial networks, the satellite-terrestrial integrated network (STIN) architecture shows great potential to find a new development path toward ubiquitous wireless networks [13]. The satellite sub-network shares the same frequency band as the terrestrial sub-network, and severe interference in and between the subnetworks is induced. Hence, analogous to terrestrial networks, it is deemed necessary to explore efficient MA strategies.

In addition to the demand for high-quality wireless connectivity, a common theme for the future trend is that sensing will play a more significant role than ever before and the demand for robust and accurate sensing capability increases [14]. Sharing of the frequency bands between radar sensors and communication systems has received considerable attention from both industry and academia, therefore motivating the research on integrated sensing and communications (ISAC) systems. ISAC systems can simultaneously perform wireless communications and remote sensing. Both functionalities are combined via shared use of the spectrum, the hardware platform and a joint signal processing framework. Such design leads to a trade-off between communication and radar performance and also calls for

flexible and robust MA strategies. Driven by the appealing benefits of RSMA in multi-antenna wireless communications, the main objective of this thesis is to investigate the application of RSMA to the aforementioned scenarios and enabling technologies in non-terrestrial communication and sensing networks, namely the multigroup multicast and multibeam satellite systems, STIN and ISAC systems, which are envisioned to play key roles in next-generation wireless networks.

The remainder of this thesis is organised as follows:

In Chapter 2, we review the fundamentals and related works of this thesis by introducing the principles of downlink RSMA and the state-of-the-art works on multigroup multicast and multibeam satellite systems, STIN and ISAC systems.

In Chapter 3, we investigate the application of RSMA for multigroup/multibeam multicast beamforming in the presence of imperfect CSIT. The effectiveness of RSMA is demonstrated in both cellular multigroup multicast and multibeam satellite systems to manage interference, taking into account various practical challenges. Physical (PHY) layer design and link-level simulations are also investigated. RSMA is demonstrated to be a very promising MA strategy for practical implementation in numerous application areas.

In Chapter 4, we introduce RSMA to STIN considering either perfect CSIT or imperfect CSIT with satellite channel phase uncertainties at the gateway (GW). Two RSMA-based STIN schemes are presented. Simulation results show the superiority of the proposed RSMA-based STIN to manage the interference in and between the satellite and terrestrial sub-networks.

In Chapter 5, we investigate the application of RSMA for ISAC systems, where the ISAC platform has a dual capability to simultaneously communicate with downlink users and probe detection signals to a moving target. Through RSMA-assisted ISAC beamforming design, RSMA is shown to be very promising for both terrestrial and

satellite ISAC systems to manage the multiuser/inter-beam interference as well as performing the radar functionality.

Finally, in Chapter 6, we conclude the thesis by summarising the achievements and discussing possible directions for future works.

1.3 List of Contributions

The novel contributions of this thesis are listed as follows:

1. Max-min fairness (MMF)-degree of freedom (DoF) of RSMA is analyzed in multigroup multicast with imperfect CSIT. RSMA is shown to provide MMF-DoF gains in both underloaded and overloaded systems.

2. An RSMA-based MMF multigroup multicast beamforming optimization problem with imperfect CSIT is formulated to investigate whether the MMF-DoF gain translates into MMF rate gain. A weighted minimum mean square error (WMMSE) algorithm is developed to solve the problem.

3. The RSMA-based multigroup multicast framework and algorithm are applied to a multibeam satellite setup. A novel RSMA-based multibeam multicast beamforming scheme is therefore studied.

4. The RSMA transmitter and receiver architecture and link-level simulation (LLS) platform are designed by considering finite length polar coding, finite alphabet modulation, adaptive modulation and coding (AMC) algorithm, etc.

5. The MMF rate/throughput performance of the proposed RSMA-based multigroup/multibeam multicast is compared with conventional strategies in both cellular and multibeam satellite systems.

6. A multiuser downlink RSMA-based STIN is presented, where the GW operates as a control center to implement centralized processing. Two integration levels of RSMA-based STIN are proposed, namely the coordinated scheme, and the cooperative scheme.

7. MMF RSMA-based STIN beamforming optimization problems are formulated considering both perfect CSIT and imperfect CSIT with satellite channel phase uncertainty. Iterative algorithms based on SCA are proposed respectively to solve the optimization problems.

8. The MMF rate performance of the proposed RSMA-based STIN is compared with several conventional baseline strategies.

9. A general RSMA-assisted ISAC system is presented, where the antenna array is shared by a co-located monostatic MIMO radar system and a multiuser communication system.

10. An RSMA-assisted ISAC beamforming optimization problem is formulated to investigate the trade-off between the radar and communication performance. An iterative algorithm based on SCA is proposed to solve the problem.

11. The performance of the proposed RSMA-assisted ISAC beamforming is compared with conventional baseline strategies considering both terrestrial and multibeam satellite ISAC systems.

1.4 Publications

The material presented in this thesis has led to the following publications:

1. L. Yin and B. Clerckx, "Rate-Splitting Multiple Access for Multigroup Multicast and Multibeam Satellite Systems," in *IEEE Transactions on Communi-*

cations, vol. 69, no. 2, pp. 976-990, Feb. 2021. [Chapter 3]

2. L. Yin and B. Clerckx, "Rate-Splitting Multiple Access for Multibeam Satellite Communications," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1-6. [Chapter 3]

3. L. Yin, O. Dizdar and B. Clerckx, "Rate-Splitting Multiple Access for Multi-group Multicast Cellular and Satellite Communications: PHY Layer Design and Link-Level Simulations," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1-6. [Chapter 3]

4. L. Yin and B. Clerckx, "Rate-Splitting Multiple Access for Satellite-Terrestrial Integrated Networks: Benefits of Coordination and Cooperation," in *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 317-332, Jan. 2023. [Chapter 4]

5. L. Yin, Y. Mao, O. Dizdar and B. Clerckx, "Rate-Splitting Multiple Access for 6G—Part II: Interplay With Integrated Sensing and Communications," in *IEEE Communications Letters*, vol. 26, no. 10, pp. 2237-2241, Oct. 2022. [Chapter 5]

6. L. Yin and B. Clerckx, "Rate-Splitting Multiple Access for Dual-Functional Radar-Communication Satellite Systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 1-6. [Chapter 5]

1.5 Notation

The following notation is used throughout the thesis. Boldface uppercase, boldface lowercase and standard letters denote matrices, column vectors, and scalars respectively. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$, where N is the size of the identity matrix. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex domains. $\mathbb{E}(\cdot)$ is

the expectation of a random variable. The real part of a complex number x is represented by $\mathcal{R}(x)$. The operators $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the Hermitian transpose respectively. $\mathrm{tr}()$ denotes the trace operator. $\mathbf{A} \succeq 0$ means that the symmetric matrix $\mathbf{A}$ is positive semidefinite. $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with diagonal entries given by the elements of $\mathbf{a}$. On the other hand, $\mathrm{diag}(\mathbf{A})$ is a vector with entries given by the diagonal elements of $\mathbf{A}$. $|\cdot|$ denotes the absolute value or the size of a set. $\|\cdot\|$ denotes the Euclidean norm.

Chapter 2

Background

In this chapter, the background knowledge and state-of-the-art works covered in this thesis are presented, including the fundamentals of downlink RSMA, multigroup multicast and multibeam satellite systems, satellite-terrestrial integrated networks and integrated sensing and communications.

2.1 Fundamentals of Downlink RSMA

We consider a MISO BC system, where K single-antenna users are simultaneously served by a base station (BS) equipped with N_t transmit antennas. As shown in Fig. 2.1, the message intended for user-k, $k \in \mathcal{K} = \{1, \cdots, K\}$ denoted as W_k is split into a common part $W_{c,k}$ and a private part $W_{p,k}$. The common parts are combined into a common message $W_c = \{W_{c,1}, \cdots, W_{c,K}\}$, which is then encoded into a single common stream $W_c \rightarrow s_c$. The private messages are encoded into corresponding private streams $W_{p,k} \rightarrow s_k, \ \forall k \in \mathcal{K}$.

By defining $\mathbf{p}_c \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{p}_k \in \mathbb{C}^{N_t \times 1}, \ \forall k \in \mathcal{K}$ as the linear precoders/beamforming

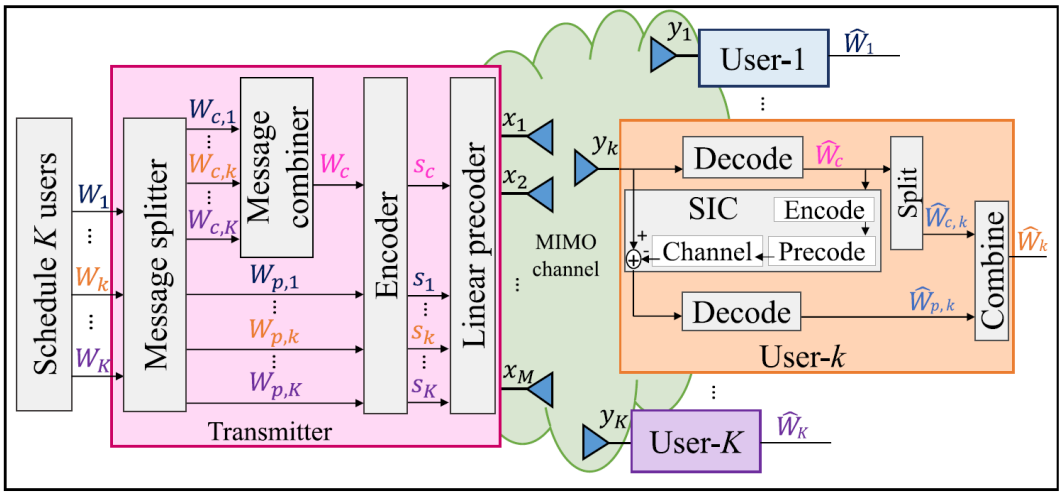

Figure 2.1: Transceiver architecture of K-user downlink RSMA.

vectors, the transmit signal $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ can be written as

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{p}_c s_c + \sum_{k=1}^{K} \mathbf{p}_k s_k, \tag{2.1}$$

where $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_K]$ denotes the beamforming matrix. The vector of symbol streams to be transmitted is $\mathbf{s} = [s_c, s_1, \cdots, s_K]^T \in \mathbb{C}^{(K+1)\times 1}$. We assume $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}$, thus the sum transmit power constraint at the transmitter is $\mathrm{tr}\left(\mathbf{P}\mathbf{P}^H\right) \leq P$, where P is the transmit power budget. The received signal at user-k is given by

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k = \mathbf{h}_k^H \mathbf{p}_c s_c + \mathbf{h}_k^H \sum_{k=1}^{K} \mathbf{p}_k s_k + n_k, \tag{2.2}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$ denotes the channel vector between the transmitter and the k-th user. $n_k \sim \mathcal{CN}\left(0, \sigma_{\mathrm{n},k}^2\right)$ is the receiver additive white Gaussian noise (AWGN) of zero mean and variance $\sigma_{\mathrm{n},k}^2$. It is assumed that $\sigma_{\mathrm{n},1}^2, \cdots, \sigma_{\mathrm{n},K}^2 = \sigma_{\mathrm{n}}^2$.

At the receiver side, each user sequentially decodes the common stream and the intended private stream to recover its message. User-k first decodes the common stream by treating the interference from all private streams as noise. Hence, the

signal-to-interference-noise ratio (SINR) of decoding s_c at user-k is expressed as

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \mathbf{p}_c\right|^2}{\sum_{i\in\mathcal{K}} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_{\mathrm{n}}^2}. \tag{2.3}$$

After successfully decoding and removing the common stream using SIC[1], user-k decodes its own private stream by treating the private streams of other users as noise. By considering perfect SIC, the SINR of decoding s_k at user-k is expressed as

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \mathbf{p}_k\right|^2}{\sum_{i\in\mathcal{K}, i\neq k} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_{\mathrm{n}}^2}. \tag{2.4}$$

Under the assumption of Gaussian signalling and infinite block length, the achievable rates for decoding the common and private streams at user-k are respectively $R_{c,k} = \log_2(1+\gamma_{c,k})$ and $R_k = \log_2(1+\gamma_k)$. To ensure the common stream s_c is successfully decoded by all users, its rate cannot exceed $R_c = \min_{k\in\mathcal{K}} R_{c,k}$. Since s_c contains messages $W_{c,1}, \cdots, W_{c,K}$ of the K users, let C_k denote the portion of rate R_c allocated to user-k for $W_{c,K}$. Then, we have $R_c = \sum_{k\in\mathcal{K}} C_k$. As a consequence, the overall achievable rate of user-k is writtn as $R_{k,tot} = C_k + R_k$.

The beamforming matrix $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_K]$ can be designed using low-complexity methods, such as zero-forcing beamforming (ZFBF) for the private streams and multicast precoders (e.g., dominant singular vector of concatenated channel matrix) for the common stream. Alternatively, the precoders can be optimized via different objectives, e.g., maximizing the weighted sum-rate (WSR) [5, 16], maximizing the energy efficiency (EE) [17], etc.

It should be noted that in this section, the fundamentals of downlink RSMA is introduced based on one-layer RSMA, the simplest and most practical RSMA implementation [2], and will be utilized throughout this thesis. Interested readers

[1]Throughout this thesis, perfect CSIR is assumed, where the common stream can be removed perfectly by SIC. For imperfect CSIR, please see [15].

are referred to [2] for a more comprehensive study on the other forms of RSMA. One-layer RSMA requires only one layer of SIC at each receiver. User grouping and ordering are not required since each user decodes the common stream before decoding its private stream. Compared with the generalized RSMA elaborated in [5], which involves multiple common streams and requires multiple SIC layers at the receivers, the encoding complexity, scheduling complexity and receiver complexity are reduced tremendously. Results in [5] show that the low complexity one-layer RSMA has a comparable rate performance to the generalized RSMA. The advantage of complexity reduction becomes more significant when the user number increases. Thanks to the inherent message splitting capability which is not featured in any other MA schemes, RSMA allows to:

1) partially decode interference and partially treat interference as noise (hence its efficiency, flexibility, reliability, and resilience),

2) reconcile the two extreme strategies of interference management and multiple MA schemes into a single framework (hence its generality/universality),

3) achieve optimal DoF in practical scenarios subject to imperfect CSIT [1].

In the literature, the benefits achieved by RSMA have been investigated in a wide range of multi-antenna scenarios, namely multiuser unicast transmission with perfect CSIT [5,9,18,19], imperfect CSIT [16,20–25], multigroup multicast transmission [26–29], as well as superimposed unicast and multicast transmission [17], etc. According to the analysis and simulations, [5] shows that RSMA is more robust to the influencing factors such as channel disparity, channel orthogonality, network load, and quality of CSIT. For imperfect CSIT, the sum-DoF and MMF-DoF of underloaded MU-MISO system are studied in [16] and [21]. RSMA is demonstrated to further exploit spatial dimensions. The superior performance of RSMA can also be seen in massive MIMO systems with residual transceiver hardware impairments [30],

mmWave communications [31] and simultaneous wireless information and power transfer (SWIPT) networks [17], etc. The performance gains of RSMA are not just limited to the assumption of Gaussian signalling and infinite block lengths, but are realized for practical systems as well in throughput performance through link-level simulations [32, 33].

2.2 Multibeam Satellite Systems

Satellite communications, appealing for its ubiquitous coverage, will play a key role in the next generation of wireless communications [34]. It not only provides connectivity in unserved areas but also decongests dense terrestrial networks. In recent years, multibeam satellite communication systems have received considerable research attention due to the full frequency reuse across multiple narrow spot beams towards higher throughput [11, 12]. Multibeam satellites are equipped with multiple antenna feeds and serve multiple user groups within multiple co-channel beams. Since the available spectrum is aggressively reused, interference management techniques become particularly important. Based on state-of-the-art technologies in DVB-S2X [35], each spot beam of the satellite serves more than one user simultaneously by transmitting a single coded frame. Multiple users within the same beam share the same precoding vector. Since different beams illuminate different groups of users [36], this promising multibeam multicasting follows the physical layer (PHY) multigroup multicast transmission.

Multicast transmission is at first considered in [37] with a single-group setup. Then, the problem is extended to multigroup multicast in [38] where the beamforming design is investigated in two optimization perspectives, namely the QoS-constrained transmit power minimization (QoS problem) and the power-constrained max-min fairness (MMF) problem. Both formulations are shown to be NP-hard, containing the

multiuser unicast and the single-group multicast as extreme cases. The combination of semi-definite relaxation (SDR) and Gaussian randomization, together with the bisection search algorithm are elaborated to generate feasible approximate solutions. Alternatively, a convex-concave procedure (CCP) [39] algorithm is demonstrated to provide better performance. However, its complexity increases dramatically as the problem size grows. In [40], a low-complexity algorithm for multigroup multicast beamforming based on alternating direction method of multipliers (ADMM) together with CCP is proposed for large-scale wireless systems. Moreover, the multigroup multicast beamforming is extended to many other scenarios, including the per-antenna power constraint addressed in [41], Cloud-radio access network (RAN) with wireless backhaul [42], coordinated beamforming in multi-cell networks [43], cache aided networks [44] and massive MIMO [45]. Since one practical application of multigroup multicast is found in multibeam satellite communication systems, in the literature of multibeam satellite systems, a generic iterative algorithm is proposed in [46] to design the precoding and power allocation alternatively in a TDM scheme considering a single user per beam. Then, multibeam multicast is considered. [35] proposes a frame-based precoding problem for multibeam multicast satellites. Optimization of the system sum rate is considered under individual power constraints via an alternating projection technique with an SDR procedure, which is adequate for small to medium-coverage areas. In [47], a two-stage low complex beamforming design for multibeam multicast satellite systems is proposed. The first stage minimizes inter-beam interference, while the second stage enhances intra-beam SINR. [36] studies the sum rate maximization problem in multigateway multibeam satellite systems considering feeder link interference. Leakage-based minimum mean square error (MMSE) and successive convex approximation (SCA)-ADMM algorithms are used to compute beamforming vectors locally with limited coordination.

All aforementioned works rely on the conventional multigroup/multibeam multicast

linear precoding (denoted as SDMA in this thesis). Each user decodes its desired stream while treating all the interference streams as noise. The advantage of this conventional scheme lies in exploiting the spatial degrees of freedom provided by multiple antennas using low-complexity transmitter-receiver architecture. However, its effectiveness severely depends on the network load and the quality of CSIT. Since the precoders are designed based on the channel knowledge, CSIT inaccuracy can result in an inter-group interference problem which is detrimental to the system performance. Another limitation is that the SDMA is able to eliminate inter-group interference only when the number of transmit antennas is sufficient. Otherwise, it fails to do so in overloaded systems [26]. For example, rate saturation occurs in overloaded systems. Departing from SDMA, the employment of RSMA in multi-group multicast beamforming is at first proposed in [26]. The key of RSMA-based multigroup multicast beamforming is to divide each group-intended message into a common part and a private part. An RSMA-based MMF problem was formulated and solved by the WMMSE approach [48]. The superiority of RSMA with perfect CSIT is shown in overloaded multigroup multicast systems.

2.3 Satellite-Terrestrial Integrated Networks

In recent years, due to the explosive growth of wireless applications and multimedia services, STIN has gained a tremendous amount of attention in both academia and industry as it can provide ubiquitous coverage and convey rich multimedia services, e.g., video on demand (VoD) streaming and TV broadcasting, etc. to users in both densely and sparsely populated areas [49]. The integration of terrestrial and satellite networks is of great potential in achieving geographic coverage, especially for remote areas where no terrestrial BS infrastructure can be employed [50, 51]. It is envisaged that the C-band ($4-8$ GHz) and S-band ($2-4$ GHz) can be shared between the

terrestrial and satellite networks. In addition, Ka band from 20 GHz to 40 GHz is foreseen to be the most promising candidate radio band for the next generation terrestrial cellular networks, and part of this band has already been allocated to the satellite networks [52].

A number of research efforts have investigated STIN systems. A coexistence framework of the satellite and terrestrial network is presented in [53] with the satellite link as primary and the terrestrial link as secondary. Transmit beamforming techniques are studied to maximize the SINR towards terrestrial users and minimize the interference towards satellite users. [54] considers a time division cooperative STIN, where a weighted MMF problem was formulated to jointly optimize the beamforming of BSs and the satellite. A multicast beamforming STIN system is investigated in [55] with the aim to maximize the sum of user minimum ratio under constraints of backhaul links and QoS. The authors generally assume the satellite channels as Rician channels. The effects of satellite antenna gain, path loss and atmospheric attenuation can be taken into account to model more practical satellite channels so as to evaluate the system performance more accurately. In this regard, [52] investigates a joint beamforming scheme for secure communication of STIN operating in mmWave frequencies. [56] focuses on the joint optimization for wireless information and power transfer (WIPT) technique in STINs. In [57], the cache-enabled low Earth orbit (LEO) satellite network is introduced, and the scheme of STIN is proposed to enable an energy-efficient RAN by offloading traffic from BSs through satellite broadcast transmission.

The above works consider conventional linear precoding and assume perfect CSIT. Each user decodes its desired stream while treating all the other interference streams as noise. The spatial degrees of freedom provided by multiple antennas are exploited, however, the effectiveness of beamforming design relies on the accuracy of CSIT significantly. In the real satellite communication environment, one practical issue is

that accurate CSI is very difficult to acquire at the GW because of the long-distance propagation delay and device mobility. Thus, robust design in the presence of imperfect CSIT has been widely studied in the literature [58–63]. [58–60] assume the satellite channel uncertainty as additive estimation error located in a bounded error region. Robust beamforming is designed based on the optimization of the worst-case situation. Yet, due to the special characteristics of satellite channels, the channel magnitude does not vary significantly due to the fact that the channel propagation is dominated by the line-of-sight component. The phase variations constitute the major source of channel uncertainty [11]. Therefore, in [61–63], beamforming is studied when considering constant channel amplitudes within the coherence time interval and independent time-varying phase components. Considering the phase-blind scenario, the achievable rate performance of RSMA in an multiuser MISO network is investigated in [64]. Apart from the difficulties in acquiring perfect CSIT, another consideration is the frame-based structure of multibeam satellite standards such as DVB-S2X [65]. Each spot beam of the satellite serves more than one user simultaneously by transmitting a single coded frame. Multiple users within the same beam share the same beamforming vector. Such multibeam multicast transmission is a promising solution for the rapidly growing content-centric applications including video streaming, advertisements, large-scale system updates and localized services, etc.

More recently, the use of RSMA in multibeam satellite and integrated satellite systems has been investigated. [25] studies RSMA in a two-beam satellite system adopting TDM in each beam. [66] focuses on the sum rate optimization and low complexity RSMA precoding design by decoupling the design of common stream and private streams. [67, 68] propose a RSMA-based multibeam multicast beamforming scheme and formulate an MMF problem with different CSIT qualities. In [69], RSMA is proven to be promising for multigateway multibeam satellite systems with feeder link interference. [70] considers a satellite and aerial integrated network

comprising a satellite and an unmanned aerial vehicle (UAV). The satellite employs multicast transmission, while the UAV uses RSMA to improve spectral efficiency. In [71], a secure beamforming scheme for STIN is presented, where the satellite serves one earth station (ES) with K eavesdroppers (Eves). RSMA is employed at the BS to achieve higher spectral efficiency. A robust beamforming scheme is proposed to maximize the secrecy energy efficiency of the ES considering Euclidean norm bounded channel uncertainty.

2.4 Integrated Sensing and Communications

ISAC has been envisioned as a key technique for future 6G wireless networks to fulfil the increasing demands on high-quality wireless connectivity as well as accurate and robust sensing capability [14]. ISAC merges wireless communications and remote sensing into a single system, where both functionalities are combined via shared use of the spectrum, the hardware platform, and a joint signal processing framework. ISAC systems are typically categorized into three types: radar-centric design, communication-centric design, and joint beamforming design [72]. This thesis will only focus on the joint beamforming design of ISAC rather than relying on existing radar or communication waveforms [73, 74].

ISAC has been considered in several promising terrestrial applications, including autonomous vehicles, human activity monitoring, indoor positioning, etc [75–77]. In [78], a novel framework is proposed for the transmit beamforming of the joint multi-antenna radar-communication (RadCom) system. The precoders are designed to formulate an appropriate desired radar beampattern, while guaranteeing the SINR requirements of the communication users. The authors in [79] propose the joint waveform design such that the multiuser interference is minimized while formulating a desired radar beampattern. [72] investigates the joint waveform design with

emphasis on optimizing the target estimation performance, measured by Cramér-Rao bound (CRB) considering both point and extended target scenarios.

Inspired by the advantages of RSMA in spectral efficiency, energy efficiency, user fairness, reliability, and QoS, performance enhancements in a wide range of network loads (underloaded and overloaded) and channel conditions, etc, the interplay between RSMA and ISAC systems is proposed in [80, 81], which demonstrates the benefits of an RSMA-assisted ISAC system with the objective of jointly maximising WSR and minimising mean square error (MSE) of beampattern approximation considering the per-antenna power constraint. As a step further, RSMA-assisted ISAC is studied in [82] considering partial CSIT and mobility of communication users as a practical application. RSMA is shown to better manage the interference and improve the trade-off between WSR and MSE of beampattern approximation compared with other MA strategies such as SDMA and NOMA. The design of RSMA-assisted ISAC with low resolution digital-to-analog converter (DAC) units is introduced in [83], where RSMA is shown to achieve improved energy efficiency by employing a smaller number of RF chains, owing to its generalized structure and improved interference management capabilities.

Chapter 3

RSMA for Multigroup Multicast and Multibeam Satellite Systems

This chapter is concerned with RSMA and its beamforming design problem to achieve MMF among multiple co-channel multicast groups with imperfect CSIT. Contrary to the conventional SDMA for multigroup multicast that relies on linear precoding and fully treating any residual interference as noise, we consider a novel multigroup multicast beamforming strategy based on RSMA. We characterize the MMF-DoF achieved by RSMA and SDMA in multigroup multicast with imperfect CSIT and demonstrate the benefits of RSMA for both underloaded and overloaded scenarios. Motivated by the DoF analysis, we then formulate a generic transmit power constrained optimization problem to achieve MMF rate performance. PHY layer design and link-level simulations are also investigated. The superiority of RSMA-based multigroup multicast beamforming compared with conventional schemes is demonstrated via simulations in both terrestrial multigroup multicast and multibeam satellite systems. In particular, due to the characteristics and challenges of multibeam satellite systems, the proposed RSMA-based strategy is shown promising to manage its inter-beam interference.

3.1 Introduction

With the proliferation of mobile data and multimedia traffic, demands for massive connectivity and content-centric services are continuously rising. Examples include audio/video streaming, advertisements, large-scale system updates, localized services and downloads, etc. Spurred by such requirements, wireless multicasting has attracted widespread research attention. It is a promising solution to deliver the same message to a group of recipients. In a more general scenario, which is known as multigroup multicasting, distinct contents are simultaneously transmitted to multiple co-channel multicast groups. Since the available spectrum is aggressively reused towards spectrum efficient and high throughput wireless communications, advanced interference mitigation techniques are of particular importance.

In this chapter, motivated by exploring the benefits of RSMA for multigroup multicast beamforming, we consider both underloaded and overloaded regimes with imperfect CSIT and its application to multibeam satellite systems. The main contributions are as follows:

- First, the MMF-DoF of RSMA and SDMA in multigroup multicast with imperfect CSIT is characterized. The MMF-DoF, also known as the MMF multiplexing gain, indicates the maximum multiplexing gain that can be simultaneously achieved by all multicast groups. It reflects the pre-log factor of MMF-rate at high SNR. This is the *first work* on DoF analysis for multigroup multicast in the presence of imperfect CSIT. In [26], MMF-DoF gains of RSMA with perfect CSIT are only observed in overloaded systems. In this chapter with an imperfect CSIT setting, RSMA is shown to provide MMF-DoF gains in both underloaded and overloaded systems. Through residual interference and group partitioning analysis, RSMA is shown to be more flexible than SDMA to overcome the residual interference caused by imperfect CSIT. By

adjusting the common stream and private streams, we can determine how much interference to be decoded and how much to be treated as noise. Due to the existence of a common part, RSMA provides extra gains and avoids the saturating performance at the high SNR regime.

- Second, motivated by the benefits of RSMA from a DoF perspective, an MMF beamforming optimization problem is formulated to investigate whether the DoF gain translates into rate gain. This is the *first work* on the optimization of RSMA-based multigroup multicast with imperfect CSIT. Solving the MMF problem with imperfect CSIT via sample average approximation (SAA) and WMMSE is for the *first time* studied. The optimum MMF Ergodic rate can be obtained by optimizing the defined short-term MMF Average rate over a long sequence of channel estimates. The formulated problem is general enough to cope with flexible power constraints, namely a total power constraint (TPC) and per-antenna power constraints (PAC). Through simulation results, the DoF benefits of RSMA translate into rate benefits at finite SNR and RSMA is shown to outperform SDMA in a wide range of setups. All the simulation results are inline with the derived theoretical MMF-DoFs results. Considering imperfect CSIT, we show that RSMA for multigroup multicast brings rate gains compared with SDMA in both underloaded and overloaded scenarios. This contrasts with the perfect CSIT setting of [26], where RSMA is shown to provide gains in the overloaded scenarios only.

- Third, the proposed RSMA framework is applied to a multibeam satellite setup and the results confirm the significant performance gains over conventional schemes. Based on state-of-the-art technologies in DVB-S2X, each spot beam of the satellite serves more than one user simultaneously by transmitting a single coded frame. This multibeam multicast scheme follows the PHY multigroup multicast transmission. Different from [84], which studies RSMA in a two-

beam satellite system adopting TDM scheme in each beam, and [85] which focuses on the sum rate optimization and low-complexity RSMA beamforming assuming perfect CSIT, we consider a novel RSMA-based multibeam multicast beamforming in this chapter and formulate a per-feed power constrained MMF problem. RSMA is shown very promising for multibeam satellite systems to manage inter-beam interference, taking into account practical challenges such as CSIT uncertainty, per-feed transmit power constraints, hot spots, uneven user distribution per beam, and overloaded regimes. Simulation results confirm the significant performance gains over conventional techniques.

- Fourth, the RSMA transmitter and receiver architecture, PHY layer and LLS platform are designed by considering finite length polar coding, finite alphabet modulation, AMC algorithm, etc. LLS results verify the effectiveness of RSMA-based multigroup multicast for practical implementation.

3.2 System Model

We consider a multigroup multicasting downlink MISO system. The transmitter is equipped with N_t antennas, serving K single-antenna users which are grouped into $M\,(1 \leq M \leq K)$ multicast groups. The users within each group desire the same multicast message. The messages are independent amongst different groups. Let $\mathcal{G}_m$ denote the set of users belonging to the m-th group, for all $m \in \mathcal{M} = \{1 \cdots M\}$. The size of group-$m$ is $G_m = |\mathcal{G}_m|$. We assume that each user belongs to only one group, thus $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$, for all $i, j \in \mathcal{M}$, $i \neq j$. Let $\mathcal{K} = \{1 \cdots K\}$ denote the set of all users, i.e., $\cup_{m \in \mathcal{M}} \mathcal{G}_m = \mathcal{K}$. In this model, the signal received at user-k writes as $y_k = \mathbf{h}_k^H \mathbf{x} + n_k$, $\forall k \in \mathcal{K}$, where $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal, $\mathbf{h}_k \in \mathbb{C}^{N_t \times 1}$ is the channel vector between the transmitter and the k-th user. $\mathbf{H} \triangleq [\mathbf{h}_1, \cdots, \mathbf{h}_K]$ is the composite channel. $n_k \sim \mathcal{CN}\left(0, \sigma_{n,k}^2\right)$ represents the AWGN at user-k, which

is independent and identically distributed (i.i.d) across users with zero mean and variance $\sigma_{n,k}^2$. Without loss of generality, unit noise variances are assumed, i.e., $\sigma_{n,k}^2 = \sigma_n^2 = 1$.

3.2.1 Transceiver Scheme

The application of RSMA to multigroup multicasting is described as follows. We assume that there are M messages $W_1, \cdots, W_M$ intended to users in $\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_M$ respectively. Each message is split into a common part and a private part, i.e., $W_m \rightarrow \{W_{m,c}, W_{m,p}\}$. All the common parts are packed together and encoded into a common stream shared by all groups, i.e., $\{W_{1,c} \cdots W_{M,c}\} \rightarrow s_c$, while the private parts are encoded into private streams for each group independently, i.e., $W_{m,p} \rightarrow s_m$. As a consequence, the vector of symbol streams to be transmitted is $\mathbf{s} = [s_c, s_1, \cdots, s_M]^T \in \mathbb{C}^{(M+1)\times 1}$, where $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}_{M+1}$. Data streams are then mapped to transmit antennas through a beamforming matrix $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots \mathbf{p}_M] \in \mathbb{C}^{N_t \times (M+1)}$. This yields a transmit signal $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ given by

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{p}_c s_c + \sum_{m=1}^{M} \mathbf{p}_m s_m, \tag{3.1}$$

where $\mathbf{p}_c \in \mathbb{C}^{N_t \times 1}$ is the common precoder, and $\mathbf{p}_m \in \mathbb{C}^{N_t \times 1}$ is the m-th group's precoder. Moreover, flexible transmit power constraints are considered in this work, including a total power constraint and per-antenna power constraints. Since the average power of transmit symbols are normalized to be one, the expression of a general transmit power constraint writes as

$$\mathbf{p}_c^H \mathbf{D}_l \mathbf{p}_c + \sum_{m=1}^{M} \mathbf{p}_m^H \mathbf{D}_l \mathbf{p}_m \leq P_l, \quad l = 1 \cdots L, \tag{3.2}$$

where L is the number of power constraints. P_l is the l-th power limit, and $\mathbf{D}_l$ is a diagonal shaping matrix changing among different demands. In particular, when the

focus is on a total transmit power constraint, let $L = 1$, $\mathbf{D}_l = \mathbf{I}$ and $P_l = P > 0$, from which P equals to the transmit SNR. However, in some practical implementations, using individual amplifiers per-antenna causes a lack of flexibility in sharing energy resources. Such a scenario is typically found in multibeam satellite communications because flexible on-board payloads are costly and complex to implement. Per-antenna available power constraints are taken into account by setting $L = N_t$, and $P_l = P/N_t$. The matrix D_l becomes a zero matrix except its l-th diagonal element equaling 1. Then, we define $\mu : \mathcal{K} \to \mathcal{M}$ as mapping a user to its corresponding group. The signal received at user-k can be expanded as

$$y_k = \mathbf{h}_k^H \mathbf{p}_c s_c + \mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)} + \mathbf{h}_k^H \sum_{j=1, j\neq\mu(k)}^{M} \mathbf{p}_j s_j + n_k, \tag{3.3}$$

where $\mu(k)$ is the group index of user-k. Each user at first decodes the common stream s_c and treats M private streams as noise. The SINR of decoding s_c at user-k is

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \mathbf{p}_c\right|^2}{\left|\mathbf{h}_k^H \mathbf{p}_{\mu(k)}\right|^2 + \sum_{j=1, j\neq\mu(k)}^{M} \left|\mathbf{h}_k^H \mathbf{p}_j\right|^2 + \sigma_n^2}. \tag{3.4}$$

Its corresponding achievable rate writes as $R_{c,k} = \log_2(1 + \gamma_{c,k})$. To guarantee that each user is capable of decoding s_c, we define a common rate R_c at which s_c is communicated

$$R_c \triangleq \min_{k\in\mathcal{K}} R_{c,k}. \tag{3.5}$$

Note that s_c is shared among groups such that $R_c \triangleq \sum_{m=1}^{M} C_m$, where C_m corresponds to group-m's portion of common rate. After the common stream s_c is decoded and removed through SIC, each user then decodes its desired private stream by treating all the other interference streams as noise. The SINR of decoding $s_{\mu(k)}$ at user-k is given by

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \mathbf{p}_{\mu(k)}\right|^2}{\sum_{j=1, j\neq\mu(k)}^{M} \left|\mathbf{h}_k^H \mathbf{p}_j\right|^2 + \sigma_n^2}. \tag{3.6}$$

Its corresponding achievable rate is $R_k = \log_2(1+\gamma_k)$. In terms of group-m, the multicast information s_m should be decoded by all users in $\mathcal{G}_m$. Thus, the shared information rate r_m is determined by the weakest user in $\mathcal{G}_m$ and defined as

$$r_m = \min_{i \in \mathcal{G}_m} R_i. \tag{3.7}$$

The m-th group-rate is composed of C_m and r_m, and writes as

$$r_{g,m}^{RS} = C_m + r_m = C_m + \min_{i \in \mathcal{G}_m} R_i. \tag{3.8}$$

In addition, the conventional linear precoding (SDMA) for multigroup multicast is revisited. Unlike RSMA, information intended for each group is encoded directly to a single stream, i.e., $W_m \to s_m,\ \forall m \in \{1 \cdots M\}$, rather than splitting into a common part and private part. The symbol vector to be transmitted is $\mathbf{s} = [s_1, \cdots, s_M]^T \in \mathbb{C}^{M\times 1}$, where $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H = \mathbf{I}\right\}$. At the receiver side, each user decodes its desired stream and treats all the interference streams as noise. Following the same multicast logic as (3.7), the m-th group rate writes as

$$r_{g,m}^{SDMA} = r_m = \min_{i \in \mathcal{G}_m} R_i. \tag{3.9}$$

Through the description above, RSMA is a more general scheme[1] which encompasses SDMA as a special case by allocating all transmit power to the private streams.

Remark 3.1: The encoding complexity and receiver complexity of RSMA are slightly higher than SDMA. For one-layer RSMA in a M-group multigoup multicast MISO BC, $M+1$ streams need to be encoded in contrast to M streams for SDMA. One-layer RSMA requires one SIC at each user while SDMA does not require any SIC.

[1]RSMA is also a more general framework that encompasses NOMA as a special case [5,9,17,26]. Since NOMA leads to a waste of spatial resources and multiplexing gain/DoF (and therefore rate loss) in multi-antenna settings at the additional expense of large receiver complexity, as demonstrated extensively in [5,9], we do not compare with NOMA in this work.

3.2.2 CSIT Uncertainty and Scaling

Imperfect CSIT is considered in this work while the channel state information at the receiver (CSIR) is assumed to be perfect[2]. To model CSIT uncertainty, channel matrix $\mathbf{H}$ is denoted as the sum of a channel estimate $\widehat{\mathbf{H}} \triangleq \left[\widehat{\mathbf{h}}_1, \cdots, \widehat{\mathbf{h}}_K\right]$ and a CSIT error $\widetilde{\mathbf{H}} \triangleq \left[\widetilde{\mathbf{h}}_1, \cdots, \widetilde{\mathbf{h}}_K\right]$, i.e. $\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}$. CSIT uncertainty can be characterized by a conditional density $f_{\mathbf{H}|\widehat{\mathbf{H}}}\left(\mathbf{H} \mid \widehat{\mathbf{H}}\right)$ [16]. Taking each channel vector separately, the CSIT error variance $\sigma_{e,k}^2 \triangleq \mathbb{E}_{\widetilde{\mathbf{h}}_k}\left\{\left\|\widetilde{\mathbf{h}}_k\right\|^2\right\}$ is allowed to decay as $O\left(P^{-\alpha_k}\right)$ [16, 21, 87, 88], where $\alpha_k \in [0, \infty)$ is the scaling factor which quantifies CSIT quality of the k-th user. Equal scaling factors among users are assumed for simplicity in this model, i.e., $\alpha_k = \alpha$. For a finite non-zero α, CSIT uncertainty decays as P grows, (e.g., by increasing the number of feedback bits). In extreme cases, $\alpha = 0$ corresponds to a non-scaling CSIT, (e.g., with a fixed number of feedback bits). $\alpha \to \infty$ represents perfect CSIT, (e.g., with infinite number of feedback bits). The scaling factor is truncated such that $\alpha \in [0, 1]$ in this context since $\alpha = 1$ corresponds to perfect CSIT in the DoF sense [16, 21].

3.3 Max-Min Fair DoF Analysis

The MMF-DoFs of RSMA and SDMA are investigated in this section to characterize the performance of both schemes. The MMF-DoF, also named MMF multiplexing gain or symmetric multiplexing gain, corresponds to the maximum multiplexing gain that can be simultaneously achieved across multicast groups. It reflects the pre-log factor of MMF-rate at high SNR. The larger MMF-DoF is, the faster MMF-rate increases with SNR. One would therefore like to use communication schemes with the largest possible DoF. Motivated by mitigating interference at receivers, the beamforming used in this section is sufficient from the DoF perspective since DoF

[2]For imperfect CSIR, please see [15, 86].

can be roughly interpreted as the number of interference-free streams simultaneously communicated in a single channel use [16, 26].

3.3.1 Max-Min Fair DoF of SDMA

We start from SDMA, and define the k-th user-DoF as $D_k \triangleq \lim_{P\to\infty} \frac{R_k(P)}{\log_2(P)}$. The m-th group-DoF is given by $d_m^{SDMA} \triangleq \lim_{P\to\infty} \frac{r_{g,m}^{SDMA}(P)}{\log_2(P)} = \min_{i\in\mathcal{G}_m} D_i$, and $d^{SDMA} \triangleq \min_{m\in\mathcal{M}} d_m^{SDMA}$ is achieved by all groups. For a given beamforming matrix $\mathbf{P} = [\mathbf{p}_1, \cdots \mathbf{p}_M] \in \mathbb{C}^{N_t\times M}$, d^{SDMA} represents the MMF-DoF. It interprets the maximum fraction of an interference-free stream that can be simultaneously communicated amongst groups. Since each user is equipped with only one antenna, we have

$$d^{SDMA} \leq d_m^{SDMA} \leq D_i \leq 1, \quad \forall\ i \in \mathcal{G}_m,\ m \in \mathcal{M}. \tag{3.10}$$

Proposition 3.1. The optimum MMF-DoF achieved by SDMA is given by

$$d^{*SDMA} = \begin{cases} \alpha, & N_t \geq K - G_1 + 1 \\ \dfrac{\alpha}{2}, & K - G_M + 1 \leq N_t < K - G_1 + 1 \\ 0, & 1 \leq N_t < K - G_M + 1. \end{cases} \tag{3.11}$$

The achievability of Proposition 3.1 is discussed as follows by providing at least one feasible beamforming matrix that achieves the DoF in (3.11). Next, results in (3.11) are derived as tight upper-bounds from the converse, which completes the proof of Proposition 3.1.

1) Achievability of Proposition 3.1:

To mitigate inter-group interference observed by each user, we aim to design the precoders such that $\widehat{\mathbf{h}}_k^H \mathbf{p}_m = 0,\ \forall\ m \in \mathcal{M},\ k \in \mathcal{K} \setminus \mathcal{G}_m$. Define $\widehat{\mathbf{H}}_m$ as the composite channel estimate of users in group-m, we have $\mathbf{p}_m \in \text{null}\big(\widehat{\mathbf{H}}_{\overline{m}}^H\big)$, where

$\widehat{\mathbf{H}}_{\overline{m}} \triangleq \left[\widehat{\mathbf{H}}_1, \cdots \widehat{\mathbf{H}}_{m-1}, \widehat{\mathbf{H}}_{m+1} \cdots, \widehat{\mathbf{H}}_M\right] \in \mathbb{C}^{N_t \times (K-G_m)}$ is a channel estimate matrix excluding $\widehat{\mathbf{H}}_m$. All the channel vectors are assumed to be independent. To satisfy $\dim\left(\text{null}\left(\widehat{\mathbf{H}}_{\overline{m}}^H\right)\right) \geq 1$, a minimum number of transmit antennas is required, as follows

$$N_t \geq K - G_m + 1. \tag{3.12}$$

(3.12) ensures sufficient N_t to place $\mathbf{p}_m$ in the null space of its unintended groups. Primary inter-group interference caused by the m-th precoder can be eliminated. Without loss of generality, group sizes are assumed in ascending order: $G_1 \leq G_2 \leq \cdots \leq G_M$. In an underloaded scenario, condition (3.12) has to hold for all $m \in \mathcal{M}$, and we rewrite it as

$$N_t \geq K - G_1 + 1. \tag{3.13}$$

When (3.13) is satisfied, the system is underloaded. Considering equal power allocation such that $\|\mathbf{p}_1\|^2 = \cdots = \|\mathbf{p}_M\|^2 = \frac{P}{M}$, the received signal of user-k and the scaling of received signal components are expressed by

$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)}}^{O(P)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{j=1, j\neq\mu(k)}^{M} \mathbf{p}_j s_j}^{O\left(P^{1-\alpha}\right)} + \overbrace{n_k}^{O\left(P^0\right)}. \tag{3.14}$$

The second term is named as residual interference caused by imperfect CSIT. All the primary inter-group interference $\widehat{\mathbf{h}}_k^H \sum_{j=1, j\neq\mu(k)}^{M} \mathbf{p}_j s_j$ has been eliminated. Since the channel state does not depend on P, we have $\|\mathbf{h}_k\|^2, \|\widehat{\mathbf{h}}_k\|^2 = O(1)$. The residual interference term scales as $O\left(P^{1-\alpha}\right)$, with the CSIT error variance decaying as $O\left(P^{-\alpha}\right)$. Note that when $\alpha = 1$, the residual interference is reduced to the noise level, and it corresponds to perfect CSIT from the DoF sense. When $\alpha \in [0, 1]$, γ_k scales as $O\left(P^{\alpha}\right)$, from which $D_k = \alpha$ at each user. For all $m \in \mathcal{M}$, $d_m^{SDMA} = \alpha$, thus the MMF-DoF $d^{SDMA} = \alpha$. When $N_t < K - G_1 + 1$, the system becomes overloaded. If reducing the spatial dimensions to $N_t < K - G_M + 1$, it is evident that

the inter-group interference caused by each precoder cannot be eliminated. Such scenario is identified as fully-overloaded [26], and its MMF-DoF collapses to 0. Next, we focus on the partially-overloaded in which $K - G_M + 1 \leq N_t < K - G_1 + 1$. We generally assume $N_t = K - G_x + 1$, where the group index $x \in (1, M]$. Following the logic of (3.12), primary inter-group interference caused by the $[x, M]$-th group can be nulled if the precoders are designed such that $\mathbf{p}_m \in \text{null}\big(\widehat{\mathbf{H}}_{\overline{m}}^H\big), \forall\ m \in [x, M]$. In addition, since $N_t = K - G_x + 1 > (K - G_x) - G_1 + 1$, the system excluding group-x can be regarded as underloaded. Thus, we design $\mathbf{p}_m \in \text{null}\big(\widehat{\mathbf{H}}_{\overline{m},\overline{x}}^H\big), \forall\ m \in \mathcal{M} \setminus x$ to remove inter-group interference among $\mathcal{M} \setminus x$. The beamforming directions described above can be concluded as

$$\mathbf{p}_m \in \begin{cases} \text{null}\big(\widehat{\mathbf{H}}_{\overline{m},\overline{x}}^H\big),\ \forall\ m \in [1, x) \\ \text{null}\big(\widehat{\mathbf{H}}_{\overline{m}}^H\big),\ \forall\ m \in [x, M]\,. \end{cases} \tag{3.15}$$

An example of power allocation is

$$\|\mathbf{p}_m\|^2 = \begin{cases} \dfrac{P^\beta}{M-1}, & \forall m \in \mathcal{M} \setminus x \\ P - P^\beta, & m \in x, \end{cases} \tag{3.16}$$

where $\beta \in [0, 1]$ is a power partition factor. User-k's received signal is given by

$$y_k = \begin{cases} \overbrace{\mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)}}^{O\left(P^\beta\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum\limits_{j=1, j\neq\mu(k), j\neq x}^{M} \mathbf{p}_j s_j}^{O\left(P^{\beta-\alpha}\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \mathbf{p}_x s_x}^{O\left(P^{1-\alpha}\right)} + \overbrace{n_k}^{O\left(P^0\right)},\ \forall\ k \in \mathcal{K} \setminus \mathcal{G}_x \\ \overbrace{\mathbf{h}_k^H \mathbf{p}_x s_x}^{O(P)} + \overbrace{\mathbf{h}_k^H \sum\limits_{j\in[1,x)} \mathbf{p}_j s_j}^{O\left(P^\beta\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum\limits_{i\in(x,M]} \mathbf{p}_i s_i}^{O\left(P^{\beta-\alpha}\right)} + \overbrace{n_k}^{O\left(P^0\right)},\ \forall\ k \in \mathcal{G}_x. \end{cases} \tag{3.17}$$

It is observed that $\mathcal{G}_x$ bear both residual interference and interference from groups $[1, x)$, while $\mathcal{K} \setminus \mathcal{G}_x$ see only residual interference. γ_k at user $k \in \mathcal{K} \setminus \mathcal{G}_x$ scales as

$O\left(P^{\beta+\alpha-1}\right)$, and γ_k at user $k \in \mathcal{G}_x$ scales as $O\left(P^{1-\beta}\right)$. Achieving max-min fair DoF requires the same DoF amongst groups. By setting $\beta = 1 - \frac{\alpha}{2}$, all users' SINRs scale as $O\left(P^{\frac{\alpha}{2}}\right)$. It turns out that $d_m^{SDMA} = \frac{\alpha}{2}$ for all $m \in \mathcal{M}$, and the MMF-DoF $d^{SDMA} = \frac{\alpha}{2}$ is achieved. Multiplexing gains are partially achieved. Importantly, such partially-overloaded scenario does not exist when the group sizes are equal.

2) Converse of Proposition 3.1:

Proposition 3.1 is further shown as a tight upper-bound for any feasible SDMA beamforming. Here, we generally assume the power allocation $\|\mathbf{p}_1\|^2, \cdots, \|\mathbf{p}_M\|^2$ scale as $O\left(P^{a_1}\right), \cdots, O\left(P^{a_M}\right)$, where $a_1, \cdots, a_M \in [0,1]$ are power partition factors. For each $m \in \mathcal{M}$, $\mathcal{I}_m \subset \mathcal{M}$ is defined as a group set with precoding vectors interfering with the m-th group, while $\mathcal{R}_m \subset \mathcal{M}$ is defined as a group set with precoding vectors that only cause residual interference to the m-th group. We define $\overline{a}_m \triangleq \max_{j\in\mathcal{I}_m} a_j$, and $\overline{\overline{a}}_m \triangleq \max_{j\in\mathcal{R}_m} a_j$. Note that $\overline{a}_m = 0$ for $\mathcal{I}_m = \emptyset$, and $\overline{\overline{a}}_m = 0$ for $\mathcal{R}_m = \emptyset$. For each $m \in \mathcal{M}$, there exists at least one user $k \in \mathcal{G}_m$ with SINR scaling as $O\big(P^{\min\left\{(a_m-\overline{a}_m)^+,\ \left(a_m-\overline{\overline{a}}_m+\alpha\right)^+\right\}}\big)$, since the received signal can be generally written as

$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)}}^{O(P^{a_m})} + \overbrace{\mathbf{h}_k^H \sum_{j\in\mathcal{I}_m} \mathbf{p}_j s_j}^{O\left(P^{\overline{a}_m}\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{i\in\mathcal{R}_m} \mathbf{p}_i s_i}^{O\left(P^{\overline{\overline{a}}_m-\alpha}\right)} + \overbrace{n_k}^{O\left(P^0\right)}. \tag{3.18}$$

According to the definition, we obtain an upper-bound for the achievable group-DoF

$$d_m^{SDMA} \leq \min\left\{ \left(a_m - \overline{a}_m\right)^+,\ \left(a_m - \overline{\overline{a}}_m + \alpha\right)^+ \right\}, \tag{3.19}$$

where $(\cdot)^+$ ensures DoF non-negativity. The achievable MMF-DoF of SDMA satisfies $d^{SDMA} \leq d_m^{SDMA}$ for all $m \in \mathcal{M}$. Next, we aim to derive its tight upper-bound d^{*SDMA} such that $d^{SDMA} \leq d^{*SDMA}$ for any feasible SDMA-based beamforming in different network load scenarios.

When the system is underloaded, it is obvious that $\mathcal{I}_m = \emptyset$ and $\mathcal{R}_m = \mathcal{M} \setminus m$ for all $m \in \mathcal{M}$. Accordingly, we have $\overline{a}_m = 0$ and $\overline{\overline{a}}_m = \max_{j\in\mathcal{M}\setminus m} a_j$. (3.19) can be rewritten as

$$d_m^{SDMA} \leq \min\left\{a_m,\ \left(a_m - \max_{j\in\mathcal{M}\setminus m} a_j + \alpha\right)^+\right\}. \tag{3.20}$$

From (3.20), we assume $a_m - \max_{j\in\mathcal{M}\setminus m} a_j + \alpha > 0$ because $a_m - \max_{j\in\mathcal{M}\setminus m} a_j + \alpha \leq 0$ limits d^{*SDMA} to 0. Then, $(\cdot)^+$ can be omitted. Since d^{SDMA} is upper-bounded by taking the average of any two group-DoFs, we have

$$\begin{aligned} d^{SDMA} &\leq \frac{d_1^{SDMA} + d_2^{SDMA}}{2} && (3.21)\\ &\leq \frac{\min\left\{a_1,\ a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha\right\} + \min\left\{a_2,\ a_2 - \max_{j\in\mathcal{M}\setminus 2} a_j + \alpha\right\}}{2} && (3.22)\\ &\leq \frac{a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha + a_2 - \max_{j\in\mathcal{M}\setminus 2} a_j + \alpha}{2} && (3.23)\\ &\leq \alpha. && (3.24) \end{aligned}$$

(3.23) follows from the fact that point-wise minimum is upper-bounded by any element in the set. (3.24) is obtained due to $a_1 \leq \max_{j\in\mathcal{M}\setminus 2} a_j$ and $a_2 \leq \max_{j\in\mathcal{M}\setminus 1} a_j$.

Next, we focus on the partially-overloaded scenario. It is sufficient to show that $d^{SDMA} \leq \frac{\alpha}{2}$ for $N_t = K - G_1$, as decreasing the number of antennas does not increase DoF. Since $N_t < K - G_1 + 1$, $\mathbf{p}_1$ leads to interference to at least one group. We denote such group index as m_1. Thus, we have $\mathcal{I}_{m_1} = 1$ and $\mathcal{R}_{m_1} = \mathcal{M} \setminus \{1, m_1\}$, i.e., $\overline{a}_{m_1} = a_1$ and $\overline{\overline{a}}_{m_1} = \max_{j\in\mathcal{M}\setminus\{1,m_1\}} a_j$. Recalling (3.19), $d_{m_1}^{SDMA}$ writes as

$$d_{m_1}^{SDMA} \leq \min\left\{\left(a_{m_1} - a_1\right)^+,\ \left(a_{m_1} - \max_{j\in\mathcal{M}\setminus\{1,m_1\}} a_j + \alpha\right)^+\right\}. \tag{3.25}$$

For group-1, it is obvious that $\mathcal{I}_1 = \emptyset$ and $\mathcal{R}_1 = \mathcal{M} \setminus 1$, i.e., $\overline{a}_1 = 0$ and $\overline{\overline{a}}_1 =$

$\max_{j\in\mathcal{M}\setminus 1} a_j$. Then, we have

$$d_1^{SDMA} \leq \min\left\{a_1,\ \left(a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha\right)^+\right\}. \tag{3.26}$$

By assuming $a_{m_1} - a_1 > 0$ and $a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha > 0$, the group-DoF $d_{m_1}^{*SDMA}$ and d_1^{*SDMA} are not limited to 0. $(\cdot)^+$ can be omitted in both inequalities. Since $a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha > 0$ leads to $a_{m_1} - a_1 < a_{m_1} - \max_{j\in\mathcal{M}\setminus\{1,m_1\}} a_j + \alpha$, (3.25) can be rewritten as $d_{m_1}^{SDMA} \leq a_{m_1} - a_1$. Following the same logic as (3.21), d^{SDMA} is upper-bounded by taking the average of d_1^{SDMA} and $d_{m_1}^{SDMA}$

$$\begin{aligned} d^{SDMA} &\leq \frac{d_1^{SDMA} + d_{m_1}^{SDMA}}{2} && (3.27)\\ &\leq \frac{\min\left\{a_1,\ a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha\right\} + a_{m_1} - a_1}{2} && (3.28)\\ &\leq \frac{a_1 - \max_{j\in\mathcal{M}\setminus 1} a_j + \alpha + a_{m_1} - a_1}{2} && (3.29)\\ &\leq \frac{\alpha}{2}. && (3.30) \end{aligned}$$

(3.29) is obtained because point-wise minimum is upper-bounded by any element in the set. (3.30) is obtained due to $a_{m_1} - \max_{j\in\mathcal{M}\setminus 1} a_j \leq 0$.

In a fully-overloaded scenario, it is sufficient to show that d^{SDMA} is upper-bounded by 0 for $N_t = K - G_M$, as further decreasing N_t does not increase DoF. In this case, we have $N_t < K - G_m + 1$ for all $m \in \mathcal{M}$. Each $\mathbf{p}_m$ causes interference to at least one group. Here, we assume $a_{m_2} = \max_{m\in\mathcal{M}} a_m$. The index of group seeing interference from $\mathbf{p}_{m_2}$ is denoted by m_3. Thus, d^{SDMA} is upper-bounded by

$$d^{SDMA} \leq d_{m_3}^{SDMA} \leq \min\left\{\left(a_{m_3} - a_{m_2}\right)^+,\ \left(a_{m_3} - \overline{\overline{a}}_{m_3} + \alpha\right)^+\right\} \leq \left(a_{m_3} - a_{m_2}\right)^+ = 0. \tag{3.31}$$

Combining the upper-bounds and achievability derived above, Proposition 3.1 is proved. When $\alpha = 1$, the results boil down to the Proposition 1 in [26] with perfect CSIT.

Remark 3.2: The basic difference between perfect and imperfect CSIT scenarios while analysing the DoF of SDMA is the existence of residual interference. For example, when we consider perfect CSIT [26], $N_t \geq K - G_m + 1$ ensures a sufficient number of transmit antennas to place the m-th precoder in the null space of all of its unintended groups. Inter-group interference caused by such precoder can be fully eliminated. However, considering imperfect CSIT here, only primary inter-group interference can be eliminated. At least one form of residual interference still exists.

From the above discussion, when the number of transmit antenna is greater than $K - G_1 + 1$, only residual interference will be seen by each user by controlling the beamforming directions and power allocation. Otherwise, the system becomes overloaded. Through beamforming and power control, the MMF-DoF does not collapse to zero directly as in multi-user unicast or equal-group multigroup multicast systems. When N_t drops below $K - G_1 + 1$, $M - 1$ groups can be regarded as underloaded, seeing only two forms of residual interference as given in the first equation of (3.17), while the remaining one group's received signal subspace is partially sacrificed. As a result, an MMF-DoF of $\frac{\alpha}{2}$ is achieved through power control. When N_t drops below $K - G_M + 1$, each multicast group sees interference from all of its unintended groups. The MMF-DoF drops to 0.

3.3.2 Max-Min Fair DoF of RSMA

In RSMA scheme, the m-th group-DoF writes as $d_m^{RS} \triangleq \lim_{P\to\infty} \frac{r_{g,m}^{RS}(P)}{\log_2(P)} = \min_{i\in\mathcal{G}_m} D_i + d_{c,m}$, where $d_{c,m} \triangleq \lim_{P\to\infty} \frac{C_m(P)}{\log_2(P)}$ is provided by the common rate portions. $d^{RS} \triangleq \min_{m\in\mathcal{M}} d_m^{RS}$ is the MMF-DoF for a given beamforming matrix $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots \mathbf{p}_M] \in \mathbb{C}^{N_t\times(M+1)}$. Thus, we have

$$d^{RS} \leq d_m^{RS} \leq D_i + d_{c,m} \leq 1, \quad \forall\ i \in \mathcal{G}_m,\ m \in \mathcal{M}. \tag{3.32}$$

Proposition 3.2. The optimum MMF-DoF achieved by RSMA is given by

$$d^{*RS} \geq \begin{cases} \frac{1-\alpha}{M}+\alpha, & \text{when } \ N_t \geq K-G_1+1,\ 0 \leq \alpha \leq 1 \\ \frac{1}{1+M-M_{\mathrm{R}}^*}, & \text{when } 1 \leq N_t < K-G_1+1,\ \frac{1}{1+M-M_{\mathrm{R}}^*} < \alpha \leq 1 \\ \alpha+\frac{1-(1+M-M_{\mathrm{R}}^*)\,\alpha}{M}, & \text{when } 1 \leq N_t < K-G_1+1,\ 0 \leq \alpha \leq \frac{1}{1+M-M_{\mathrm{R}}^*}. \end{cases} \tag{3.33}$$

Note that M_{R}^* is the maximum number of groups which can be regarded as underloaded, and served by RSMA when the system is overloaded. The inequality indicates that the results provided here are achievable, yet not necessarily optimum.

1) Achievability of Proposition 3.2:

When the system is underloaded, i.e., $N_t \geq K-G_1+1$, we design $\mathbf{p}_m \in \mathrm{null}\big(\widehat{\mathbf{H}}_{\overline{m}}^H\big)$, which follows the same logic as SDMA. The direction of $\mathbf{p}_c$ is chosen randomly. Consider the power allocation such that $\|\mathbf{p}_1\|^2 = \cdots = \|\mathbf{p}_M\|^2 = \frac{P^\delta}{M}$, and $\|\mathbf{p}_c\|^2 = P - P^\delta$, where $\delta \in [0,1]$ is a power partition factor. The received signal writes as

$$y_k = \overbrace{\mathbf{h}_k^H \mathbf{p}_c s_c}^{O(P)} + \overbrace{\mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)}}^{O\left(P^\delta\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{j=1, j\neq\mu(k)}^{M} \mathbf{p}_j s_j}^{O\left(P^{\delta-\alpha}\right)} + \overbrace{n_k}^{O\left(P^0\right)}. \tag{3.34}$$

It can be observed that s_c is firstly decoded at each user with SINR $\gamma_{c,k}$ scaling as $O\left(P^{1-\delta}\right)$. The common stream can provide a DoF of $1-\delta$. Since $R_c = \sum_{m=1}^{M} C_m$, sharing R_c equally amongst groups leads to max-min fairness, and $d_{c,m} = \frac{1-\delta}{M}$ is achieved by each group. After removing s_c, each user then decodes $s_{\mu(k)}$ with γ_k scaling as $O\left(P^{\min\{\alpha,\delta\}}\right)$. For all $k \in \mathcal{K}$, we have $D_k = \min\{\alpha, \delta\}$. Therefore, the MMF-DoF $d^{RS} = \min_{m\in\mathcal{M}} d_m^{RS} = \frac{1-\delta}{M} + \min\{\alpha,\delta\}$ can be achieved. By setting $\delta = \alpha$, d^{RS} reaches its maximum value at $\frac{1-\alpha}{M} + \alpha$.

Next, in overloaded scenarios, i.e., $1 \leq N_t < K - G_1 + 1$, we consider a special case of RSMA where groups are divided into two subsets, namely $\mathcal{M}_{\mathrm{R}} \subseteq \mathcal{M}$ and $\mathcal{M}_{\mathrm{C}} = \mathcal{M} \setminus \mathcal{M}_{\mathrm{R}}$. Specifically, $\mathcal{M}_{\mathrm{R}}$ is a subset which can be treated as underloaded and served by RSMA, while $\mathcal{M}_{\mathrm{C}}$ are the remaining groups and served by degraded beamforming. Based on this mixed scheme, messages are split such that $W_m \to \{W_{m,c}, W_{m,p}\}$ for all $m \in \mathcal{M}_{\mathrm{R}}$, and $W_m \to \{W_{m,c}\}$ for all $m \in \mathcal{M}_{\mathrm{C}}$. Such scheme leads to $\|\mathbf{p}_m\|^2 = 0$ for all $m \in \mathcal{M}_{\mathrm{C}}$. The size of $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{C}}$ are denoted by $M_{\mathrm{R}} = |\mathcal{M}_{\mathrm{R}}|$ and $M_{\mathrm{C}} = |\mathcal{M}_{\mathrm{C}}| = M - M_{\mathrm{R}}$ respectively. To gain insight into the subset partition, we define

$$N_L = \begin{cases} K - G_1 - \sum_{j=L+1}^{M} G_j + 1, \ L \in \{1, \cdots, M-1\} \\ K - G_1 + 1, \ L = M. \end{cases} \tag{3.35}$$

According to (3.12), N_L is the minimum number of transmitting antennas required to regard groups $\{1, \cdots, L\}$ as underloaded while disregarding all the remaining groups. Conversely, if N_t satisfies $N_L \leq N_t < N_{L+1}$, L is interpreted as the maximum number of M_{R}. We can define it as

$$M_{\mathrm{R}}^* = \begin{cases} M, \ N_t \geq N_M \\ L, \ N_L \leq N_t < N_{L+1}, \forall L \in \{1, \cdots, M-1\}. \end{cases} \tag{3.36}$$

For all $m \in \mathcal{M}_{\mathrm{R}}$, beamforming directions are designed as $\mathbf{p}_m \in \mathrm{null}\left(\widehat{\mathbf{H}}^H_{\{\overline{m}, \overline{\mathcal{M}_{\mathrm{C}}}\}}\right)$. $\mathbf{p}_c$'s direction is set randomly. Consider the power allocation $\|\mathbf{p}_m\|^2 = \frac{P^\delta}{M_{\mathrm{R}}}$ for all $m \in \mathcal{M}_{\mathrm{R}}$, and $\|\mathbf{p}_c\|^2 = P - P^\delta$, where $\delta \in [0, 1]$.

The received signal of user-k is written as

$$y_k = \begin{cases} \overbrace{\mathbf{h}_k^H \mathbf{p}_c s_c}^{O(P)} + \overbrace{\mathbf{h}_k^H \mathbf{p}_{\mu(k)} s_{\mu(k)}}^{O\left(P^{\delta}\right)} + \overbrace{\widetilde{\mathbf{h}}_k^H \sum_{j \in \mathcal{M}_{\mathrm{R}} \setminus \mu(k)} \mathbf{p}_j s_j}^{O\left(P^{\delta-\alpha}\right)} + \overbrace{\sigma_n^2}^{O\left(P^0\right)}, & \forall\, k \in \{\mathcal{G}_m \mid m \in \mathcal{M}_{\mathrm{R}}\} \\ \overbrace{\mathbf{h}_k^H \mathbf{p}_c s_c}^{O(P)} + \overbrace{\mathbf{h}_k^H \sum_{j \in \mathcal{M}_{\mathrm{R}}} \mathbf{p}_j s_j}^{O\left(P^{\delta}\right)} + \overbrace{\sigma_n^2}^{O\left(P^0\right)}, & \forall\, k \in \{\mathcal{G}_m \mid m \in \mathcal{M}_{\mathrm{C}}\}. \end{cases} \tag{3.37}$$

Firstly, s_c is decoded at each user by treating all the other streams as noise. $\gamma_{c,k}$ is observed to scale as $O\left(P^{1-\delta}\right)$ for $k \in \mathcal{K}$. Thus, the common stream achieves a DoF of $1-\delta$. Since the common rate $R_c = \sum_{m=1}^{M} C_m$ is divided amongst $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{C}}$, we introduce a fraction $z \in [0,1]$ of the common rate such that $\sum_{m \in \mathcal{M}_{\mathrm{R}}} C_m = z R_c$, and $\sum_{m \in \mathcal{M}_{\mathrm{C}}} C_m = (1-z) R_c$. This leads to $d_{c,m} = \frac{z(1-\delta)}{M_{\mathrm{R}}}$ for $m \in \mathcal{M}_{\mathrm{R}}$ and $d_{c,m} = \frac{(1-z)(1-\delta)}{M-M_{\mathrm{R}}}$ for $m \in \mathcal{M}_{\mathrm{C}}$. After removing s_c through SIC, it can be seen that γ_k scales as $O\left(P^{\min\{\alpha,\delta\}}\right)$ in the first subset $\mathcal{M}_{\mathrm{R}}$. Hence, we have $D_k = \min\{\alpha, \delta\}$ for all $k \in \{\mathcal{G}_m \mid m \in \mathcal{M}_{\mathrm{R}}\}$. The group-DoF d_m^{RS} is given by

$$d_m^{RS} = \begin{cases} \dfrac{z(1-\delta)}{M_{\mathrm{R}}} + \min\{\alpha, \delta\}, & \forall\, m \in \mathcal{M}_{\mathrm{R}} \\ \dfrac{(1-z)(1-\delta)}{M - M_{\mathrm{R}}}, & \forall\, m \in \mathcal{M}_{\mathrm{C}}. \end{cases} \tag{3.38}$$

To achieve max-min fairness, equal group-DoFs between $\mathcal{M}_{\mathrm{R}}$ and $\mathcal{M}_{\mathrm{C}}$ are required. On one hand, we assume $\delta \geq \alpha$, and the equation can be written as

$$\frac{z(1-\delta)}{M_{\mathrm{R}}} + \alpha = \frac{(1-z)(1-\delta)}{M - M_{\mathrm{R}}}. \tag{3.39}$$

Note that there are two variables δ and z on both sides of (3.39). Since the two variables cannot be solved simultaneously, we fix one variable to maximize at least one side of (3.39) while reserving the other variable on both sides. For example, let

$\delta = \alpha$ in this case, and then calculate the remaining variable z according to

$$\frac{z\,(1-\alpha)}{M_{\mathrm{R}}} + \alpha = \frac{(1-z)\,(1-\alpha)}{M - M_{\mathrm{R}}}. \tag{3.40}$$

$z = \frac{[1-(1+M-M_{\mathrm{R}})\alpha]M_{\mathrm{R}}}{(1-\alpha)M}$ is obtained. Substitute it into arbitrary side of (3.40), and the group-DoF $d_m^{RS} = \alpha + \frac{1-(1+M-M_{\mathrm{R}})\alpha}{M}$ for all $m \in \mathcal{M}$ is derived. Moreover, a corresponding condition $0 \leq \alpha \leq \frac{1}{1+M-M_{\mathrm{R}}}$ is obtained by considering $0 \leq z = \frac{[1-(1+M-M_{\mathrm{R}})\alpha]M_{\mathrm{R}}}{(1-\alpha)M} \leq 1$. The MMF-DoF is achieved as $d^{RS} = \min_{m\in\mathcal{M}} d_m^{RS} = \alpha + \frac{1-\left(1+M-M_{\mathrm{R}}^*\right)\alpha}{M}$, when $0 \leq \alpha \leq \frac{1}{1+M-M_{\mathrm{R}}^*}$.

On the other hand, we assume $\delta < \alpha$. The equation in (3.39) is rewritten as

$$\frac{z\,(1-\delta)}{M_{\mathrm{R}}} + \delta = \frac{(1-z)\,(1-\delta)}{M - M_{\mathrm{R}}}. \tag{3.41}$$

There are still two variables δ and z in (3.41). In this case, we can set $z = 0$ to maximize the right side of (3.41) and calculate δ according to

$$\delta = \frac{1-\delta}{M - M_{\mathrm{R}}}. \tag{3.42}$$

By substituting the solution $\delta = \frac{1}{1+M-M_{\mathrm{R}}}$ into arbitrary side of (3.42), the group-DoF $d_m^{RS} = \frac{1}{1+M-M_{\mathrm{R}}}$ for all $m \in \mathcal{M}$ is derived. Since $\delta = \frac{1}{1+M-M_{\mathrm{R}}} < \alpha$, we obtain the corresponding condition $\frac{1}{1+M-M_{\mathrm{R}}} < \alpha \leq 1$ for this case.

Above all, the achievable MMF-DoF of RSMA is summarized in Proposition 3.2. When $\alpha = 1$, such result boils down to the achievability of Proposition 3 in [26] with perfect CSIT. In overloaded scenarios, it is noteworthy that the d^{RS} with $\frac{1}{1+M-M_{\mathrm{R}}^*} < \alpha \leq 1$ is not a function of α and is the same as that achieved with perfect CSIT. Thus, one can relax the CSIT quality up to $\frac{1}{1+M-M_{\mathrm{R}}^*}$ without affecting the MMF-DoF. However, in the other case when $0 \leq \alpha \leq \frac{1}{1+M-M_{\mathrm{R}}^*}$, d^{RS} diminishes as the CSIT quality reduces.

2) Insight:

From (3.37), the interference seen by each user $k \in \{\mathcal{G}_m \mid m \in \mathcal{M}_{\mathrm{R}}\}$ after SIC scales as $O\left(P^{\delta-\alpha}\right)$. As discussed above, we have two assumptions, namely $\delta \geq \alpha$ and $\delta < \alpha$. When $\delta \geq \alpha$, this residual interference cannot be ignored. By setting the power partition factor $\delta \to \alpha$, we can reduce it to the noise level and at the same time increase $\gamma_{c,k}$ which scales as $O\left(P^{1-\delta}\right)$ for all $k \in \mathcal{K}$. To achieve max-min fairness, the common rate factor z is then managed to obtain equal group-DoFs among groups in $\mathcal{M}_R$ and $\mathcal{M}_C$. $0 \leq \alpha \leq \frac{1}{1+M-M_{\mathrm{R}}^*}$ is derived as a corresponding range of this case. The MMF-DoF reduces as α goes down. Otherwise, when $\delta < \alpha$, such interference is always at the noise level. By setting $z \to 0$, all the common rate R_c contributes to C_m, for all $m \in \mathcal{M}_C$. The RSMA scheme used by $\mathcal{M}_R$ boils down to SDMA. Meanwhile, the group-DoFs of all $m \in \mathcal{M}_{\mathrm{C}}$ are maximized. Then, we further manage the power partition factor δ to achieve max-min fairness amongst all groups. $\frac{1}{1+M-M_{\mathrm{R}}^*} < \alpha \leq 1$ is derived as the corresponding range. In this case, changing α will no longer affect MMF-DoF because the interference seen by each user $k \in \{\mathcal{G}_m \mid m \in \mathcal{M}_{\mathrm{R}}\}$ after SIC is always at the noise level. The MMF-DoF performance remains the same as that achieved with perfect CSIT. Such behaviour is not observed in partially-overloaded SDMA. It can be observed in (3.17) that the power of interference seen by each user $k \in \mathcal{K} \setminus \mathcal{G}_x$ and $k \in \mathcal{G}_x$ scales as $O\left(P^{1-\alpha}\right)$ and $O\left(P^{\beta}\right)$ respectively. α will always affect MMF-DoF as $O\left(P^{1-\alpha}\right)$ cannot be ignored unless considering perfect CSIT. To get more insight into the gains provided by RSMA over SDMA, we substitute (3.36) into (3.33) and yield (3.43).

By comparing (3.43) with (3.11), we can see that the achievable MMF-DoF of RSMA is always superior to d^{*SDMA}, and hence $d^{*RS} \geq d^{*SDMA}$ is guaranteed. The gain of RSMA over SDMA is $\frac{1-\alpha}{M}$ when the system is underloaded. Once $N_t \geq N_M$ is violated, the range of partially-overloaded SDMA $K - G_M + 1 \leq N_t < K - G_1 + 1$, (i.e., $N_{M-1} + G_1 \leq N_t < N_M$) locates within the range $N_{M-1} \leq N_t < N_M$. For any

$$d^{*RS} \geq \begin{cases} \dfrac{1-\alpha}{M}+\alpha, & \text{when } N_t \geq N_M \\ \dfrac{1}{2}, & \text{when } N_{M-1} \leq N_t < N_M,\ \dfrac{1}{2} < \alpha \leq 1 \\ \alpha + \dfrac{1-2\alpha}{M}, & \text{when } N_{M-1} \leq N_t < N_M,\ 0 \leq \alpha \leq \dfrac{1}{2} \\ \vdots & \\ \dfrac{1}{M-1}, & \text{when } N_2 \leq N_t < N_3,\ \dfrac{1}{M-1} < \alpha \leq 1 \\ \alpha + \dfrac{1-(M-1)\,\alpha}{M}, & \text{when } N_2 \leq N_t < N_3,\ 0 \leq \alpha \leq \dfrac{1}{M-1} \\ \dfrac{1}{M}, & \text{when } 1 \leq N_t < N_2,\ \dfrac{1}{M} < \alpha \leq 1 \\ \dfrac{1}{M}, & \text{when } 1 \leq N_t < N_2,\ 0 \leq \alpha \leq \dfrac{1}{M} \end{cases} \tag{3.43}$$

$0 \leq \alpha \leq 1$, the achievable MMF-DoF of RSMA is still greater than SDMA. Once N_t drops below $N_{M-1}+G_1$, by taking $N_t = N_{M-1}$ as an example, the number of transmit antenna is not sufficient to eliminate any inter-group interference through SDMA beamforming. d^{*SDMA} collapses to 0. For RSMA, d^{*RS} is kept by exploiting all the M_{R}^* streams and transmitting the remaining stream through degraded beamforming. This is carried on until RSMA reducing to a single-stream degraded beamforming. A single DoF is split amongst all the groups. Therefore, $d^{*RS} \geq \frac{1}{M} > 0$ is guaranteed. For the particular case where all the group sizes are equal, (i.e., $G_m = G,\ \forall m \in \mathcal{M}$), there is no partially-overloaded scenario in (3.11). When N_t drops below $K - G + 1$, d^{*SDMA} decreases from α to 0 directly. However, the expression of d^{*RS} remains the same as (3.43), which is always greater than $\frac{1}{M}$.

Remark 3.3: The obtained MMF-DoFs of different strategies are listed in Table 3.1, where the first row represents underloaded and the others are the results of overloaded systems. From the above discussion, the MMF-DoF analysis in the underloaded regime is similar when considering RSMA and SDMA. Each user sees only residual interference by managing the beamforming directions and power allocation. A gain of $\frac{1-\alpha}{M}$ is obtained by applying RSMA. Thus, we can conclude that in the presence of

imperfect CSIT, there is an MMF-DoF gain of RSMA over SDMA when the system is underloaded. This contrasts with perfect CSIT scenarios where both underloaded SDMA and RSMA can achieve full MMF-DoF of 1*. Overloaded RSMA is more challenging since both residual interference and group partitioning method should be considered. [26] considers a special case where the groups are partitioned into two subsets, namely* $\mathcal{M}_D \subseteq \mathcal{M}$ *which are served using SDMA, and* $\mathcal{M}_C \subseteq \mathcal{M} \setminus \mathcal{M}_D$ *served by degraded beamforming. The number of groups in* $\mathcal{M}_D$ *is set as the maximum number of groups that can be served by interference-free SDMA (i.e., achieving a group-DoF of* 1 *each). However, in this work considering imperfect CSIT, SDMA can no longer reach an MMF-DoF of* 1*. As shown in Table 3.1, the maximum achievable MMF-DoF is* α *when the system is underloaded, while RSMA outperforms SDMA slightly. Thus, we consider a different subset partitioning in this work where the groups are divided into* $\mathcal{M}_R \subseteq \mathcal{M}$ *and* $\mathcal{M}_C \subseteq \mathcal{M} \setminus \mathcal{M}_R$*. The number of groups in* $\mathcal{M}_R$ *is chosen as the maximum number of groups which can be served by RSMA and achieve an MMF-DoF of* $\frac{1-\alpha}{M} + \alpha$*.* $\mathcal{M}_C$ *is still served by degraded beamforming. Accordingly, from the results summarised in Table 3.1, RSMA is shown to provide MMF-DoF gains and outperform SDMA in overloaded systems.*

All the discussions above motivate the use of RSMA from a DoF perspective. However, DoF is an asymptotically high SNR metric. It remains to be seen whether the DoF gain translates into rate gain. To that end, the design of RSMA for rate maximization at finite SNR needs to be investigated. Beamforming schemes that achieve Proposition 3.1 and Proposition 3.2 are not necessarily optimum from an MMF-rate sense. Therefore, the beamforming directions, power allocation and rate partition can be elaborated by formulating MMF-rate optimization problems as we see in the next section. Importantly, the DoF analysis provides fundamental grounds, helps to draw insights into the performance limits of various strategies and guides the design of efficient strategies (rate-splitting in this case).

Table 3.1: Achievable MMF-DoF of different strategies

	Perfect CSIT [16]		**Imperfect CSIT [this chapter]**	
Strategy	SDMA	RSMA	SDMA	RSMA
$N_t \geq N_M$	1	1	α	$\frac{1-\alpha}{M} + \alpha$
$N_{M-1} + G_1 \leq N_t < N_M$[1]	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{\alpha}{2}$	$\begin{cases} \frac{1}{2}, & \frac{1}{2} < \alpha \leq 1 \\ \alpha + \frac{1-2\alpha}{M}, & 0 \leq \alpha \leq \frac{1}{2} \end{cases}$
$N_{M-1} \leq N_t < N_{M-1} + G_1$	0	$\frac{1}{2}$	0	$\begin{cases} \frac{1}{2}, & \frac{1}{2} < \alpha \leq 1 \\ \alpha + \frac{1-2\alpha}{M}, & 0 \leq \alpha \leq \frac{1}{2} \end{cases}$
$N_{M-2} \leq N_t < N_{M-1}$	0	$\frac{1}{3}$	0	$\begin{cases} \frac{1}{3}, & \frac{1}{3} < \alpha \leq 1 \\ \alpha + \frac{1-3\alpha}{M}, & 0 \leq \alpha \leq \frac{1}{3} \end{cases}$
$\vdots$ $1 \leq N_t < N_2$	0	$\frac{1}{M}$	0	$\vdots$ $\frac{1}{M}$

[1] The second line of this table (partially-overloaded scenario) does not exist when the group sizes are equal. When N_t drops below $K - G + 1$, d^{*SDMA} decreases to 0 directly.

3.4 Max-Min Fair Problem Formulation

Now, we formulate an optimization problem to design precoders to achieve MMF among multiple co-channel multicast groups subject to a flexible power constraint with imperfect CSIT. MMF Ergodic rate is the metric for both RSMA and SDMA. It reflects long-term MMF rate performance over varying channel states. Given a long sequence of channel estimates, the MMF Ergodic rate can be measured by updating precoders based on each short-term MMF Average rate. For a given channel estimate $\widehat{\mathbf{H}}$, the Average rate is defined as the expected performance over CSIT error distribution. The Average rates for the common and private streams of user-k are short-term measures given by

$$\overline{R}_{c,k}(\widehat{\mathbf{H}}) = \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\big\{R_{c,k}(\mathbf{H},\widehat{\mathbf{H}}) \mid \widehat{\mathbf{H}}\big\}, \tag{3.44}$$

$$\overline{R}_{k}(\widehat{\mathbf{H}}) = \mathbb{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\big\{R_{k}(\mathbf{H},\widehat{\mathbf{H}}) \mid \widehat{\mathbf{H}}\big\}. \tag{3.45}$$

Note that Average rates should not be confused with Ergodic rates. Ergodic rates capture the long-term performance over all channel states, while Average rates measure the short-term expected performance over CSIT error distribution for a given channel state estimate. According to the law of total expectation and the definition of Average rate, the Ergodic rates for the common and private streams of user-k are expressed by

$$\mathbb{E}_{\{\mathbf{H},\widehat{\mathbf{H}}\}}\big\{R_{c,k}(\mathbf{H},\widehat{\mathbf{H}})\big\} = \mathbb{E}_{\widehat{\mathbf{H}}}\big\{\mathbb{E}_{\{\mathbf{H}|\widehat{\mathbf{H}}\}}\big\{R_{c,k}(\mathbf{H},\widehat{\mathbf{H}}) \mid \widehat{\mathbf{H}}\big\}\big\} = \mathbb{E}_{\widehat{\mathbf{H}}}\{\overline{R}_{c,k}(\widehat{\mathbf{H}})\}, \tag{3.46}$$

$$\mathbb{E}_{\{\mathbf{H},\widehat{\mathbf{H}}\}}\big\{R_{k}(\mathbf{H},\widehat{\mathbf{H}})\big\} = \mathbb{E}_{\widehat{\mathbf{H}}}\big\{\mathbb{E}_{\{\mathbf{H}|\widehat{\mathbf{H}}\}}\big\{R_{k}(\mathbf{H},\widehat{\mathbf{H}}) \mid \widehat{\mathbf{H}}\big\}\big\} = \mathbb{E}_{\widehat{\mathbf{H}}}\{\overline{R}_{k}(\widehat{\mathbf{H}})\}. \tag{3.47}$$

It turns out that measuring Ergodic rates is transformed into measuring Average rates over the variation of $\widehat{\mathbf{H}}$. Therefore, the MMF Ergodic rate maximization problem is decomposed to an MMF Average rate maximization problem for each $\widehat{\mathbf{H}}$. The MMF Ergodic rate of RSMA can be characterized by $\mathbb{E}_{\widehat{\mathbf{H}}}\{\mathcal{F}_{RS}\}$, where $\mathcal{F}_{RS}$ is

the MMF Average rate maximization problem for a given channel estimate $\widehat{\mathbf{H}}$.

$$\mathcal{F}_{RS}: \quad \max_{\overline{\mathbf{c}},\mathbf{P}} \min_{m\in\mathcal{M}} \left(\overline{C}_m + \min_{i\in\mathcal{G}_m} \overline{R}_i\right) \tag{3.48}$$

$$s.t. \quad \overline{R}_{c,k} \geq \sum_{m=1}^{M} \overline{C}_m, \quad \forall k \in \mathcal{K} \tag{3.49}$$

$$\overline{C}_m \geq 0, \quad \forall m \in \mathcal{M} \tag{3.50}$$

$$\mathbf{p}_c^H \mathbf{D}_l \mathbf{p}_c + \sum_{m=1}^{M} \mathbf{p}_m^H \mathbf{D}_l \mathbf{p}_m \leq P_l, \quad l = 1 \cdots L \tag{3.51}$$

The average common rate vector $\overline{\mathbf{c}} = \left[\overline{C}_1, \cdots, \overline{C}_M\right]$ and the beamforming matrix $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots \mathbf{p}_M]$ are jointly optimized to achieve MMF performance. Since the average common rate is defined by $\overline{R}_c = \sum_{m=1}^{M} \overline{C}_m = \min_{k\in\mathcal{K}} \overline{R}_{c,k}$, we use constraint (3.49) to ensure that the common stream s_c is decoded by each user. Constraint (3.50) implies that each portion of $\overline{R}_c$ is non-negative and (3.51) is the transmit power constraint.

Similarly, the corresponding SDMA-based MMF Average rate maximization problem is formulated as

$$\mathcal{F}_{SDMA}: \quad \max_{\mathbf{P}} \min_{m\in\mathcal{M}} \left(\min_{i\in\mathcal{G}_m} \overline{R}_i\right) \tag{3.52}$$

$$s.t. \quad \sum_{m=1}^{M} \mathbf{p}_m^H \mathbf{D}_l \mathbf{p}_m \leq P_l, \quad l = 1 \cdots L \tag{3.53}$$

where the beamforming matrix $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_M]$ is optimized to solve $\mathcal{F}_{SDMA}$. (3.53) is the transmit power constraint. SDMA is a sub-scheme of RSMA by switching off (i.e., allocating zero power to) the common stream. Solving $\mathcal{F}_{SDMA}$ is a special case of $\mathcal{F}_{RS}$ by fixing $\overline{\mathbf{c}} = 0$ and $\|\mathbf{p}_c\|^2 = 0$. We will focus on solving the RSMA-based problem in the following discussion.

Sample average approximation (SAA) is then adopted to convert $\mathcal{F}_{RS}$ into a deterministic problem denoted by $\mathcal{F}_{RS}^{(S)}$. For a given channel estimate $\widehat{\mathbf{H}}$ and

sample index set $\mathfrak{S} \triangleq \{1, \cdots, S\}$, we construct S channel samples denoted as $\mathbb{H}^{(S)} \triangleq \left\{\mathbf{H}^{(s)} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}^{(s)} \mid \widehat{\mathbf{H}}, s \in \mathfrak{S}\right\}$ containing S i.i.d realizations drawn from a conditional distribution with density $f_{\mathbf{H}|\widehat{\mathbf{H}}}(\mathbf{H} \mid \widehat{\mathbf{H}})$. These realizations are available at the transmitter and can be used to approximate the Average rates experienced by each user through sample average functions (SAFs). When $S \to \infty$, according to the strong law of large numbers, we have

$$\overline{R}_{c,k} = \lim_{S\to\infty} \overline{R}_{c,k}^{(S)} = \lim_{S\to\infty} \frac{1}{S} \sum_{s=1}^{S} R_{c,k}(\mathbf{H}^{(s)}), \tag{3.54}$$

$$\overline{R}_{k} = \lim_{S\to\infty} \overline{R}_{k}^{(S)} = \lim_{S\to\infty} \frac{1}{S} \sum_{s=1}^{S} R_{k}(\mathbf{H}^{(s)}), \tag{3.55}$$

where $R_{c,k}(\mathbf{H}^{(s)})$ and $R_k(\mathbf{H}^{(s)})$, $s \in \mathfrak{S}$ are the common and private rates associated with the s-th channel realization. Accordingly, the SAA problem can be written as

$$\mathcal{F}_{RS}^{(S)}: \quad \max_{\bar{\mathbf{c}},\mathbf{P}} \min_{m\in\mathcal{M}} \left(\overline{C}_m + \min_{i\in\mathcal{G}_m} \overline{R}_i^{(S)}\right) \tag{3.56}$$

$$s.t. \quad \overline{R}_{c,k}^{(S)} \geq \sum_{m=1}^{M} \overline{C}_m, \quad \forall k \in \mathcal{K} \tag{3.57}$$

$$\overline{C}_m \geq 0, \quad \forall m \in \mathcal{M} \tag{3.58}$$

$$\mathbf{p}_c^H \mathbf{D}_l \mathbf{p}_c + \sum_{m=1}^{M} \mathbf{p}_m^H \mathbf{D}_l \mathbf{p}_m \leq P_l, \quad l = 1 \cdots L \tag{3.59}$$

Note that $\mathcal{F}_{RS}^{(S)}$ is a non-convex optimization problem which is challenging to solve. Next, we turn to solve the SAA problem using the WMMSE approach.

3.5 The WMMSE approach

The WMMSE approach, initially proposed in [48], is effective in solving problems containing non-convex superimposed rate expressions, i.e., RSMA-based sum rate maximization problems [16]. In this section, we further modify this approach so as

to solve the formulated SAA problem to achieve max-min fairness in RSMA-based multigroup multicast with imperfect CSIT. To begin with, the relationship between rate and WMMSE is derived, enabling the formulation of the equivalent problem.

3.5.1 Rate-WMMSE Relationship

At the k-th user, we denote the estimate of the common stream s_c by $\widehat{s}_{c,k} = g_{c,k} y_k$, where $g_{c,k}$ is a scalar equalizer. After s_c is successfully decoded by all receivers and removed from the received signal y_k, the estimate of $s_{\mu(k)}$ is obtained at user-k such that $\widehat{s}_{\mu(k)} = g_k(y_k - \mathbf{h}_k^H \mathbf{p}_c s_c)$, where g_k is the corresponding equalizer.

The common and private MSEs are defined as $\varepsilon_{c,k} = \mathbb{E}\{|\widehat{s}_{c,k} - s_{c,k}|^2\}$ and $\varepsilon_k = \mathbb{E}\{|\widehat{s}_{\mu(k)} - s_{\mu(k)}|^2\}$. The expectations are taken over the distributions of the input signals and the noise. By substituting the signal expressions into the definitions, the MSEs can be expressed by

$$\varepsilon_{c,k} = |g_{c,k}|^2 T_{c,k} - 2\mathcal{R}\{g_{c,k} \mathbf{h}_k^H \mathbf{p}_c\} + 1, \tag{3.60}$$

$$\varepsilon_k = |g_k|^2 T_k - 2\mathcal{R}\{g_k \mathbf{h}_k^H \mathbf{p}_{\mu(k)}\} + 1, \tag{3.61}$$

where the k-th user's received power is given by $T_{c,k} = |\mathbf{h}_k^H \mathbf{p}_c|^2 + |\mathbf{h}_k^H \mathbf{p}_{\mu(k)}|^2 + \sum_{j=1, j\neq\mu(k)}^{M} |\mathbf{h}_k^H \mathbf{p}_j|^2 + \sigma_n^2$. The power of observation after SIC writes as $T_k = T_{c,k} - |\mathbf{h}_k^H \mathbf{p}_c|^2$. Furthermore, we define $I_{c,k}$ as the interference plus noise portion in $T_{c,k}$ which is equal to T_k, and define $I_k = T_k - |\mathbf{h}_k^H \mathbf{p}_{\mu(k)}|^2$ as the interference plus noise portion in T_k. To minimize the MSEs over equalizers, we let $\frac{\partial \varepsilon_{c,k}}{\partial g_{c,k}} = 0$ and $\frac{\partial \varepsilon_k}{\partial g_k} = 0$. This yields the optimum equalizers given by

$$g_{c,k}^{MMSE} = \mathbf{p}_c^H \mathbf{h}_k T_{c,k}^{-1} \quad \text{and} \quad g_k^{MMSE} = \mathbf{p}_{\mu(k)}^H \mathbf{h}_k T_k^{-1}. \tag{3.62}$$

By substituting (3.62) into (3.60) and (3.61), the MMSEs with optimum equalizers,

i.e., the well-known MMSE equalizers are given by

$$\varepsilon_{c,k}^{MMSE} = \min_{g_{c,k}} \varepsilon_{c,k} = T_{c,k}^{-1} I_{c,k}, \tag{3.63}$$

$$\varepsilon_{k}^{MMSE} = \min_{g_{k}} \varepsilon_{k} = T_{k}^{-1} I_{k}. \tag{3.64}$$

It is evident that the SINRs can be expressed in the form of MMSEs such that

$$\gamma_{c,k} = (1/\varepsilon_{c,k}^{MMSE}) - 1 \quad \text{and} \quad \gamma_{k} = (1/\varepsilon_{k}^{MMSE}) - 1. \tag{3.65}$$

The corresponding rate expressions write as

$$R_{c,k} = -\log_2(\varepsilon_{c,k}^{MMSE}) \quad \text{and} \quad R_{k} = -\log_2(\varepsilon_{k}^{MMSE}). \tag{3.66}$$

Next, we introduce the augmented WMSEs from which the Rate-WMMSE relationship is derived. The common and private augmented WMSEs of user-k are respectively defined as

$$\xi_{c,k} = u_{c,k}\varepsilon_{c,k} - \log_2(u_{c,k}) \quad \text{and} \quad \xi_{k} = u_{k}\varepsilon_{k} - \log_2(u_{k}), \tag{3.67}$$

where $u_{c,k}$ and u_k denote auxiliary positive weights. By substituting optimum equalizers to WMSEs, we obtain

$$\xi_{c,k}(g_{c,k}^{MMSE}) = \min_{g_{c,k}} \xi_{c,k} = u_{c,k}\varepsilon_{c,k}^{MMSE} - \log_2(u_{c,k}), \tag{3.68}$$

$$\xi_{k}(g_{k}^{MMSE}) = \min_{g_{k}} \xi_{c,k} = u_{k}\varepsilon_{k}^{MMSE} - \log_2(u_{k}). \tag{3.69}$$

Moreover, let $\frac{\partial \xi_{c,k}\left(g_{c,k}^{MMSE}\right)}{\partial u_{c,k}} = 0$ and $\frac{\partial \xi_{k}\left(g_{k}^{MMSE}\right)_{k}}{\partial u_{k}} = 0$ to minimize the WMSEs over both equalizers and weights. This yields the optimum MMSE weights

$$u_{c,k}^{MMSE} = (\varepsilon_{c,k}^{MMSE})^{-1} \quad \text{and} \quad u_{k}^{MMSE} = (\varepsilon_{k}^{MMSE})^{-1}. \tag{3.70}$$

We substitute (3.70) into (3.68), (3.69), hence leading to the Rate-WMMSEs relationship

$$\xi_{c,k}^{MMSE} = \min_{g_{c,k},u_{c,k}} \xi_{c,k} = 1 + \log_2(\varepsilon_{c,k}^{MMSE}) = 1 - R_{c,k}, \tag{3.71}$$

$$\xi_{c,k}^{MMSE} = \min_{g_{c,k},u_{c,k}} \xi_{c,k} = 1 + \log_2(\varepsilon_{k}^{MMSE}) = 1 - R_{k}. \tag{3.72}$$

With respect to imperfect CSIT, a deterministic SAF version of the Rate-WMMSE relationship is constructed such that

$$\overline{\xi}_{c,k}^{MMSE(S)} = \min_{\mathbf{g}_{c,k},\mathbf{u}_{c,k}} \overline{\xi}_{c,k}^{(S)} = 1 - \overline{R}_{c,k}^{(S)}, \tag{3.73}$$

$$\overline{\xi}_{k}^{MMSE(S)} = \min_{\mathbf{g}_{k},\mathbf{u}_{k}} \overline{\xi}_{k}^{(S)} = 1 - \overline{R}_{k}^{(S)}. \tag{3.74}$$

This relationship holds for the whole set of stationary points [16]. For a given channel estimate, $\overline{\xi}_{c,k}^{MMSE(S)}$ and $\overline{\xi}_{k}^{MMSE(S)}$ represent the Average WMMSEs. We have $\overline{\xi}_{c,k}^{MMSE(S)} = \frac{1}{S}\sum_{s=1}^{S} \xi_{c,k}^{MMSE(s)}$ and $\overline{\xi}_{k}^{MMSE(S)} = \frac{1}{S}\sum_{s=1}^{S} \xi_{k}^{MMSE(s)}$, where $\xi_{c,k}^{MMSE(s)}$ and $\xi_{k}^{MMSE(s)}$ are associated with the s-th realization in $\mathbb{H}^{(S)}$. The sets of optimum equalizers are defined as $\mathbf{g}_{c,k}^{MMSE} = \{g_{c,k}^{MMSE(s)} \mid s \in \mathfrak{S}\}$ and $\mathbf{g}_{k}^{MMSE} = \{g_{k}^{MMSE(s)} \mid s \in \mathfrak{S}\}$. Following the same manner, the sets of optimum weights are $\mathbf{u}_{c,k}^{MMSE} = \{u_{c,k}^{MMSE(s)} \mid s \in \mathfrak{S}\}$ and $\mathbf{u}_{k}^{MMSE} = \{u_{k}^{MMSE(s)} \mid s \in \mathfrak{S}\}$. Each optimum element in these sets is associated with the s-th realization in $\mathbb{H}^{(S)}$. From the perspective of each user, the composite optimum equalizers and composite optimum weights are respectively

$$\mathbf{G}^{MMSE} = \left\{\mathbf{g}_{c,k}^{MMSE}, \mathbf{g}_{k}^{MMSE} \mid k \in \mathcal{K}\right\}, \tag{3.75}$$

$$\mathbf{U}^{MMSE} = \left\{\mathbf{u}_{c,k}^{MMSE}, \mathbf{u}_{k}^{MMSE} \mid k \in \mathcal{K}\right\}. \tag{3.76}$$

Note that the WMSEs are convex in each of their corresponding variables (e.g., equalizers, weights or precoding matrix) when fixing the other two. This block-wise convexity, preserved under superimposed expressions, together with the Rate-WMMSE relationship is the key to WMMSE approach [26]. Now, we can transform

$\mathcal{F}_{RS}^{(S)}$ into an equivalent WWMSE problem.

$$\mathcal{W}_{RS}^{(S)}: \quad \max_{\overline{\mathbf{c}},\mathbf{P},\mathbf{G},\mathbf{U},\overline{r}_g,\overline{\mathbf{r}}} \overline{r}_g \tag{3.77}$$

$$s.t. \quad \overline{C}_m + \overline{r}_m \geq \overline{r}_g, \quad \forall m \in \mathcal{M} \tag{3.78}$$

$$1 - \overline{\xi}_i^{(S)} \geq \overline{r}_m, \ \forall i \in \mathcal{G}_m, \quad \forall m \in \mathcal{M} \tag{3.79}$$

$$1 - \overline{\xi}_{c,k}^{(S)} \geq \sum_{m=1}^{M} \overline{C}_m, \quad \forall k \in \mathcal{K} \tag{3.80}$$

$$\overline{C}_m \geq 0, \quad \forall m \in \mathcal{M} \tag{3.81}$$

$$\mathbf{p}_c^H \mathbf{D}_l \mathbf{p}_c + \sum_{m=1}^{M} \mathbf{p}_m^H \mathbf{D}_l \mathbf{p}_m \leq P_l, \quad l = 1 \cdots L \tag{3.82}$$

where $\overline{r}_g$ and $\overline{\mathbf{r}} = [\overline{r}_1, \cdots, \overline{r}_M]$ are auxiliary variables. Furthermore, if the solution $(\mathbf{P}^*, \mathbf{G}^*, \mathbf{U}^*, \overline{r}_g^*, \overline{\mathbf{r}}^*, \overline{\mathbf{c}}^*)$ satisfies the KKT optimality conditions of $\mathcal{W}_{RS}^{(S)}$, $(\mathbf{P}^*, \overline{\mathbf{c}}^*)$ will satisfy the KKT optimality conditions of $\mathcal{F}_{RS}^{(S)}(P)$. Considering the block-wise convexity property, we use an Alternating Optimization (AO) algorithm illustrated below to solve $\mathcal{W}_{RS}^{(S)}$.

3.5.2 Alternating Optimization Algorithm

Each iteration of the AO algorithm is composed of two steps.

1) Updating $\mathbf{G}$ *and* $\mathbf{U}$:

During the n-th iteration, all the equalizers and weights are updated according to a given beamforming matrix such that $\mathbf{G} = \mathbf{G}^{MMSE}\left(\mathbf{P}^{[n-1]}\right)$ and $\mathbf{U} = \mathbf{U}^{MMSE}\left(\mathbf{P}^{[n-1]}\right)$, where $\mathbf{P}^{[n-1]}$ is the given beamforming matrix obtained from the previous iteration. To facilitate the $\mathbf{P}$ updating problem in the next step, we introduce several expressions calculated by updated $\mathbf{G}$ and $\mathbf{U}$ [16] to express the Average

WMSEs.

$$t_{c,k}^{(s)} = u_{c,k}^{(s)}\left|g_{c,k}^{(s)}\right|^2 \quad \text{and} \quad t_k^{(s)} = u_k^{(s)}\left|g_{c,k}^{(s)}\right|^2 \tag{3.83}$$

$$\Psi_{c,k}^{(s)} = t_{c,k}^{(s)}\mathbf{h}_k^{(s)}\mathbf{h}_k^{(s)H} \quad \text{and} \quad \Psi_k^{(s)} = t_k^{(s)}\mathbf{h}_k^{(s)}\mathbf{h}_k^{(s)H} \tag{3.84}$$

$$\mathbf{f}_{c,k}^{(s)} = u_{c,k}^{(s)}\mathbf{h}_k^{(s)}g_{c,k}^{(s)H} \quad \text{and} \quad \mathbf{f}_k^{(s)} = u_k^{(s)}\mathbf{h}_k^{(s)}g_k^{(s)H} \tag{3.85}$$

$$v_{c,k}^{(s)} = \log_2\left(u_{c,k}^{(s)}\right) \quad \text{and} \quad v_k^{(s)} = \log_2\left(u_k^{(s)}\right). \tag{3.86}$$

Therefore, by taking the averages over S realizations, the corresponding SAFs are $\overline{t}_{c,k}^{(S)}$, $\overline{t}_k^{(S)}$, $\overline{\Psi}_{c,k}^{(S)}$, $\overline{\Psi}_k^{(S)}$, $\overline{\mathbf{f}}_{c,k}^{(S)}$, $\overline{\mathbf{f}}_k^{(S)}$, $\overline{v}_{c,k}^{(S)}$, $\overline{v}_k^{(S)}$, from which leads to the Average WMSEs coupled with updated $\mathbf{G}$ and $\mathbf{U}$.

$$\overline{\xi}_{c,k}^{(S)} = \mathbf{p}_c^H\overline{\Psi}_{c,k}^{(S)}\mathbf{p}_c + \sum_{m=1}^{M}\mathbf{p}_m^H\overline{\Psi}_{c,k}^{(S)}\mathbf{p}_m + \sigma_n^2\overline{t}_{c,k}^{(S)} - 2\mathcal{R}\left\{\overline{\mathbf{f}}_{c,k}^{(S)H}\mathbf{p}_c\right\} + \overline{u}_{c,k}^{(S)} - \overline{v}_{c,k}^{(S)}, \tag{3.87}$$

$$\overline{\xi}_k^{(S)} = \sum_{m=1}^{M}\mathbf{p}_m^H\overline{\Psi}_k^{(S)}\mathbf{p}_m + \sigma_n^2\overline{t}_k^{(S)} - 2\mathcal{R}\left\{\overline{\mathbf{f}}_k^{(S)H}\mathbf{p}_{\mu(k)}\right\} + \overline{u}_k^{(S)} - \overline{v}_k^{(S)}. \tag{3.88}$$

2) Updating $\mathbf{P}$:

In this step, we fix $\mathbf{G}$, $\mathbf{U}$, and update $\mathbf{P}$ together with all the auxiliary variables. By substituting the Average WMSEs coupled with updated $\mathbf{G}$ and $\mathbf{U}$ into $\overline{\mathcal{W}}$, the problem of updating $\mathbf{P}$ based on updated $\mathbf{G}$ and $\mathbf{U}$ is formulated in $\overline{\mathcal{W}}^{(S)[n]}$. This is a convex optimization problem which can be solved using interior-point methods. The steps are summarized in Algorithm 1.

$$\mathcal{W}_{RS}^{(S)[n]} \qquad \max_{\overline{\mathbf{c}},\mathbf{P},\overline{r}_g,\overline{\mathbf{r}}} \overline{r}_g \tag{3.89}$$

$$s.t. \qquad \overline{C}_m + \overline{r}_m \geq \overline{r}_g, \quad \forall m \in \mathcal{M} \tag{3.90}$$

$$1 - \overline{r}_m \geq \sum_{m=1}^{M}\mathbf{p}_m^H\overline{\Psi}_k^{(S)}\mathbf{p}_m + \sigma_n^2\overline{t}_k^{(S)} - 2\mathcal{R}\left\{\overline{\mathbf{f}}_k^{(S)H}\mathbf{p}_{\mu(k)}\right\} + \overline{u}_k^{(S)} - \overline{v}_k^{(S)}, \ \forall i \in \mathcal{G}_m, \ \forall m \in \mathcal{M} \tag{3.91}$$

$$1-\sum_{m=1}^{M}\overline{C}_m \geq \mathbf{p}_c^H\overline{\Psi}_{c,k}^{(S)}\mathbf{p}_c+\sum_{m=1}^{M}\mathbf{p}_m^H\overline{\Psi}_{c,k}^{(S)}\mathbf{p}_m$$
$$+\sigma_n^2\overline{t}_{c,k}^{(S)}-2\mathcal{R}\{\overline{\mathbf{f}}_{c,k}^{(S)H}\mathbf{p}_c\}+\overline{u}_{c,k}^{(S)}-\overline{v}_{c,k}^{(S)},\ \forall k\in\mathcal{K} \tag{3.92}$$

$$\overline{C}_m\geq 0,\quad \forall m\in\mathcal{M} \tag{3.93}$$

$$\mathbf{p}_c^H\mathbf{D}_l\mathbf{p}_c+\sum_{m=1}^{M}\mathbf{p}_m^H\mathbf{D}_l\mathbf{p}_m\leq P_l,\quad l=1\cdots L \tag{3.94}$$

Through the AO algorithm, variables in the equivalent WMMSE problem are optimized iteratively in an alternating manner. The proposed algorithm is guaranteed to converge as the objective function is bounded above for the given power constraints. The objective function $\overline{r}_g$ increases until convergence as the iteration process goes on.

Algorithm 1 Alternating Optimization

Initialize: $n\leftarrow 0,\ \mathbf{P},\ \mathcal{W}_{RS}^{(S)[n]}\leftarrow 0$
while $\left|\mathcal{W}_{RS}^{(S)[n]}-\mathcal{W}_{RS}^{(S)[n-1]}\right|<\varepsilon$ **do**
  $n\leftarrow n+1,\ \mathbf{P}^{[n-1]}\leftarrow\mathbf{P}$
  $\mathbf{G}\leftarrow\mathbf{G}^{MMSE}\left(\mathbf{P}^{[n-1]}\right)$ (3.75)
  $\mathbf{U}\leftarrow\mathbf{U}^{MMSE}\left(\mathbf{P}^{[n-1]}\right)$ (3.76)
  update $\overline{t}_{c,k}^{(S)},\ \overline{t}_k^{(S)},\ \overline{\Psi}_{c,k}^{(S)},\ \overline{\Psi}_k^{(S)},\ \overline{\mathbf{f}}_{c,k}^{(S)},\ \overline{\mathbf{f}}_k^{(S)},\ \overline{v}_{c,k}^{(S)},\ \overline{v}_k^{(S)},\ \overline{u}_{c,k}^{(S)},\ \overline{u}_{c,k}^{(S)}$ (3.83)-(3.86)
  $\mathbf{P}\leftarrow$ Solve $\arg\mathcal{W}_{RS}^{(S)[n]}$
end while

3.6 Simulation Results and Analysis

In this section, the performance of the proposed RSMA-assisted multigroup multicast beamforming strategy is evaluated through simulation results by considering the scenarios of both Rayleigh fading channels (representative of cellular terrestrial systems) and multibeam satellite systems. Additionally, we evaluate the throughput performance by link-level simulations.

3.6.1 Performance Over Rayleigh Channels

The performance of RSMA and SDMA are both evaluated over Rayleigh fading channels (representative of conventional cellular terrestrial systems) when considering a total transmit power constraint. During simulation, entries of $\mathbf{H}$ are independently drawn from $\mathcal{CN}(0,1)$. Following the CSIT uncertainty model, entries of $\widetilde{\mathbf{H}}$ are also i.i.d complex Gaussian drawn from $\mathcal{CN}(0,\sigma_e^2)$, where $\sigma_e^2 = N_t^{-1}\sigma_{e,k}^2 = P^{-\alpha}$. Herein, we evaluate the MMF Ergodic rate by averaging over 100 channel estimates. For each given channel estimate $\widehat{\mathbf{H}} = \mathbf{H} - \widetilde{\mathbf{H}}$, its corresponding MMF Average rate is approximated by SAA method and the sample size S is set to be 1000. $\mathbb{H}^{(S)}$ is the set of conditional realizations available at the transmitter. The s-th conditional realization in $\mathbb{H}^{(S)}$ is given by $\mathbf{H}^{(s)} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}^{(s)}$, where $\widetilde{\mathbf{H}}^{(s)}$ follows the above CSIT error distribution.

We firstly consider an underloaded system with $N_t = 6$ transmit antennas, $G = 3$ groups and $K = 6$ users. The group sizes are respectively $G_1 = 1,\ G_2 = 2,\ G_3 = 3$. Fig. 3.1 presents the MMF Ergodic rate of RSMA and SDMA versus an increasing SNR under various CSIT qualities. For perfect CSIT, beaming an interference-free stream to each group simultaneously is possible since the system is underloaded. Both RSMA and SDMA achieve full MMF-DoF and the performance of such two schemes are nearly identical. However, RSMA shows a little improvement in the rate sense compared with SDMA due to its more flexible architecture. For imperfect CSIT, the superiority of RSMA over SDMA becomes more evident. It can be observed in Fig. 3.1 that the MMF-DoF disparity between RSMA and SDMA gradually appears as the CSIT uncertainty increases. The MMF-DoFs of SDMA and RSMA in Fig. 3.1 are respectively α and $\frac{1-\alpha}{M} + \alpha$, which follow the results in Table 3.1. This implies that the common stream of RSMA can provide a DoF gain of $\frac{1-\alpha}{M}$ and consequently MMF rate gains in underloaded regimes.

In Fig. 3.2, we reduce the number of transmit antennas to 4 and the system becomes

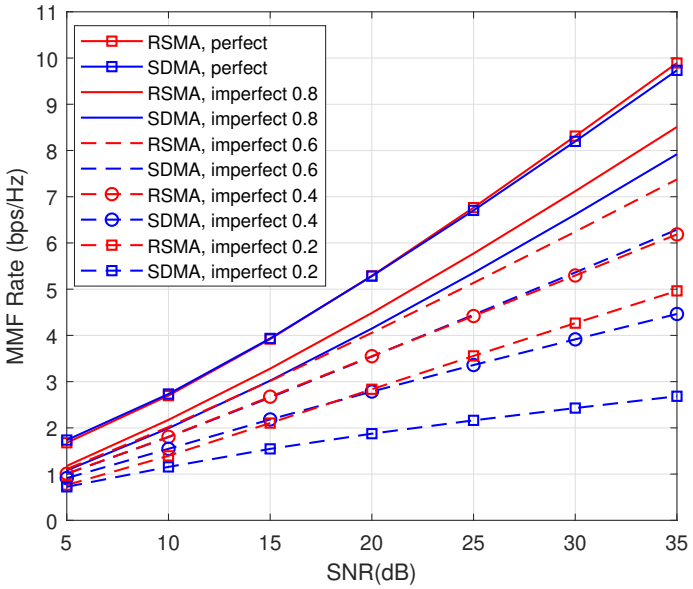

Figure 3.1: MMF rate performance. $N_t = 6$ antennas, $K = 6$ users, $M = 3$ groups, $G_1, G_2, G_3 = 1, 2, 3$ users.

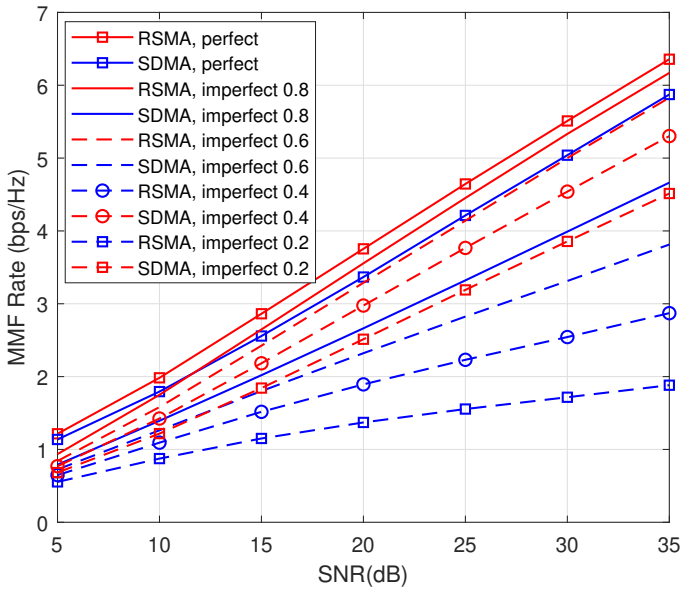

Figure 3.2: MMF rate performance. $N_t = 4$ antennas, $K = 6$ users, $M = 3$ groups, $G_1, G_2, G_3 = 1, 2, 3$ users.

partially-overloaded ($K - G_3 + 1 \leq N_t < K - G_1 + 1$). When considering perfect CSIT, RSMA and SDMA achieve identical MMF-DoFs at $\frac{1}{2}$. It follows the perfect CSIT results in Table 3.1. Meanwhile, it also follows the results of imperfect CSIT by setting $\alpha = 1$. Multiplexing gains are partially achieved. A small rate gap between the two schemes is observed although their MMF-DoFs are equal. Next, it comes to imperfect CSIT. We can see that the merit of RSMA over SDMA becomes more obvious compared with the underloaded regime. From Fig. 3.2, the MMF-DoFs of SDMA (blue curves) are approximately $\frac{\alpha}{2}$, which match the theoretical result in (3.11). CSIT imperfectness can affect the system's performance significantly. Considering RSMA, we have $M_{\mathrm{R}}^* = 2$ as a result of $N_2 \leq N_t < N_3$ in this specific setup. Substituting $M_{\mathrm{R}}^* = 2$ and $M = 3$ into (3.33) or the overloaded results in Table 3.1, we obtain

$$d^{*RS} \geq \begin{cases} \frac{1}{2}, & 0.5 < \alpha \leq 1 \\ \alpha + \frac{1-2\alpha}{3}, & 0 \leq \alpha \leq 0.5. \end{cases} \tag{3.95}$$

In addition, we have $d^{*SDMA} = \frac{\alpha}{2}$. Such DoF performance is exhibited in Fig. 3.2. All simulation results are inline with the theoretical MMF-DoFs in Table 3.1. Due to the benefits of RSMA, the system is able to maintain its MMF-DoFs at $\frac{1}{2}$ for all $0.5 < \alpha \leq 1$ in this example. When $0 \leq \alpha \leq 0.5$, the MMF-DoFs decrease slightly to $\alpha + \frac{1-2\alpha}{3}$, which is still greater than the $\frac{\alpha}{2}$ achieved by SDMA. Compared with the underloaded scenario in Fig. 3.1, the gaps between RSMA (red curves) and SDMA (blue curves) increase. In other words, the superiority of RSMA over SDMA becomes more apparent when the system is partially-overloaded.

Furthermore, we keep the same setting as in Fig. 3.2 but change the group sizes to be symmetric, i.e., $G_1 = 2,\ G_2 = 2,\ G_3 = 2$. It is noted that the system at present becomes fully-overloaded ($1 \leq N_t < K - G_3 + 1$). As illustrated in Fig. 3.3, RSMA outperforms SDMA to a great extent in both perfect CSIT and imperfect CSIT

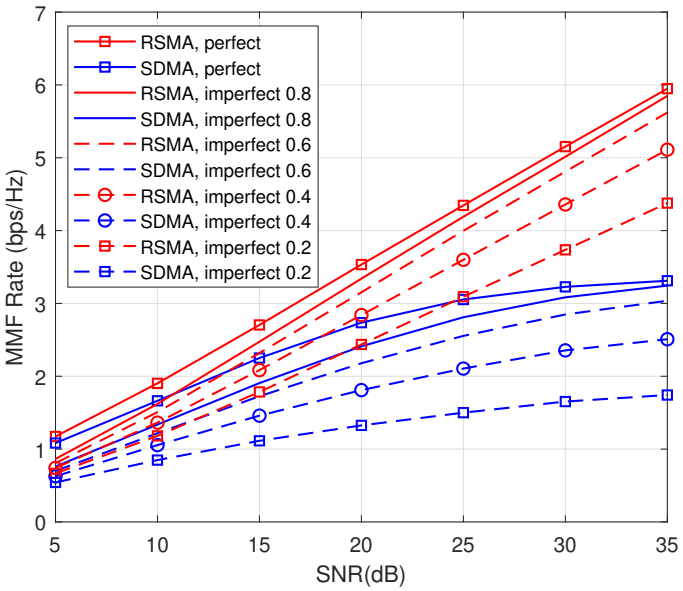

Figure 3.3: MMF rate performance. $N_t = 4$ antennas, $K = 6$ users, $M = 3$ groups, $G_1, G_2, G_3 = 2, 2, 2$ users.

scenarios. RSMA maintains the same MMF-DoFs as in Fig. 3.2. However, all the multiplexing gains of SDMA are sacrificed and collapse to 0. The corresponding MMF rate performance of SDMA gradually saturates as SNR grows, thus resulting in severe rate limitation.

Through the simulation results over Rayleigh fading channels, it is demonstrated that RSMA-based multigroup multicast beamforming is more robust to CSIT imperfectness than the conventional SDMA scheme. RSMA is able to further exploit spatial multiplexing gains and achieve higher MMF rate performance in various setups. In particular, RSMA provides significant gains compared with SDMA in overloaded regimes with imperfect CSIT.

Above all, the gains of RSMA for multigroup multicast in the presence of imperfect CSIT are shown via simulations in both underloaded and overloaded deployments. This contrasts with [26] where gains in the presence of perfect CSIT were demonstrated primarily in the overloaded scenarios.

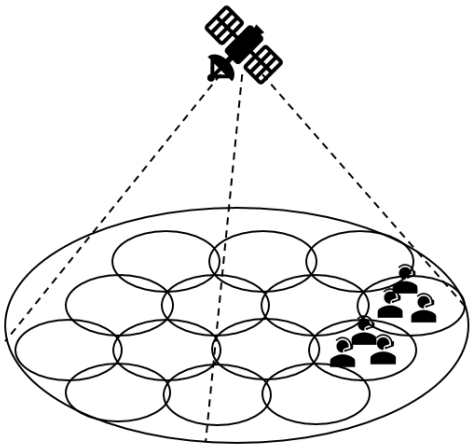

Figure 3.4: Architecture of multibeam satellite systems.

3.6.2 Application to Multibeam Satellite Systems

In order to show the versatility of RSMA, the application of RSMA-based multigroup multicast beamforming to multibeam satellite systems is addressed in this section. Here, we focus on a Ka-band multibeam satellite system with multiple single-antenna terrestrial users served by a geostationary orbit (GEO) satellite as shown in Fig. 3.4. A single gateway is employed in this system, and the feeder link between the gateway and the satellite is assumed to be noiseless. Let N_t denote the number of antenna feeds. The array fed reflector can transform N_t feed signals into M transmitted signals (i.e., one signal per beam) to be radiated over the multibeam coverage area [89]. Considering single feed per beam architecture which is popular in modern satellites such as Eutelsat Ka-Sat [34, 36], only one feed is required to generate one beam (i.e., $N_t = M$). Since the multibeam satellite system is in practice user overloaded, we assume that ρ ($\rho > 1$) users are served simultaneously by each beam. Users per beam are uniformly distributed within the satellite coverage area. Ideally, the user selection and beamforming can be jointly designed. However, this is out of the scope of this thesis and can be explored in future work. $K = \rho N_t$ is the total number of users.

Table 3.2: Simulation parameters [Chapter 3]

Parameter	Value
Frequency band (carrier frequency)	Ka (20 GHz)
Satellite height	35786 km (GEO)
User link bandwidth	500 MHz
3 dB angle	0.4°
Maximum beam gain	52 dBi
User terminal antenna gain	41.7 dBi
System noise temperature	517 K

1) Multibeam Satellite Channel:

The main difference between satellite and terrestrial communications lies in the channel characteristics including free space loss, radiation pattern and atmospheric fading. The satellite channel $\mathbf{H} \in \mathbb{C}^{N_t \times K}$ is a matrix composed of receive antenna gain, free space loss and satellite multibeam antenna gain. Its (n, k)-th entry can be modeled as

$$H_{n,k} = \frac{\sqrt{G_R G_{n,k}}}{4\pi \frac{d_k}{\lambda} \sqrt{\kappa T_{sys} B_w}} \tag{3.96}$$

where G_R is the user terminal antenna gain, d_k is the distance between user-k and the satellite, λ is the carrier wavelength, κ is the Boltzmann constant, T_{sys} is the receiving system noise temperature and B_w denotes the user link bandwidth. $G_{n,k}$ is the multibeam antenna gain from the n-th feed to the k-th user. It mainly depends on the satellite antenna radiation pattern and user locations.

2) Performance Over Satellite Channels:

Then, we evaluate the application of RSMA in multibeam satellite communications. Results of MMF problems are obtained by averaging 100 satellite channel realizations. Since non-flexible on-board payloads prevent power sharing between beams, per-feed power constraints are adopted. System parameters are summarized in Table 3.2. Fig. 3.5 shows the curves of MMF rates among $N_t = 7$ beams versus an increasing per-feed available transmit power. We assume two users per beam, i.e., $\rho = 2$.

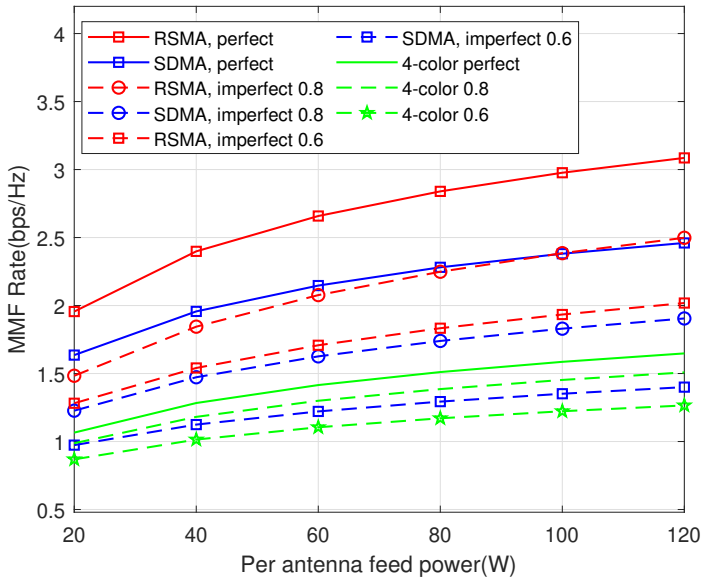

Figure 3.5: MMF rate versus per-feed available power. $N_t = 7$ antennas, $K = 14$ users, $\rho = 2$ users.

For perfect CSIT, RSMA achieves around 25% gains over SDMA. For imperfect CSIT, RSMA is seen to outperform SDMA with 31% and 44% gains respectively when $\alpha = 0.8$ and $\alpha = 0.6$. Accordingly, the advantage of employing RSMA in multigroup multicast beamforming is still observed in multibeam satellite systems. Through partially decoding the interference and partially treating the interference as noise, RSMA is more robust to the CSIT uncertainty and overloaded regime than SDMA. Such benefit of RSMA exactly tackles the challenges of multibeam satellite communications. The conventional 4-colour scheme performs the worst compared with full frequency reuse schemes.

Here, we set the per-feed available transmit power to be 80 Watts. As CSIT error scaling factor drops, the MMF rate gap between RSMA and SDMA increases gradually, which implies the gains of our proposed RSMA scheme become more and more apparent as the CSIT quality decreases. In addition, the impact of user number per frame is also studied. Since all the users within a beam share the same beamforming vector, the beam rate is determined by the user with the lowest SINR.

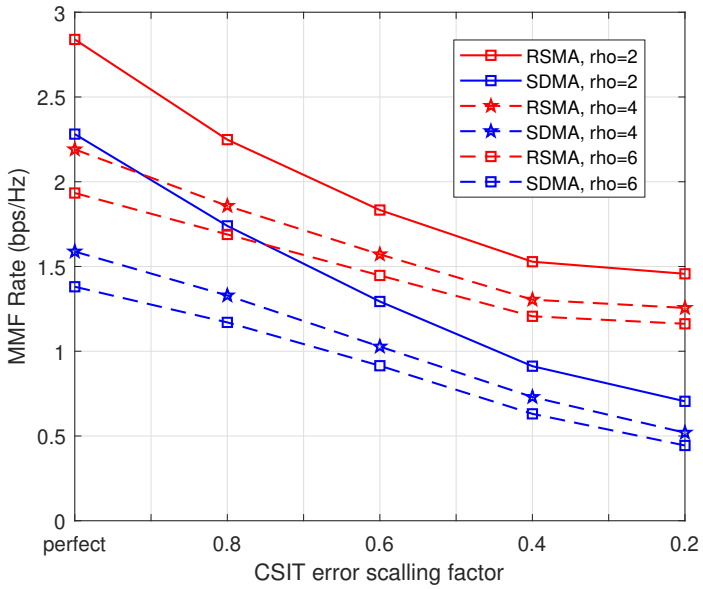

Figure 3.6: MMF rate versus CSIT error scalling factor α. $N_t = 7$ antennas, $\rho = 2,\ 4,\ 6$ users, $P/N_t = 80$ W.

Considering $\rho = 2,\ 4,\ 6$ users per frame, it is clear that increasing the number of users per frame results in system performance degradation for both RSMA and SDMA.

Moreover, the impact of different transmit power constraints is studied. Based on the fair per-antenna power constraint assumption, each transmit antenna cannot radiate a power more than P/N_t. Compared with the total transmit power constraint, the existence of per-antenna power constraint will inevitably restrict the flexibility of beamforming design. Taking imperfect CSIT with $\alpha = 0.8$ as an example, Fig. 3.7 respectively shows the MMF rates when considering total power constraint and per-antenna power constraint. It is noticed that the practical per-antenna power constraint reduces MMF rate performance slightly in both RSMA and SDMA.

Finally, we consider a hot spot user configuration rather than the uniform user configuration. In Fig. 3.8, the performance of a hot spot configuration, (e.g., with 8 users in the central beam and 1 user each in the other beams) is compared with the above uniform setting. We can observe that the MMF rate improvement provided by

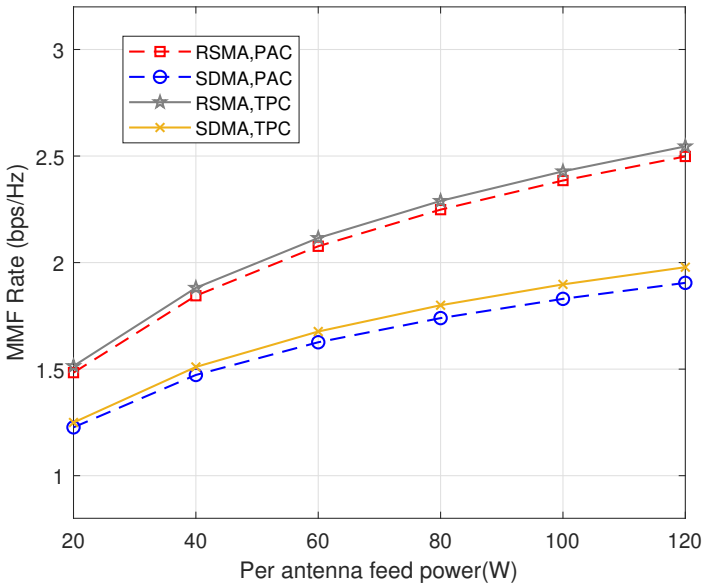

Figure 3.7: MMF rate constrained by PAC/ TPC. $N_t = 7$ antennas, $K = 14$ users, $\rho = 2$ users, imperfect CSIT: $\alpha = 0.8$.

RSMA is more obvious than SDMA, which means that RSMA is better at managing interference in such a hot spot scenario. Specifically, for perfect CSIT, RSMA outperforms SDMA with 42% gains. For imperfect CSIT, RSMA achieves higher gains at around 54%.

3.6.3 Link-Level Simulations

In this section, by leveraging the results of the MMF optimization problem with assumptions of Gaussian inputs and infinite block length, we further investigate the RSMA PHY layer design for multigroup multicast with finite length polar coding, finite alphabet modulation and an AMC algorithm. In [90], the uncoded link-level performance of RSMA-based multiuser MISO systems is investigated. With channel coding taken into consideration, [32] designs the basic transmitter and receiver architecture for RMSA in a MISO BC with two single-antenna users. Here, we use the same transceiver architecture as [32] and conduct LLS to show explicit throughput gain of RSMA multigroup multicast in both cellular and multibeam

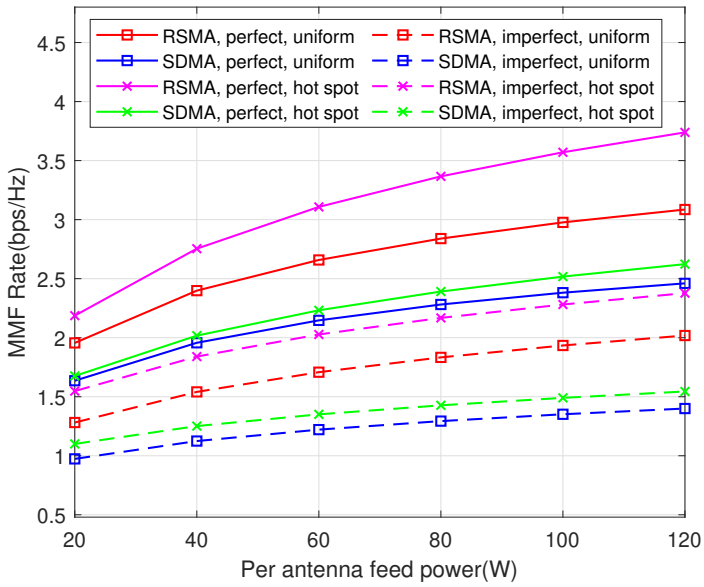

Figure 3.8: MMF rate versus per-feed available power. $N_t = 7$ antennas, $K = 14$ users, imperfect CSIT: $\alpha = 0.6$, hot spot $G = [8, 1, 1, 1, 1, 1, 1]$.

satellite systems.

The transmitter and receiver architecture for RSMA multigroup multicast is depicted in Fig. 3.9. We use finite alphabet modulation symbols carrying codewords from finite-length polar code codebooks as channel inputs. The overall framework follows the architecture in [32] where a two-user MISO BC system is considered. For more detailed explanations of each module, please refer to Appendix A.

Thus, we can demonstrate the performance improvements achieved by RSMA over SDMA for multigroup multicast by LLS results and compare the obtained throughput levels with the Shannon bounds obtained in the previous sections. The PHY-layer design follows the architecture described in Fig. 3.9. Appropriate modulation schemes and coding rates are selected by the AMC algorithm.

In LLS, we define throughput as the number of bits which can be transmitted correctly at a single channel use. All MMF throughput levels are obtained by averaging over 100 Monte-Carlo realizations. The number of channel uses in the l-th

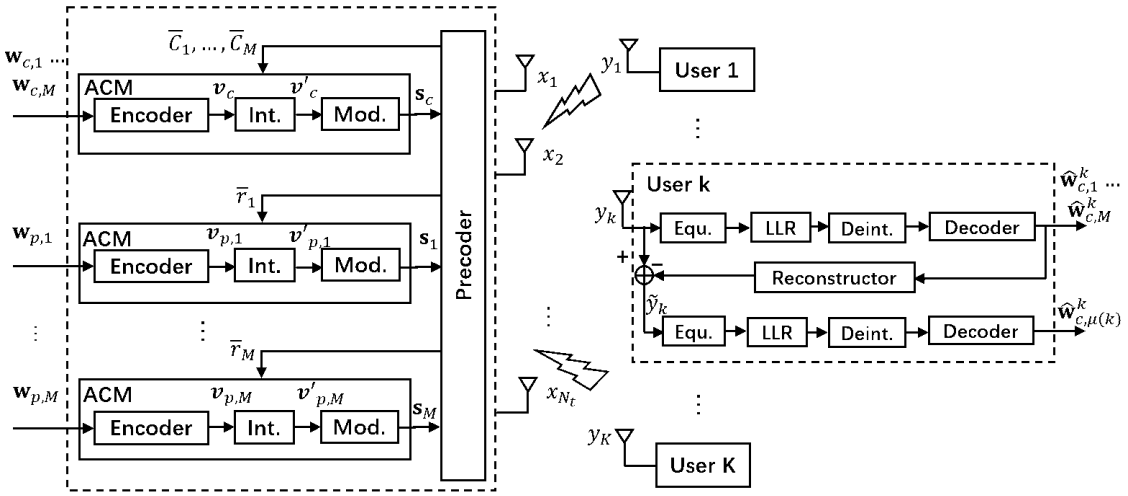

Figure 3.9: Transceiver architecture of RSMA multigroup multicast.

Monte-Carol realization is denoted by $S^{(l)}$. $D_{s,k}^{(l)}$ denotes the number of successfully recovered information bits by user-k for all $k \in \mathcal{K}$. Thus, the MMF throughput can be written as

$$\text{MMF Throughput [bps/Hz]} = \frac{\min_{k\in\mathcal{K}} \sum_l D_{s,k}^{(l)}}{\sum_l S^{(l)}}. \tag{3.97}$$

Without loss of generality, we assume $S^{(l)} = 256$ for all $l = 1, \cdots, 100$ Monte-Carlo realizations. The maximum code rate is set as $\beta = 0.9$.

First, we consider a cellular terrestrial multigroup multicast system with $K = 6$ users equally divided into $M = 3$ multicast groups. Independent and identically distributed Rayleigh fading channels are adopted. When the number of transmit antenna $N_t = 6$, the system is underloaded. Fig. 3.10 and Fig. 3.11 respectively show the Shannon bounds and throughput levels achieved by RSMA and SDMA with imperfect CSIT $\alpha = 0.8$ and $\alpha = 0.6$. It can be clearly observed that RSMA has a significant LLS throughput gain over SDMA. The trend of throughput levels is consistent with that of Shannon bounds. The performance improvements achieved by RSMA compared with SDMA are demonstrated in the PHY-layer design and LLS platform. Moreover, as the CSIT error scaling factor drops from 0.8 to 0.6, the CSIT uncertainty increases, thus leading to lower throughput values.

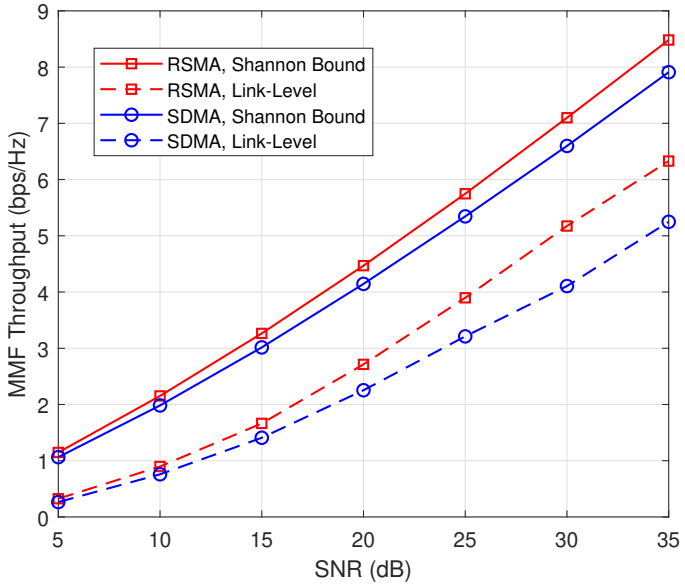

Figure 3.10: MMF throughput versus SNR, $\alpha = 0.8$, $N_t = 6$ antennas, $K = 6$ users, 2 users per group.

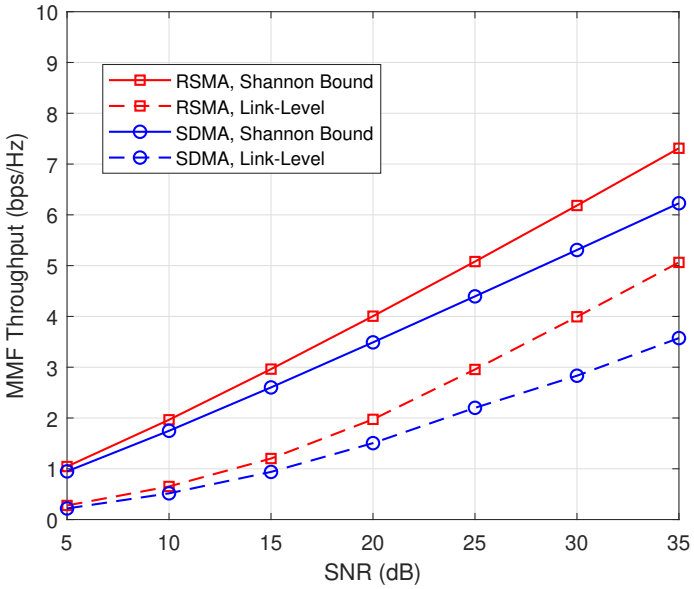

Figure 3.11: MMF throughput versus SNR, $\alpha = 0.6$, $N_t = 6$ antennas, $K = 6$ users, 2 users per group.

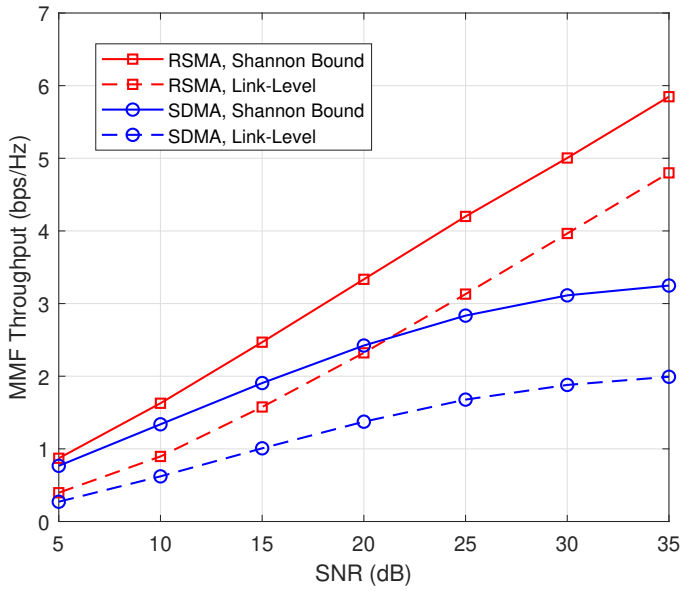

Figure 3.12: MMF throughput versus SNR, $\alpha = 0.8$, $N_t = 4$ antennas, $K = 6$ users, 2 users per group.

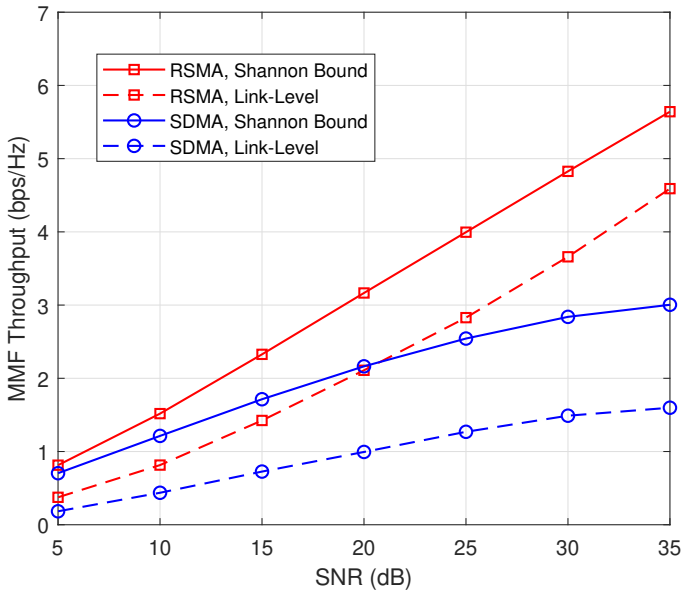

Figure 3.13: MMF throughput versus SNR, $\alpha = 0.6$, $N_t = 4$ antennas, $K = 6$ users, 2 users per group.

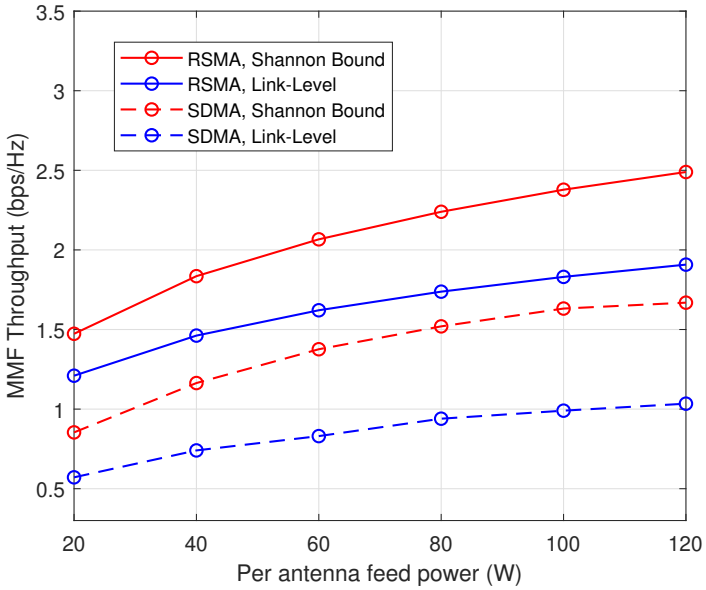

Figure 3.14: MMF throughput versus per-feed available power, $\alpha = 0.8$, $N_t = 7$ antennas, $K = 14$ users, 2 users per group.

Next, Fig. 3.12 and Fig. 3.13 depict the Shannon bounds and throughput levels when the number of transmit antenna N_t is 4. Now the system becomes overloaded, and all multiplexing gains of SDMA are sacrificed and collapse to 0 [68]. The curve of SDMA Shannon bound gradually saturates as SNR grows. The rate gain of RSMA over SDMA is more obvious. By LLS, the MMF throughput levels of both RSMA and SDMA follow the trend of Shannon bounds with comparable gaps. The throughput of RSMA outperforms SDMA significantly in the presence of considered imperfect CSIT $\alpha = 0.8$ and $\alpha = 0.6$.

Finally, we consider the same multibeam satellite system as discussed in Section 3.6.2, where a GEO satellite equipped with $N_t = 7$ antennas serves $K = 14$ single-antenna users simultaneously. Single feed per beam architecture is used such that only one feed is required to generate one beam (i.e., $N_t = M$). $\rho = \frac{K}{M} = 2$ users are served simultaneously by each beam. Fig. 3.14 illustrates the Shannon bounds and throughput levels achieved by RSMA and SDMA versus an increasing per-antenna transmit power budget with imperfect CSIT $\alpha = 0.8$ We can still observe the

matching trends of the Shannon bounds and throughput curves in this satellite setup. The effectiveness of using RSMA in multibeam satellite systems compared with conventional SDMA is demonstrated by LLS.

3.7 Summary

In this chapter, we focus on the application of RSMA for multigroup multicast beamforming in the presence of imperfect CSIT. Through MMF-DoF analysis, RSMA is shown to provide gains in both underloaded and overloaded systems compared with the conventional SDMA. A generic MMF optimization problem is formulated and solved by developing a modified WMMSE approach together with an AO algorithm. The effectiveness of adopting RSMA for multigroup multicast and multibeam satellite communications is evaluated through simulations in a wide range of setups, taking into account CSIT uncertainty and practical challenges. Additionally, the RSMA transmitter and receiver architecture and LLS platform are designed. According to numerical link-level results, we can conclude that RSMA is very promising for practical implementation to tackle the challenges of modern communication systems in numerous application areas.

Chapter 4

RSMA for Satellite-Terrestrial Integrated Networks

In this chapter, we investigate the joint beamforming design problem to achieve max-min rate fairness in a STIN where the satellite provides wide coverage to multibeam multicast satellite users (SUs), and the terrestrial BS serves multiple cellular users (CUs) in a densely populated area. Both the satellite and BS operate in the same frequency band. We present two RSMA-based STIN schemes, namely the coordinated scheme relying on CSI sharing and the cooperative scheme relying on CSI and data sharing. The objective is to maximize the minimum fairness rate amongst all SUs and CUs subject to transmit power constraints at the satellite and the BS. A joint beamforming algorithm is proposed to reformulate the original problem into an approximately equivalent convex one, which can be iteratively solved. Moreover, an expectation-based robust joint beamforming algorithm is proposed against the practical environment when the satellite channel phase uncertainties are considered. Simulation results demonstrate the effectiveness and robustness of the proposed RSMA schemes for STIN and exhibit significant performance gains compared with various baseline strategies.

4.1 Introduction

The concept of STIN has been proposed in the literature [91–93]. The satellite sub-network shares the same frequency band with the terrestrial sub-network through dynamic spectrum access technology to enhance spectrum utilization, thereby achieving higher spectrum efficiency and throughput. However, aggressive frequency reuse can induce severe interference within and between the sub-networks. In this chapter, we will concentrate on RSMA-based joint beamforming schemes to efficiently mitigate the interference of STIN.

Motivated by the benefits of RSMA presented in Chapter 3, in this chapter, we further investigate the application of RSMA into STIN to manage the interference within and between both sub-networks. Practical challenges are considered, such as the per-feed transmit power constraints, CSIT uncertainty, and multibeam multicast transmission due to the existing satellite communication standards [65]. The main contributions of this chapter are summarized as follows.

- First, we present a multiuser downlink framework for the integrated network where the satellite exploits multibeam multicast communication to serve SUs, while the terrestrial BS employs uniform planar array (UPA) and serves cellular users (CUs) in a densely populated area. We take into account multibeam satellite characteristics, including the array pattern, path loss and rain attenuation, thus building a more realistic channel model to evaluate the system performance. The GW operates as a control center to implement centralized processing and control the whole network. Based on such framework, the joint beamforming design arises so that the satellite and terrestrial sub-system can share the same radio spectrum resources and cooperate with each other. RSMA is used at both the satellite and the BS to mitigate the interference including inter-beam interference, intra-cell interference and interference between the two

sub-systems. We investigate two scenarios of RSMA-based STIN, namely the *coordinated scheme*, and the *cooperative scheme*. For the *coordinated scheme*, the satellite and BS exchange CSI of both direct and interfering links at the GW, and coordinate beamforming to manage the interference. For the *cooperative scheme*, the satellite and BS exchange both CSI and data at the GW. All propagation links (including interfering ones) are exploited to carry useful data upon appropriate beamforming. This differs from the prior RSMA-based STIN paper [71], where RSMA is utilized only at the terrestrial sub-system, and the benefits of coordination and cooperation are not investigated.

- Second, for both *coordinated scheme* and *cooperative scheme*, we respectively formulate optimization problems to maximize the minimum fairness rate of the RSMA-based STIN amongst all users subject to the constraint of per-feed transmit power at the satellite and the constraint of sum transmit power at the BS. Such problems upgrade the application of RSMA to multibeam satellite communications and terrestrial networks to a more general case, therefore leading to a joint beamforming design so that the two sub-systems can cooperate with each other. This is the first work on the joint beamforming design of RSMA-based *coordinated* STIN and *cooperative* STIN. Since the original optimization problem is non-convex, we apply the SCA to reformulate the original problem into an approximately equivalent convex one, which belongs to a second-order cone program (SOCP) and can be solved iteratively. The *cooperative scheme* is shown to outperform the *coordinated scheme* due to data exchange between the satellite and BS at the GW. Multiple baseline strategies are considered, including SDMA, NOMA, a two-step beamforming and fractional frequency reuse. Simulation results demonstrate the superiority of the proposed RSMA-based *cooperative scheme* and *coordinated scheme* compared with the baseline strategies.

- Third, since it is in general very challenging to acquire accurate satellite CSI at the GW due to the round-trip delay and device mobility, we develop an expectation-based robust beamforming design against satellite channel phase uncertainty. For both RSMA-based *coordinated* STIN and *cooperative* STIN, non-convex MMF problems are formulated. To tackle the non-convexity of the robust design, a novel iterative algorithm is proposed using SCA combining with a penalty function. Simulation results verify the effectiveness and robustness of the proposed RSMA schemes for STIN.

4.2 System Model

As illustrated in Fig. 4.1, we consider a STIN system employing full frequency reuse, where all SUs and CUs operate in the same frequency band. A GEO satellite is equipped with an array-fed reflector antenna. It provides services to SUs that lack terrestrial access in sparsely populated or remote areas. By assuming a single feed per beam architecture, the array-fed reflector antenna comprises a feed array with N_s feeds and generates N_s adjacent beams. Within the multibeam coverage area, we assume K_s SUs, and $\rho = \frac{K_s}{N_s}$ users in each beam. Since the SUs of each beam are served simultaneously by transmitting a single coded frame following DVB-S2X, the GEO satellite implements multibeam multicast transmission. Meanwhile, the terrestrial BS[1] equipped with N_t-antenna UPA serves densely populated areas in the same frequency band. $K_t \leq N_t$ unicast CUs are assumed. User mobility is not considered in this work. Spectrum sharing is able to improve spectrum efficiency, which also leads to interference in and between the terrestrial and satellite sub-networks. As shown in Fig. 4.1, the GW acts as a control center to collect and manage various kinds of information, implement centralized processing and

[1]In this chapter, a unique BS is considered. The setting of multiple BSs is not considered here and is left for future studies.

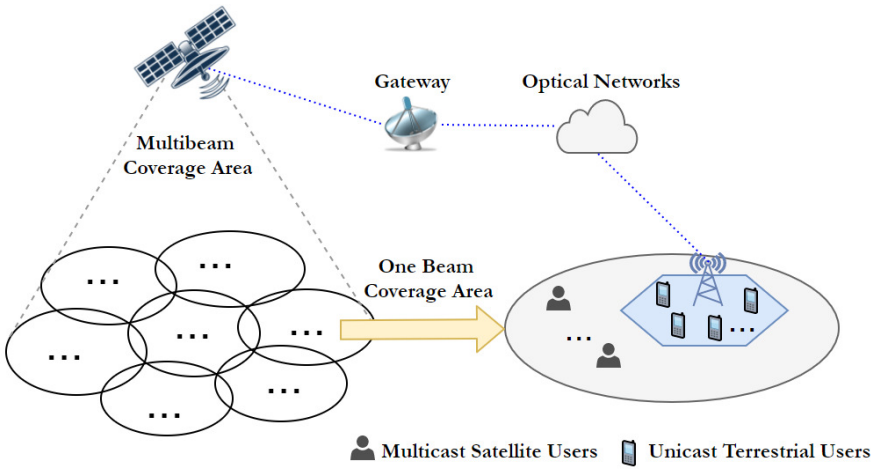

Figure 4.1: Model of a satellite-terrestrial integrated network.

control the whole STIN. Optimal resource allocation and interference management on the satellite and BS can be jointly implemented at the GW[2] to improve system performance.

4.2.1 Channel Model

As illustrated in Fig. 4.2, we assume UPA at the BS with dimension $N_t = N_1 \times N_2$. N_1 and N_2 are respectively the number of array elements uniformly employed along the X-axis and the Z-axis. The distances between adjacent array elements are identical, thus $d_1 = d_2 = d$. Due to the characteristic of radio wave propagation at high-frequency bands, the terrestrial channels can be expressed by a model consisting of L scatters. Each scatter contributes to a single propagation path. Mathematically,the downlink channel between the BS and CU k_t is given by [96]

$$\mathbf{h}_{k_t} = \sqrt{\frac{1}{L}} \sum_{l=1}^{L} \alpha_{k_t,l} \mathbf{a}_{\text{UPA}} \left(\theta_{k_t,l}, \varphi_{k_t,l}\right), \tag{4.1}$$

[2]Complete CSI of the STIN system is required at the GW, leading to significant CSI feedback overhead. To reduce the feedback overhead in STIN systems, several techniques can be used including e.g., compressed sensing, codebook-based feedback, spatially correlated feedback and CSI prediction using Kalman filtering or deep learning-based methods [94, 95]. However, this problem exceeds the scope of this thesis.

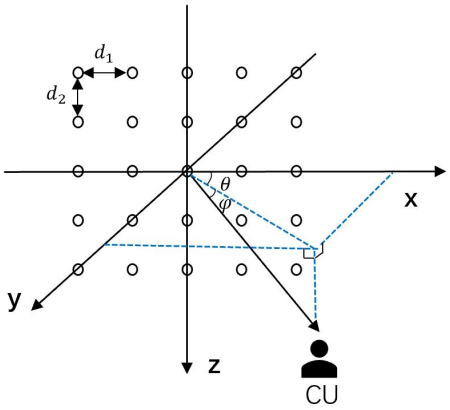

Figure 4.2: Geometry of uniform planar array employed at the BS.

where $\alpha_{k_t,l}$ is the complex channel gain of the l-th path. Each $\alpha_{k_t,l}$ is assumed to follow independent and identical distribution (i.i.d) $\mathcal{CN}(0,1)$. By denoting $\theta_{k_t,l}$ and $\varphi_{k_t,l}$ as the azimuth and elevation angles of the l-th path, the vector $\mathbf{a}_{\text{UPA}}(\theta_{k_t,l}, \varphi_{k_t,l})$ can be expressed as a function of the Cartesian coordinates of the transmit arrays as follows

$$\mathbf{a}_{\text{UPA}}(\theta_{k_t,l}, \varphi_{k_t,l}) = e^{\left(j\frac{2\pi}{\lambda}\left[\bar{\mathbf{r}}_1,\cdots,\bar{\mathbf{r}}_{N_t}\right]^T\left[\cos\theta_{k_t,l}\cos\varphi_{k_t,l},\ \ \sin\theta_{k_t,l}\cos\varphi_{k_t,l},\ \ \cos\varphi_{k_t,l}\right]^T\right)}. \tag{4.2}$$

where $[\bar{\mathbf{r}}_1, \cdots, \bar{\mathbf{r}}_{N_t}] \in \mathcal{R}^{3\times N_t}$ have columns representing the Cartesian coordinates of the UPA array elements. The terrestrial channel matrix between the BS and all CUs is denoted by $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_{K_t}] \in \mathbb{C}^{N_t\times K_t}$.

Considering the free space loss, radiation pattern and rain attenuation of satellite channels, the downlink channel from the satellite to SU-k_s can be modelled the same as in Section 3.6.2.

The satellite channel matrix between the satellite and all SUs is denoted by $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_{K_s}] \in \mathbb{C}^{N_s\times K_s}$. Similarly, when we consider $n_s \in \{1, \cdots, N_s\}$ and $k_t \in \{1, \cdots, K_t\}$, the interfering satellite channel matrix between the satellite and all CUs is denoted by $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_{K_t}] \in \mathbb{C}^{N_s\times K_t}$.

4.2.2 Coordinated scheme and Cooperative Scheme

We consider two levels of integration between the satellite and terrestrial BS.

1) Coordinated Scheme:

First, we consider the basic level of integration where the CSI of both direct and interfering links of the whole network is available at the GW, while data is not exchanged between the satellite and BS at the GW. We call such scheme a *coordinated scheme*. It allows the satellite and BS to coordinate power allocation and beamforming directions to suppress interference. Different multiple access strategies can be exploited at the satellite and BS, such as RSMA, SDMA, NOMA, etc. Here, we elaborate on the scenario where RSMA[3] is used at both the BS and satellite. To that end, the unicast messages $W_1, \cdots, W_{K_t}$ intended to CUs indexed by $\mathcal{K}_t = \{1, \cdots, K_t\}$ are split into common parts and private parts, i.e., $W_{k_t} \rightarrow \{W_{c,k_t}, W_{p,k_t}\}, \forall k_t \in \mathcal{K}_t$. All common parts are combined into W_c and encoded into a common stream s_c to be decoded by all CUs. All private parts are independently encoded into private streams $s_1, \cdots, s_{K_t}$. The vector of BS streams $\mathbf{s} = [s_c, s_1, \cdots, s_{K_t}]^T \in \mathbb{C}^{(K_t+1)\times 1}$ is therefore created, and we suppose it obeying $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}$. For the satellite, multicast messages $M_1, \cdots M_{N_s}$ are intended to the beams indexed by $\mathcal{N}_s = \{1, \cdots, N_s\}$. Each message $M_{n_s}, \forall n_s \in \mathcal{N}_s$ is split into a common part M_{c,n_s} and a private part M_{p,n_s}. All common parts are combined as M_c and encoded into m_c, while all private parts are independently encoded into $m_1, \cdots, m_{N_s}$. The vector of satellite streams $\mathbf{m} = [m_c, m_1, \cdots, m_{N_s}]^T \in \mathbb{C}^{(N_s+1)\times 1}$ is obtained, and we assume it satisfying $\mathbb{E}\left\{\mathbf{m}\mathbf{m}^H\right\} = \mathbf{I}$. Both $\mathbf{s}$ and $\mathbf{m}$ are linearly precoded. The transmitted signals at the satellite and BS are respectively

$$\mathbf{x}^{sat} = \mathbf{w}_c m_c + \sum_{n_s=1}^{N_s} \mathbf{w}_{n_s} m_{n_s} \quad \text{and} \quad \mathbf{x}^{bs} = \mathbf{p}_c s_c + \sum_{k_t=1}^{K_t} \mathbf{p}_{k_t} s_{k_t}, \tag{4.3}$$

[3]RSMA has been shown analytically as a general multiple access strategy, which boils down to SDMA and NOMA when allocating powers to the different types of message streams [9].

where $\mathbf{W} = [\mathbf{w}_c, \mathbf{w}_1, \cdots, \mathbf{w}_{N_s}] \in \mathbb{C}^{N_s \times (N_s+1)}$ and $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_{K_t}] \in \mathbb{C}^{N_t \times (K_t+1)}$ are defined as the beamforming matrices at the satellite and BS. m_c and s_c are superimposed on top of the private signals. Even though power-sharing mechanisms among beams can be implemented by using, e.g., multi-port amplifiers [97], the deployment of satellite payloads allowing flexible power allocation will require costly and complex radio-frequency designs. Thus, a per-feed transmit power constraint is considered, which is given by $(\mathbf{W}\mathbf{W}^H)_{n_s,n_s} \leq \frac{P_s}{N_s}$, $\forall n_s \in \mathcal{N}_s$. The sum transmit power constraint of BS is given by $\mathrm{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t$. Based on the channel models defined above, the received signal at each SU-k_s writes as

$$y_{k_s}^{sat} = \mathbf{f}_{k_s}^H \mathbf{w}_c m_c + \mathbf{f}_{k_s}^H \sum_{i=1}^{N_s} \mathbf{w}_i m_i + n_{k_s}^{sat}. \tag{4.4}$$

Since we assume all SUs are located outside the BS service area, each SU sees multibeam interference and no interference from the BS. The received signal at each CU-k_t writes as

$$y_{k_t}^{bs} = \mathbf{h}_{k_t}^H \mathbf{p}_c s_c + \mathbf{h}_{k_t}^H \sum_{j=1}^{K_t} \mathbf{p}_j s_j + \mathbf{z}_{k_t}^H \mathbf{w}_c m_c + \mathbf{z}_{k_t}^H \sum_{i=1}^{N_s} \mathbf{w}_i m_i + n_{k_t}^{bs}. \tag{4.5}$$

Each CU suffers from intra-cell interference and from satellite interference. $\mathbf{z}_1, \cdots, \mathbf{z}_{Kt}$ represent satellite interfering channels. $n_{k_s}^{sat}$ and $n_{k_t}^{bs}$ are the AWGN with zero mean and variance $\sigma_{k_s}^{sat2}$ and $\sigma_{k_t}^{bs2}$ respectively. For both SUs and CUs, the common stream is firstly decoded while treating the other interference as noise. The SINRs of decoding the common stream at SU-k_s and CU-k_t are given by

$$\gamma_{c,k_s}^{sat} = \frac{\left|\mathbf{f}_{k_s}^H \mathbf{w}_c\right|^2}{\sum_{i=1}^{N_s} \left|\mathbf{f}_{k_s}^H \mathbf{w}_i\right|^2 + \sigma_{k_s}^{sat2}}, \tag{4.6}$$

$$\gamma_{c,k_t}^{bs} = \frac{\left|\mathbf{h}_{k_t}^H \mathbf{p}_c\right|^2}{\sum_{j=1}^{K_t} \left|\mathbf{h}_{k_t}^H \mathbf{p}_j\right|^2 + \left|\mathbf{z}_{k_t}^H \mathbf{w}_c\right|^2 + \sum_{i=1}^{N_s} \left|\mathbf{z}_{k_t}^H \mathbf{w}_i\right|^2 + \sigma_{k_t}^{bs2}}. \tag{4.7}$$

Given perfect CSIT, the achievable rate of the common streams are $R_{c,k_s}^{sat} = \log_2(1 +$

γ_{c,k_s}^{sat}) and $R_{c,k_t}^{bs} = \log_2(1+\gamma_{c,k_t}^{bs})$. To guarantee that each SU is capable of decoding m_c, and each CU is capable of decoding s_c, they must be transmitted at rates that do not exceed

$$R_c^{sat} = \min_{k_s \in \mathcal{K}_s} \left\{ R_{c,k_s}^{sat} \right\} = \sum_{n_s=1}^{N_s} C_{n_s}^{sat} \quad \text{and} \quad R_c^{bs} = \min_{k_t \in \mathcal{K}_t} \left\{ R_{c,k_t}^{bs} \right\} = \sum_{k_t=1}^{K_t} C_{k_t}^{bs}, \tag{4.8}$$

where $C_{n_s}^{sat}$ is the portion of the common part of the n_s-th beam's message. $C_{k_t}^{bs}$ is the portion of the common part of the k_t-th CU's message. After the common stream is re-encoded, precoded and subtracted from the received signal through SIC, each user then decodes its desired private stream. We define $\mu : \mathcal{K}_s \to \mathcal{N}_s$ as mapping a SU to its corresponding beam. The SINRs of decoding $m_{\mu(k_s)}$ at SU-k_s and decoding s_{k_t} at CU-k_t are given by

$$\gamma_{k_s}^{sat} = \frac{\left| \mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)} \right|^2}{\sum_{i=1, i \neq \mu(k_s)}^{N_s} \left| \mathbf{f}_{k_s}^H \mathbf{w}_i \right|^2 + \sigma_{k_s}^{sat2}}, \tag{4.9}$$

$$\gamma_{k_t}^{bs} = \frac{\left| \mathbf{h}_{k_t}^H \mathbf{p}_{k_t} \right|^2}{\sum_{j=1, j \neq k_t}^{K_t} \left| \mathbf{h}_{k_t}^H \mathbf{p}_j \right|^2 + \left| \mathbf{z}_{k_t}^H \mathbf{w}_c \right|^2 + \sum_{i=1}^{N_s} \left| \mathbf{z}_{k_t}^H \mathbf{w}_i \right|^2 + \sigma_{k_t}^{bs2}}. \tag{4.10}$$

The achievable rates of the private streams are respectively $R_{k_s}^{sat} = \log_2(1+\gamma_{k_s}^{sat})$ and $R_{k_t}^{bs} = \log_2(1+\gamma_{k_t}^{bs})$. Thus, the achievable rates of the n_s-th beam and k_t-th CU respectively write as

$$R_{\text{tot},n_s}^{sat} = C_{n_s}^{sat} + \min_{i \in \mathcal{G}_{n_s}} R_i^{sat} \quad \text{and} \quad R_{\text{tot},k_t}^{bs} = C_{k_t}^{bs} + R_{k_t}^{bs}, \tag{4.11}$$

where $\mathcal{G}_{n_s}$ denotes the set of SUs belonging to the n_s-th beam.

2) Cooperative Scheme

Second, we consider a higher level of integration, i.e., *cooperative scheme* where both CSI and data are exchanged between the satellite and BS at the GW. In this scenario, all downlink messages $W_1, \cdots, W_{K_t}$ intended to CUs, and multicast

messages $M_1, \cdots, M_{N_s}$ intended to SUs are transmitted at both the satellite and BS. All propagation links (including interfering ones) are exploited to carry useful data upon appropriate beamforming. We still consider RSMA to manage interference in this *cooperative STIN*, including inter-beam interference, intra-cell interference and interference between the satellite and terrestrial sub-networks. Each message is split into a common part and a private part. All common parts are encoded together into a super common stream shared by all users in the system. As a result, the symbol stream to be transmitted is given by $\acute{\mathbf{s}} = [\acute{s}_c, \acute{m}_1, \cdots, \acute{m}_{N_s}, \acute{s}_1, \cdots, \acute{s}_{K_t}]^T \in \mathcal{C}^{N_s+K_t+1}$. Throughout this work, we use "$\acute{}$" to differentiate notations in the *cooperative scheme* and the above *coordinated scheme*. The transmitted signals at the satellite writes as

$$\acute{\mathbf{x}}^{sat} = \acute{\mathbf{w}}_c \acute{s}_c + \sum_{i=1}^{N_s} \acute{\mathbf{w}}_i^{sat} \acute{m}_i + \sum_{j=1}^{K_t} \acute{\mathbf{w}}_j^{bs} \acute{s}_j, \tag{4.12}$$

where $\acute{\mathbf{W}} = \left[\acute{\mathbf{w}}_c, \acute{\mathbf{w}}_1^{sat}, \cdots, \acute{\mathbf{w}}_{N_s}^{sat}, \acute{\mathbf{w}}_1^{bs}, \cdots, \acute{\mathbf{w}}_{K_t}^{bs}\right]$ is the beamforming matrix, and the superscripts of $\acute{\mathbf{w}}_i^{sat}$ and $\acute{\mathbf{w}}_j^{bs}$ are used to differentiate the precoder of satellite data and BS data. The per-feed transmit power constraint writes as $(\acute{\mathbf{W}}\acute{\mathbf{W}}^H)_{n_s,n_s} \leq \frac{P_s}{N_s}$, $\forall n_s \in \mathcal{N}_s$. Similarly, the transmitted signal at the BS writes as

$$\acute{\mathbf{x}}^{bs} = \acute{\mathbf{p}}_c \acute{s}_c + \sum_{i=1}^{N_s} \acute{\mathbf{p}}_i^{sat} \acute{m}_i + \sum_{j=1}^{K_t} \acute{\mathbf{p}}_j^{bs} \acute{s}_j, \tag{4.13}$$

where $\acute{\mathbf{P}} = \left[\acute{\mathbf{p}}_c, \acute{\mathbf{p}}_1^{sat}, \cdots, \acute{\mathbf{p}}_{N_s}^{sat}, \acute{\mathbf{p}}_1^{bs}, \cdots, \acute{\mathbf{p}}_{K_t}^{bs}\right]$ is the beamforming matrix, and the sum transmit power constraint of the BS is $\mathrm{tr}(\acute{\mathbf{P}}\acute{\mathbf{P}}^H) \leq P_t$. Accordingly, the received signal at SU-k_s is given by

$$\acute{y}_{k_s}^{sat} = \mathbf{f}_{k_s}^H \acute{\mathbf{w}}_c \acute{s}_c + \mathbf{f}_{k_s}^H \sum_{i=1}^{N_s} \acute{\mathbf{w}}_i^{sat} \acute{m}_i + \mathbf{f}_{k_s}^H \sum_{j=1}^{K_t} \acute{\mathbf{w}}_j^{bs} \acute{s}_j + \acute{n}_{k_s}^{sat}. \tag{4.14}$$

The received signal at CU-k_t is given by

$$\begin{aligned} \acute{y}_{k_t}^{bs} = \mathbf{h}_{k_t}^H \acute{\mathbf{p}}_c \acute{s}_c + \mathbf{h}_{k_t}^H \sum_{j=1}^{K_t} \acute{\mathbf{p}}_j^{bs} \acute{s}_j + \mathbf{h}_{k_t}^H \sum_{i=1}^{N_s} \acute{\mathbf{p}}_i^{sat} \acute{m}_i \\ + \mathbf{z}_{k_t}^H \acute{\mathbf{w}}_c \acute{s}_c + \mathbf{z}_{k_t}^H \sum_{i=1}^{N_s} \acute{\mathbf{w}}_i^{sat} \acute{m}_i + \mathbf{z}_{k_t}^H \sum_{j=1}^{K_t} \acute{\mathbf{w}}_j^{bs} \acute{s}_j + \acute{n}_{k_t}^{bs}. \end{aligned} \tag{4.15}$$

To simplify (4.15), aggregate channels and aggregate beamforming vectors are defined by

$$\mathbf{g}_{k_t} = \left[\mathbf{z}_{k_t}^H, \mathbf{h}_{k_t}^H\right]^H \in \mathbb{C}^{(N_s+N_t)\times 1}, \ \forall k_t \in \mathcal{K}_t, \tag{4.16}$$

$$\mathbf{v}_c = \left[\mathbf{w}_c^{*H}, \mathbf{p}_c^{*H}\right]^H \in \mathbb{C}^{(N_s+N_t)\times 1}, \tag{4.17}$$

$$\mathbf{v}_{n_s}^{sat} = \left[\acute{\mathbf{w}}_{n_s}^{satH}, \acute{\mathbf{p}}_{n_s}^{satH}\right]^H \in \mathbb{C}^{(N_s+N_t)\times 1}, \ \forall n_s \in \mathcal{N}_s, \tag{4.18}$$

$$\mathbf{v}_{k_t}^{bs} = \left[\acute{\mathbf{w}}_{k_t}^{bsH}, \acute{\mathbf{p}}_{k_t}^{bsH}\right]^H \in \mathbb{C}^{(N_s+N_t)\times 1}, \ \forall k_t \in \mathcal{K}_t. \tag{4.19}$$

The received signal at CU-k_t can be rewritten as

$$\acute{y}_{k_t}^{bs} = \mathbf{g}_{k_t}^H \mathbf{v}_c \acute{s}_c + \mathbf{g}_{k_t}^H \sum_{j=1}^{K_t} \mathbf{v}_j^{bs} \acute{s}_j + \mathbf{g}_{k_t}^H \sum_{i=1}^{N_s} \mathbf{v}_i^{sat} \acute{m}_i + \acute{n}_{kt}^{bs}. \tag{4.20}$$

Satellite interfering links are exploited to carry terrestrial data so as to improve the performance of STIN. The aggregate beamforming vectors are collected into a matrix

$$\mathbf{V} = [\mathbf{v}_c, \mathbf{v}_1^{sat}, \cdots, \mathbf{v}_{N_s}^{sat}, \mathbf{v}_1^{bs}, \cdots, \mathbf{v}_{K_t}^{bs}] \in \mathbb{C}^{(N_s+N_t)\times(N_s+K_t+1)}, \tag{4.21}$$

which can also be denoted by $\mathbf{V} = [\acute{\mathbf{W}}^H, \acute{\mathbf{P}}^H]^H$. For both SUs and CUs, the common stream is firstly decoded and removed from the received signal through SIC. The

SINRs of decoding $\acute{s}_c$ at the k_s-th SU and the k_t-th CU are respectively

$$\acute{\gamma}_{c,k_s}^{sat} = \frac{\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_c\right|^2}{\sum_{i=1}^{N_s}\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_i^{sat}\right|^2 + \sum_{j=1}^{K_t}\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_j^{bs}\right|^2 + \sigma_{k_s}^{sat2}}, \tag{4.22}$$

$$\acute{\gamma}_{c,k_t}^{bs} = \frac{\left|\mathbf{g}_{k_t}^H \mathbf{v}_c\right|^2}{\sum_{j=1}^{K_t}\left|\mathbf{g}_{k_t}^H \mathbf{v}_j^{bs}\right|^2 + \sum_{i=1}^{N_s}\left|\mathbf{g}_{k_t}^H \mathbf{v}_i^{sat}\right|^2 + \sigma_{k_t}^{bs2}}. \tag{4.23}$$

The corresponding achievable rates are $\acute{R}_{c,k_s}^{sat} = \log_2(1+\acute{\gamma}_{c,k_s}^{sat})$ and $\acute{R}_{c,k_t}^{bs} = \log_2(1+\acute{\gamma}_{c,k_t}^{bs})$. Since $\acute{s}_c$ is decoded by all users in the system, we define the common rate as

$$\acute{R}_c = \min_{k_s\in\mathcal{K}_s, k_t\in\mathcal{K}_t}\left\{\acute{R}_{c,k_s}^{sat}, \acute{R}_{c,k_t}^{bs}\right\} = \sum_{n_s=1}^{N_s}\acute{C}_{n_s}^{sat} + \sum_{k_t=1}^{K_t}\acute{C}_{k_t}^{bs}. \tag{4.24}$$

Note that $\acute{s}_c$ is shared amongst all satellite beams and CUs. $\acute{C}_{n_s}^{sat}$ and $\acute{C}_{k_t}^{bs}$ respectively correspond to the beam-n_s's and CU-k_t's portion of common rate. After removing $\acute{s}_c$ using SIC, each user then decodes its desired private stream. The SINRs of decoding private streams are

$$\acute{\gamma}_{k_s}^{sat} = \frac{\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_{\mu(k_s)}^{sat}\right|^2}{\sum_{i=1,i\neq\mu(k_s)}^{N_s}\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_i^{sat}\right|^2 + \sum_{j=1}^{K_t}\left|\mathbf{f}_{k_s}^H \acute{\mathbf{w}}_j^{bs}\right|^2 + \sigma_{k_s}^{sat2}}, \tag{4.25}$$

$$\acute{\gamma}_{k_t}^{bs} = \frac{\left|\mathbf{g}_{k_t}^H \mathbf{v}_{k_t}^{bs}\right|^2}{\sum_{j=1,j\neq k_t}^{K_t}\left|\mathbf{g}_{k_t}^H \mathbf{v}_j^{bs}\right|^2 + \sum_{i=1}^{N_s}\left|\mathbf{g}_{k_t}^H \mathbf{v}_i^{sat}\right|^2 + \sigma_{k_t}^{bs2}}. \tag{4.26}$$

$\acute{R}_{k_s}^{sat} = \log_2(1+\acute{\gamma}_{k_s}^{sat})$ and $\acute{R}_{k_t}^{bs} = \log_2(1+\acute{\gamma}_{k_t}^{bs})$ are the achievable rates of the private streams. Thus, the achievable rates of the n_s-th beam and k_t-th CU respectively write as

$$\acute{R}_{\mathrm{tot},n_s}^{sat} = \acute{C}_{n_s}^{sat} + \min_{i\in\mathcal{G}_{n_s}}\acute{R}_i^{sat} \quad \text{and} \quad \acute{R}_{\mathrm{tot},k_t}^{bs} = \acute{C}_{k_t}^{bs} + \acute{R}_{k_t}^{bs}. \tag{4.27}$$

From the above expressions, we can regard the satellite and BS working together as a super "BS" but subject to their respective power constraints to serve the CUs and SUs. The super common stream contains parts of the unicast messages intended

to the CUs, and parts of the multicast messages intended to the SUs. At each user side, the super common stream is at first decoded and then removed through SIC. Accordingly, the interference is partially decoded. Each user then decodes its private stream and treats the remaining interference as noise. Such scheme has the capability to better manage interference including not only inter-beam interference, intra-cell interference, but also interference between the satellite and terrestrial sub-networks.

Remark 4.1: With the assumption of Gaussian signalling and infinite block length, there is no decoding error in SIC. Decoding errors in SIC would only occur if we depart from Shannon assumptions and assume finite constellations and finite block lengths. We consider one-layer RSMA for either coordinated scheme and cooperative scheme. Only one layer of SIC is required at each terminal. The receiver complexity does not depend on the number of served users. The generalized RSMA and hierarchical RSMA described in [5] is able to provide more room for achievable rate enhancements at the expense of more layers of SIC at receivers. However, its implementation can be complex due to the large number of SIC layers and common messages involved. The receiver complexity of generalized RSMA and hierarchical RSMA increases with the number of served users. Moreover, ordering and grouping are not required in this one-layer RSMA architecture since all users decode the common stream before decoding their private streams. Both scheduling complexity and receiver complexity are reduced tremendously. Readers are referred to [5] and [98] for more details on complexity issues.

4.3 Proposed Joint Beamforming Scheme

In this section, the problem of interest is to design a joint beamforming scheme to maximize the minimum fairness rate amongst all unicast CUs and multibeam

multicast SUs subject to transmit power constraints. We respectively consider the scenarios of RSMA-based *coordinated STIN* and *cooperative STIN* with perfect CSI at the GW.

4.3.1 Joint Beamforming Design for Coordinated STIN

For RSMA-based *coordinated STIN*, the optimization problem to maximize the minimum fairness rate can be formulated as

$$\mathcal{P}_1: \max_{\mathbf{W},\mathbf{P},\mathbf{c}^{sat},\mathbf{c}^{bs}} \min_{n_s\in\mathcal{N}_s,k_t\in\mathcal{K}_t} \left\{R^{bs}_{\text{tot},k_t}, R^{sat}_{\text{tot},n_s}\right\} \tag{4.28}$$

$$s.t. \quad R^{bs}_{c,k_t} \geq \sum_{j=1}^{K_t} C^{bs}_j, \quad \forall k_t \in \mathcal{K}_t \tag{4.29}$$

$$C^{bs}_{k_t} \geq 0, \quad \forall k_t \in \mathcal{K}_t \tag{4.30}$$

$$\text{tr}(\mathbf{P}\mathbf{P}^H) \leq P_t \tag{4.31}$$

$$R^{sat}_{c,k_s} \geq \sum_{j=1}^{N_s} C^{sat}_j, \quad \forall k_s \in \mathcal{K}_s \tag{4.32}$$

$$C^{sat}_{n_s} \geq 0, \quad \forall n_s \in \mathcal{N}_s \tag{4.33}$$

$$(\mathbf{W}\mathbf{W}^H)_{n_s,n_s} \leq \frac{P_s}{N_s}, \quad \forall n_s \in \mathcal{N}_s \tag{4.34}$$

where $\mathbf{c}^{sat} = [C^{sat}_1, \cdots, C^{sat}_{N_s}]^T$, $\mathbf{c}^{bs} = [C^{bs}_1, \cdots, C^{bs}_{K_t}]^T$ are the vectors of common rate portions. (4.29) guarantees that the common stream s_c can be decoded by all CUs. (4.31) is the sum transmit power constraint of the BS. Similarly, (4.32) ensures the common stream m_c to be decoded by all SUs. (4.34) represents the per-feed transmit power constraint of the satellite. (4.30) and (4.33) guarantee the non-negativity of all common rate portions.

Note that the formulated problem is non-convex, we exploit an SCA-based method to convexify the non-convex constraints and approximate the non-convex problem

to a convex one. First, we introduce an equivalent reformulation of $\mathcal{P}_1$, which is

$$\mathcal{E}_1: \quad \max_{\mathbf{W},\mathbf{P},\mathbf{c}^{sat},\mathbf{c}^{bs},q,\mathbf{r},\boldsymbol{\alpha}} q \tag{4.35}$$

$$s.t. \quad C_{k_t}^{bs} + \alpha_{k_t} \geq q, \quad \forall k_t \in \mathcal{K}_t \tag{4.36}$$

$$R_{k_t}^{bs} \geq \alpha_{k_t}, \quad \forall k_t \in \mathcal{K}_t \tag{4.37}$$

$$C_{n_s}^{sat} + r_{k_s} \geq q, \quad \forall k_s \in \mathcal{G}_{n_s} \tag{4.38}$$

$$R_{k_s}^{sat} \geq r_{k_s}, \quad \forall k_s \in \mathcal{K}_s \tag{4.39}$$

$$(4.29) - (4.34)$$

where q, $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_{K_t}]^T$, $\mathbf{r} = [r_1, \cdots, r_{K_s}]^T$ are introduced auxiliary variables. To deal with the non-convexity of (4.29), (4.32), (4.37), (4.39), we further introduce new auxiliary variables $\mathbf{a} = [a_1, \cdots, a_{K_t}]^T$, $\mathbf{a}_c = [a_{c,1}, \cdots, a_{c,K_t}]^T$, $\mathbf{b} = [b_1, \cdots, b_{K_s}]^T$ and $\mathbf{b}_c = [b_{c,1}, \cdots, b_{c,K_s}]^T$. The problem $\mathcal{E}_1$ can be rewritten as

$$\mathcal{S}_1: \quad \max_{q,\mathbf{W},\mathbf{P},\mathbf{c}^{sat},\mathbf{c}^{bs},\mathbf{r},\boldsymbol{\alpha},\mathbf{a},\mathbf{a}_c,\mathbf{b},\mathbf{b}_c} q \tag{4.40}$$

$$s.t. \quad \log(1 + a_{k_t}) \geq \alpha_{k_t} \log 2, \quad \forall k_t \in \mathcal{K}_t \tag{4.41}$$

$$\gamma_{k_t}^{bs} \geq a_{k_t}, \quad \forall k_t \in \mathcal{K}_t \tag{4.42}$$

$$\log(1 + b_{k_s}) \geq r_{k_s} \log 2, \quad \forall k_s \in \mathcal{K}_s \tag{4.43}$$

$$\gamma_{k_s}^{sat} \geq b_{k_s}, \quad \forall k_s \in \mathcal{K}_s \tag{4.44}$$

$$\log(1 + a_{c,k_t}) \geq \sum_{j=1}^{K_t} C_j^{bs} \log 2, \quad \forall k_t \in \mathcal{K}_t \tag{4.45}$$

$$\gamma_{c,k_t}^{bs} \geq a_{c,k_t}, \quad \forall k_t \in \mathcal{K}_t \tag{4.46}$$

$$\log(1 + b_{c,k_s}) \geq \sum_{j=1}^{N_s} C_j^{sat} \log 2, \quad \forall k_s \in \mathcal{K}_s \tag{4.47}$$

$$\gamma_{c,k_s}^{sat} \geq b_{c,k_s}, \quad \forall k_s \in \mathcal{K}_s \tag{4.48}$$

$$(4.30), (4.31), (4.33), (4.34), (4.36), (4.38)$$

where (4.41) - (4.48) are obtained by extracting the SINRs from the rate expressions $R_{k_t}^{bs}, R_{k_s}^{sat}, R_{c,k_t}^{bs}, R_{c,k_s}^{sat}$ in (4.29), (4.32), (4.37), (4.39) of Problem $\mathcal{E}_1$. Since the constraints of $\mathcal{S}_1$ hold with equality at optimality, the equivalence between $\mathcal{P}_1$ and $\mathcal{S}_1$ can be guaranteed. Now, the non-convexity of $\mathcal{S}_1$ comes from (4.42), (4.44), (4.46) and (4.48) which contain SINR expressions. (4.42) can be expanded as

$$\sum_{j=1,j\neq k_t}^{K_t} \left|\mathbf{h}_{k_t}^H \mathbf{p}_j\right|^2 + \left|\mathbf{z}_{k_t}^H \mathbf{w}_c\right|^2 + \sum_{i=1}^{N_s} \left|\mathbf{z}_{k_t}^H \mathbf{w}_i\right|^2 + \sigma_{k_t}^{bs2} \leq \frac{\left|\mathbf{h}_{k_t}^H \mathbf{p}_{k_t}\right|^2}{a_{k_t}}, \tag{4.49}$$

where the right-hand side quadratic-over-linear function is convex. We approximate it with its lower bound, which is obtained by the first-order Taylor approximation around the point $\left(\mathbf{p}_{k_t}^{[n]}, a_{k_t}^{[n]}\right)$. Then, we have

$$\begin{aligned}\frac{\left|\mathbf{h}_{k_t}^H \mathbf{p}_{k_t}\right|^2}{a_{k_t}} &\geq \frac{2\mathcal{R}\left(\mathbf{p}_{k_t}^{[n]H} \mathbf{h}_{k_t} \mathbf{h}_{k_t}^H \mathbf{p}_{k_t}\right)}{a_{k_t}^{[n]}} - \frac{\mathbf{p}_{k_t}^{[n]H} \mathbf{h}_{k_t} \mathbf{h}_{k_t}^H \mathbf{p}_{k_t}^{[n]}}{\left(a_{k_t}^{[n]}\right)^2} a_{k_t} \\ &\triangleq \widehat{f}_1\left(\mathbf{p}_{k_t}, a_{k_t}; \mathbf{p}_{k_t}^{[n]}, a_{k_t}^{[n]}\right)\end{aligned} \tag{4.50}$$

where n represents the n-th SCA iteration. Replacing the linear approximation $\widehat{f}_1\left(\mathbf{p}_{k_t}, a_{k_t}; \mathbf{p}_{k_t}^{[n]}, a_{k_t}^{[n]}\right)$ with the right-hand side of (4.49) yields

$$\sum_{j=1,j\neq k_t}^{K_t} \left|\mathbf{h}_{k_t}^H \mathbf{p}_j\right|^2 + \left|\mathbf{z}_{k_t}^H \mathbf{w}_c\right|^2 + \sum_{i=1}^{N_s} \left|\mathbf{z}_{k_t}^H \mathbf{w}_i\right|^2 + \sigma_{k_t}^{bs2} - \widehat{f}_1\left(\mathbf{p}_{k_t}, a_{k_t}; \mathbf{p}_{k_t}^{[n]}, a_{k_t}^{[n]}\right) \leq 0. \tag{4.51}$$

Similarly, the constraint (4.44) can be expanded as

$$\sum_{i=1,i\neq\mu(k_s)}^{N_s} \left|\mathbf{f}_{k_s}^H \mathbf{w}_i\right|^2 + \sigma_{k_s}^{sat2} \leq \frac{\left|\mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)}\right|^2}{b_{k_s}}. \tag{4.52}$$

We approximate its right-hand side around the point $\left(\mathbf{w}_{\mu(k_s)}^{[n]}, b_{k_s}^{[n]}\right)$, and obtain

$$\frac{\left|\mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)}\right|^2}{b_{k_s}} \geq \frac{2\mathcal{R}\left(\mathbf{w}_{\mu(k_s)}^{[n]H} \mathbf{f}_{k_s} \mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)}\right)}{b_{k_s}^{[n]}} - \frac{\mathbf{w}_{\mu(k_s)}^{[n]H} \mathbf{f}_{k_s} \mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)}^{[n]}}{\left(b_{k_s}^{[n]}\right)^2} b_{k_s}$$
$$\triangleq \widehat{f}_2\left(\mathbf{w}_{\mu(k_s)}, b_{k_s}; \mathbf{w}_{\mu(k_s)}^{[n]}, b_{k_s}^{[n]}\right). \tag{4.53}$$

Replacing $\widehat{f}_2\left(\mathbf{w}_{\mu(k_s)}, b_{k_s}; \mathbf{w}_{\mu(k_s)}^{[n]}, b_{k_s}^{[n]}\right)$ with the right-hand side of (4.52) yields

$$\sum_{i=1, i\neq\mu(k_s)}^{N_s} \left|\mathbf{f}_{k_s}^H \mathbf{w}_i\right|^2 + \sigma_{k_s}^{sat2} - \widehat{f}_2\left(\mathbf{w}_{\mu(k_s)}, b_{k_s}; \mathbf{w}_{\mu(k_s)}^{[n]}, b_{k_s}^{[n]}\right) \leq 0. \tag{4.54}$$

Following the same logic, (4.46) and (4.48) are respectively approximated by

$$\left|\mathbf{h}_{k_t}^H \mathbf{p}_{k_t}\right|^2 + \sum_{j=1, j\neq k_t}^{K_t} \left|\mathbf{h}_{k_t}^H \mathbf{p}_j\right|^2 + \left|\mathbf{z}_{k_t}^H \mathbf{w}_c\right|^2 + \sum_{i=1}^{N_s} \left|\mathbf{z}_{k_t}^H \mathbf{w}_i\right|^2 + \sigma_{k_t}^{bs2}$$
$$- \widehat{f}_3\left(\mathbf{p}_c, a_{c,k_t}; \mathbf{p}_c^{[n]}, a_{c,k_t}^{[n]}\right) \leq 0, \tag{4.55}$$
$$\left|\mathbf{f}_{k_s}^H \mathbf{w}_{\mu(k_s)}\right|^2 + \sum_{i=1, i\neq\mu(k_s)}^{N_s} \left|\mathbf{f}_{k_s}^H \mathbf{w}_i\right|^2 + \sigma_{k_s}^{sat2} - \widehat{f}_4\left(\mathbf{w}_c, b_{c,k_s}; \mathbf{w}_c^{[n]}, b_{c,k_s}^{[n]}\right) \leq 0, \tag{4.56}$$

where $\widehat{f}_3\left(\mathbf{p}_c, a_{c,k_t}; \mathbf{p}_c^{[n]}, a_{c,k_t}^{[n]}\right)$ and $\widehat{f}_4\left(\mathbf{w}_c, b_{c,k_s}; \mathbf{w}_c^{[n]}, b_{c,k_s}^{[n]}\right)$ are linear lower bound expressions given by

$$\widehat{f}_3\left(\mathbf{p}_c, a_{c,k_t}; \mathbf{p}_c^{[n]}, a_{c,k_t}^{[n]}\right) \triangleq \frac{2\mathcal{R}\left(\mathbf{p}_c^{[n]H} \mathbf{h}_{k_t} \mathbf{h}_{k_t}^H \mathbf{p}_c\right)}{a_{c,k_t}^{[n]}} - \frac{\mathbf{p}_c^{[n]H} \mathbf{h}_{k_t} \mathbf{h}_{k_t}^H \mathbf{p}_c^{[n]}}{\left(a_{c,k_t}^{[n]}\right)^2} a_{c,k_t}, \tag{4.57}$$
$$\widehat{f}_4\left(\mathbf{w}_c, b_{c,k_s}; \mathbf{w}_c^{[n]}, b_{c,k_s}^{[n]}\right) \triangleq \frac{2\mathcal{R}\left(\mathbf{w}_c^{[n]H} \mathbf{f}_{k_s} \mathbf{f}_{k_s}^H \mathbf{w}_c\right)}{b_{c,k_s}^{[n]}} - \frac{\mathbf{w}_c^{[n]H} \mathbf{f}_{k_s} \mathbf{f}_{k_s}^H \mathbf{w}_c^{[n]}}{\left(b_{c,k_s}^{[n]}\right)^2} b_{c,k_s}. \tag{4.58}$$

Although (4.41), (4.43), (4.45) and (4.47) are convex constraints, which are solvable through the CVX toolbox in Matlab, the log terms belong to generalized nonlinear convex program with high computational complexity. Aiming at more efficient implementation, [99] approximates the log constraints to a set of SOC constraints, which introduce a great number of slack variables and result in an increase of

per-iteration complexity. Here, we use the property that $x \log(1+x)$ is convex as in [100], and approximate (4.41), (4.43), (4.45), (4.47) without additional slack variables. Since $a_{k_t} \geq 0$, the constraint (4.41) can be rewritten as

$$a_{k_t} \log(1 + a_{k_t}) \geq a_{k_t} \alpha_{k_t} \log 2. \tag{4.59}$$

Its left-hand side is convex, so we compute the first-order Taylor approximation of $a_{k_t} \log(1 + a_{k_t})$ around the point $a_{k_t}^{[n]}$ as

$$\begin{aligned} a_{k_t} \log(1 + a_{k_t}) &\geq a_{k_t}^{[n]} \log\left(1 + a_{k_t}^{[n]}\right) + \left(a_{k_t} - a_{k_t}^{[n]}\right) \left[\frac{a_{k_t}^{[n]}}{1 + a_{k_t}^{[n]}} + \log\left(1 + a_{k_t}^{[n]}\right)\right] \\ &= a_{k_t} v_{k_t}^{[n]} - u_{k_t}^{[n]}, \end{aligned} \tag{4.60}$$

where $v_{k_t}^{[n]}$ and $u_{k_t}^{[n]}$ are expressions of $a_{k_t}^{[n]}$ given by

$$v_{k_t}^{[n]} = \frac{a_{k_t}^{[n]}}{a_{k_t}^{[n]} + 1} + \log\left(1 + a_{k_t}^{[n]}\right) \quad \text{and} \quad u_{k_t}^{[n]} = \frac{\left(a_{k_t}^{[n]}\right)^2}{a_{k_t}^{[n]} + 1}. \tag{4.61}$$

Now, (4.59) can be rewritten by $a_{k_t} v_{k_t}^{[n]} - u_{k_t}^{[n]} \geq a_{k_t} \alpha_{k_t} \log 2$, which is SOC representable [101] as

$$\left\| \left[a_{k_t} + \alpha_{k_t} \log 2 - v_{k_t}^{[n]} \quad 2\sqrt{u_{k_t}^{[n]}} \right] \right\|_2 \leq a_{k_t} - \alpha_{k_t} \log 2 + v_{k_t}^{[n]}. \tag{4.62}$$

Similarly, the constraint (4.43), (4.45), (4.47) can be replaced by

$$\left\| \left[b_{k_s} + r_{k_s} \log 2 - \bar{v}_{k_s}^{[n]} \quad 2\sqrt{\bar{u}_{k_s}^{[n]}} \right] \right\|_2 \leq b_{k_s} - r_{k_s} \log 2 + \bar{v}_{k_s}^{[n]}, \tag{4.63}$$

$$\left\| \left[a_{c,k_t} + \sum_{j=1}^{K_t} C_j^{bs} \log 2 - v_{c,k_t}^{[n]} \quad 2\sqrt{u_{c,k_t}^{[n]}} \right] \right\|_2 \leq a_{c,k_t} - \sum_{j=1}^{K_t} C_j^{bs} \log 2 + v_{c,k_t}^{[n]}, \tag{4.64}$$

$$\left\| \left[b_{c,k_s} + \sum_{j=1}^{N_s} C_j^{sat} \log 2 - \bar{v}_{c,k_s}^{[n]} \quad 2\sqrt{\bar{u}_{c,k_s}^{[n]}} \right] \right\|_2 \leq b_{c,k_s} - \sum_{j=1}^{N_s} C_j^{sat} \log 2 + \bar{v}_{c,k_s}^{[n]}. \tag{4.65}$$

The expressions of $\bar{v}_{k_s}^{[n]}$, $\bar{u}_{k_s}^{[n]}$, $v_{c,k_t}^{[n]}$, $u_{c,k_t}^{[n]}$, $\bar{v}_{c,k_s}^{[n]}$, $\bar{u}_{c,k_s}^{[n]}$ are respectively

$$\begin{aligned}
\bar{v}_{k_s}^{[n]} &= \frac{b_{k_s}^{[n]}}{b_{k_s}^{[n]}+1} + \log\left(1+b_{k_s}^{[n]}\right) \text{ and } \bar{u}_{k_s}^{[n]} = \frac{\left(b_{k_s}^{[n]}\right)^2}{b_{k_s}^{[n]}+1}, \\
v_{c,k_t}^{[n]} &= \frac{a_{c,k_t}^{[n]}}{a_{c,k_t}^{[n]}+1} + \log\left(1+a_{c,k_t}^{[n]}\right) \text{ and } u_{c,k_t}^{[n]} = \frac{\left(a_{c,k_t}^{[n]}\right)^2}{a_{c,k_t}^{[n]}+1}, \\
\bar{v}_{c,k_s}^{[n]} &= \frac{b_{c,k_s}^{[n]}}{b_{c,k_s}^{[n]}+1} + \log\left(1+b_{c,k_s}^{[n]}\right) \text{ and } \bar{u}_{c,k_s}^{[n]} = \frac{\left(b_{c,k_s}^{[n]}\right)^2}{b_{c,k_s}^{[n]}+1}.
\end{aligned} \tag{4.66}$$

By replacing the constraints (45)-(52) with (55), (58), (59), (60), (66)-(69), we obtain

$$\begin{aligned}
\mathcal{A}_1: \quad & \max_{q,\mathbf{W},\mathbf{P},\mathbf{c}^{sat},\mathbf{c}^{bs},\mathbf{r},\boldsymbol{\alpha},\mathbf{a},\mathbf{a}_c,\mathbf{b},\mathbf{b}_c} q \\
s.t. \quad & (4.51), (4.55), (4.62), (4.64), \quad \forall k_t \in \mathcal{K}_t \\
& (4.54), (4.56), (4.63), (4.65), \quad \forall k_s \in \mathcal{K}_s \\
& (4.30), (4.31), (4.33), (4.34), (4.36), (4.38)
\end{aligned} \tag{4.67}$$

The n-th iteration of the problem $\mathcal{A}_1$ belongs to SOCP and can be efficiently solved by the standard solvers in CVX. In each iteration, the problem defined around the solution of the previous iteration is solved. Variables are updated iteratively until a stopping criterion is satisfied. We summarize the procedure of this RSMA-based joint beamforming scheme in Algorithm 2. ε is the tolerance value. The optimal solution of Problem $\mathcal{A}_1$ at iteration-n is a feasible solution of the problem at iteration-$(n+1)$. As a consequence, the objective variable q increases monotonically. It is bounded above by the transmit power constraints. The proposed Algorithm 2 is guaranteed to converge while the global optimality of the achieved solution can not be guaranteed. The solution of the proposed SCA-based algorithm converges to the set of KKT points (which is also known as the stationary points) of problem $\mathcal{P}_1$ [102].

Algorithm 2 Proposed Joint Beamforming Scheme

Initialize: $n \leftarrow 0$, $\mathbf{W}^{[n]}, \mathbf{P}^{[n]}, \mathbf{a}^{[n]}, \mathbf{a}_c^{[n]}, \mathbf{b}^{[n]}, \mathbf{b}_c^{[n]}, q^{[n]}$;
repeat
Solve the problem $\mathcal{A}_1$ at $\left(\mathbf{W}^{[n]}, \mathbf{P}^{[n]}, \mathbf{a}^{[n]}, \mathbf{a}_c^{[n]}, \mathbf{b}^{[n]}, \mathbf{b}_c^{[n]}\right)$ to get the optimal solution $\left(\breve{\mathbf{W}}, \breve{\mathbf{P}}, \breve{\mathbf{a}}, \breve{\mathbf{a}}_c, \breve{\mathbf{b}}, \breve{\mathbf{b}}_c, \breve{q}\right)$;
$n \leftarrow n+1$;
Update $\mathbf{W}^{[n]} \leftarrow \breve{\mathbf{W}}, \mathbf{P}^{[n]} \leftarrow \breve{\mathbf{P}}, \mathbf{a}^{[n]} \leftarrow \breve{\mathbf{a}}, \mathbf{a}_c^{[n]} \leftarrow \breve{\mathbf{a}}_c, \mathbf{b}^{[n]} \leftarrow \breve{\mathbf{b}}, \mathbf{b}_c^{[n]} \leftarrow \breve{\mathbf{b}}_c, q^{[n]} \leftarrow \breve{q}$;
until $\left|q^{[n]} - q^{[n-1]}\right| < \varepsilon$;

4.3.2 Joint Beamforming Design for Cooperative STIN

When RSMA-based *cooperative STIN* is considered, the optimization problem of max-min fairness rate among all users is given by

$$\mathcal{P}_2: \quad \max_{\acute{\mathbf{W}}, \acute{\mathbf{P}}, \acute{\mathbf{c}}} \min_{n_s \in \mathcal{N}_s, k_t \in \mathcal{K}_t} \left\{\acute{R}_{k_t}^{bs}, \acute{R}_{k_s}^{sat}\right\} \tag{4.68}$$

$$s.t. \quad \acute{R}_{c,k_t}^{bs} \geq \sum_{j=1}^{K_t} \acute{C}_j^{bs} + \sum_{j=1}^{N_s} \acute{C}_j^{sat}, \quad \forall k_t \in \mathcal{K}_t \tag{4.69}$$

$$\acute{C}_{k_t}^{bs} \geq 0 \quad \forall k_t \in \mathcal{K}_t \tag{4.70}$$

$$\mathrm{tr}\left(\acute{\mathbf{P}}\acute{\mathbf{P}}^H\right) \leq P_t \tag{4.71}$$

$$\acute{R}_{c,k_s}^{sat} \geq \sum_{j=1}^{K_t} \acute{C}_j^{bs} + \sum_{j=1}^{N_s} \acute{C}_j^{sat}, \quad \forall k_s \in \mathcal{K}_s \tag{4.72}$$

$$\acute{C}_{n_s}^{sat} \geq 0, \quad \forall n_s \in \mathcal{N}_s \tag{4.73}$$

$$\left(\acute{\mathbf{W}}\acute{\mathbf{W}}^H\right)_{n_s,n_s} \leq \frac{P_s}{N_s}, \quad \forall n_s \in \mathcal{N}_s \tag{4.74}$$

where $\acute{\mathbf{c}} = [\acute{C}_1^{sat}, \cdots, \acute{C}_{N_s}^{sat}, \acute{C}_1^{bs}, \cdots, \acute{C}_{K_t}^{bs}]^T$ is the vector of all common rate portions. (4.69) and (4.72) guarantee that the common stream of the whole system $\acute{s}_c$ can be decoded by all SUs and CUs. (4.70) and (4.73) ensure non-negativity of each element in $\acute{\mathbf{c}}$. (4.71) and (4.74) are respectively the sum transmit power constraint of the BS and per-feed transmit power constraints of the satellite. The formulated MMF problem for *cooperative STIN* is also non-convex. Note that the main difference

between $\mathcal{P}_1$ and $\mathcal{P}_2$ lies in the transmit data information sharing in $\mathcal{P}_2$. One super common stream is transmitted at both the satellite and BS instead of transmitting individual common streams. The achievable rate expressions and beamforming matrices of *cooperative STIN* have been given in Section 4.2.2. We can still use the SCA-based algorithm to solve $\mathcal{P}_2$. Here, we omit the detailed problem transformation and optimization framework, which follow the same procedure as that for $\mathcal{P}_1$.

4.4 Robust Joint Beamforming Scheme

Here, we further investigate the beamforming design for RSMA-based *coordinated STIN* and *cooperative STIN* considering satellite channel phase uncertainty. The CSIT of terrestrial channels is assumed to be perfect. From the satellite channel model, we can observe that the amplitudes of the channel vector components are determined by some constant coefficients during the coherence time interval, including the free space loss, satellite antenna gain and rain attenuation [62]. However, the satellite channel phases vary rapidly due to a series of time-varying factors, such as the use of different local oscillators (LO) on-board, the rain, cloud and gaseous absorption, and the use of low-noise block (LNB) at receivers [11, 62]. Therefore, within a coherence time interval, the phase of the channel vector from the satellite to SU-k_s at time instant t_1 can be modeled as

$$\boldsymbol{\phi}_{k_s}(t_1) = \boldsymbol{\phi}_{k_s}(t_0) + \mathbf{e}_{k_s}, \tag{4.75}$$

where $\boldsymbol{\phi}_{k_s}(t_0)$ represents the phase vector, which is estimated at the previous time instant t_0 and fed back to the GW. $\mathbf{e}_{k_s} = [e_{k_s,1}, e_{k_s,2}, \cdots, e_{k_s,N_s}]^T$ is the phase uncertainty following the distribution $\mathbf{e}_{k_s} \sim \mathcal{N}(\mathbf{0}, \delta^2\mathbf{I})$, with i.i.d Gaussian random entries. For ease of notation, we can generally indicate $\boldsymbol{\phi}_{k_s}(t_1)$ and $\boldsymbol{\phi}_{k_s}(t_0)$ by $\boldsymbol{\phi}_{k_s}$ and $\widehat{\boldsymbol{\phi}}_{k_s}$ respectively. Since we assume blueconstant channel amplitudes within the

coherence time interval, the channel vector from the satellite to SU-k_s is written as

$$\mathbf{f}_{k_s} = \widehat{\mathbf{f}}_{k_s} \odot \mathbf{x}_{k_s} = \mathrm{diag}\big(\widehat{\mathbf{f}}_{k_s}\big)\mathbf{x}_{k_s}, \tag{4.76}$$

where $\mathbf{x}_{k_s} = \exp\{j\mathbf{e}_{k_s}\}$ is a random vector. We further assume that the channel estimate $\widehat{\mathbf{f}}_{k_s}$ and the correlation matrix of $\mathbf{x}_{k_s}$ denoted by $\mathbf{X}_{k_s} = \mathbb{E}\left\{\mathbf{x}_{k_s}\mathbf{x}_{k_s}^H\right\}$ are known at the GW [62]. For the interfering channels, by defining $\mathbf{y}_{k_t} = \exp\left\{j\mathbf{e}'_{k_t}\right\}$ and $\mathbf{e}'_{k_t} = \left[e'_{k_t,1}, e'_{k_t,2}, \cdots, e'_{k_t,N_s}\right]^T$ following $\mathbf{e}'_{k_t} \sim \mathcal{N}(\mathbf{0}, \delta^2\mathbf{I})$, the channel vector from the satellite to CU-k_t write as

$$\mathbf{z}_{k_t} = \widehat{\mathbf{z}}_{k_t} \odot \mathbf{y}_{k_t} = \mathrm{diag}\big(\widehat{\mathbf{z}}_{k_t}\big)\mathbf{y}_{k_t}, \tag{4.77}$$

where the channel estimate $\widehat{\mathbf{z}}_{k_t}$ and the correlation matrix $\mathbf{Y}_{k_t} = \mathbb{E}\left\{\mathbf{y}_{k_t}\mathbf{y}_{k_t}^H\right\}$ are available at the GW. Hence, we concentrate on the expectation-based robust beamforming design. The MMF optimization problem for RSMA-based *coordinated STIN* considering satellite phase uncertainty remains the same as $\mathcal{P}_1$ in Section 4.3.1, By introducing auxiliary variables q, $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_{K_t}]^T$, $\mathbf{r} = [r_1, \cdots, r_{K_s}]^T$, $W = \{\mathbf{W}_c, \mathbf{W}_1, \cdots, \mathbf{W}_{N_s}\}$ and $P = \{\mathbf{P}_c, \mathbf{P}_1, \cdots, \mathbf{P}_{K_t}\}$, the original $\mathcal{P}_1$ can be equivalently transformed into semi-definite programming (SDP) form with rank-one constraints

$$\mathcal{D}_1: \quad \max_{W,P,\mathbf{c}^{sat},\mathbf{c}^{bs},q,\mathbf{r},\boldsymbol{\alpha}} q \tag{4.78}$$

$$s.t. \quad \mathrm{tr}\left(\mathbf{P}_c\right) + \sum_{k_t=1}^{K_t} \mathrm{tr}\left(\mathbf{P}_{k_t}\right) \leq P_t \tag{4.79}$$

$$\Big[\mathbf{W}_c + \sum_{i=1}^{N_s} \mathbf{W}_i\Big]_{n_s,n_s} \leq \frac{P_s}{N_s}, \quad n_s \in \mathcal{N}_s \tag{4.80}$$

$$\mathbf{W}_c \succeq 0, \ \mathbf{W}_{n_s} \succeq 0, \quad \forall n_s \in \mathcal{N}_s \tag{4.81}$$

$$\mathbf{P}_c \succeq 0, \ \mathbf{P}_{k_t} \succeq 0, \quad \forall k_t \in \mathcal{K}_t \tag{4.82}$$

$$\text{rank}\left(\mathbf{W}_c\right)=1,\ \text{rank}\left(\mathbf{W}_{n_s}\right)=1,\ \forall n_s\in\mathcal{N}_s \tag{4.83}$$

$$\text{rank}\left(\mathbf{P}_c\right)=1,\ \text{rank}\left(\mathbf{P}_{k_t}\right)=1,\quad \forall k_t\in\mathcal{K}_t \tag{4.84}$$

$$(4.29),(4.30),(4.32),(4.33),(4.36)-(4.39)$$

where $\mathbf{W}_c=\mathbf{w}_c\mathbf{w}_c^H$, $\left\{\mathbf{W}_{n_s}=\mathbf{w}_{n_s}\mathbf{w}_{n_s}^H\right\}_{n_s=1}^{N_s}$, $\mathbf{P}_c=\mathbf{p}_c\mathbf{p}_c^H$, $\left\{\mathbf{P}_{k_t}=\mathbf{p}_{k_t}\mathbf{p}_{k_t}^H\right\}_{k_t=1}^{K_t}$. (4.79) and (4.80) are transmit power constraints. All the rate expressions in this section are redefined by the Ergodic form $R\triangleq\mathbb{E}\left\{\log_2\left(1+\text{SINR}\right)\right\}$, as the metric of average robust design. By taking (4.39) as an example, $R_{k_s}^{sat}$ can be approximated by

$$\begin{aligned}R_{k_s}^{sat}&=\mathbb{E}\left\{\log_2\left(1+\gamma_{k_s}^{sat}\right)\right\}\\&\approx\log_2\left(\frac{\mathbb{E}\left\{\text{tr}\left(\mathbf{F}_{k_s}\mathbf{W}_{\mu(k_s)}\right)\right\}+\sum_{i=1,i\neq\mu(k_s)}^{N_s}\mathbb{E}\left\{\text{tr}\left(\mathbf{F}_{k_s}\mathbf{W}_i\right)\right\}+\sigma_{k_s}^{sat2}}{\sum_{i=1,i\neq\mu(k_s)}^{N_s}\mathbb{E}\left\{\text{tr}\left(\mathbf{F}_{k_s}\mathbf{W}_i\right)\right\}+\sigma_{k_s}^{sat2}}\right)\\&=\log_2\left(\frac{\text{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_{\mu(k_s)}\right)+\sum_{i=1,i\neq\mu(k_s)}^{N_s}\text{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right)+\sigma_{k_s}^{sat2}}{\sum_{i=1,i\neq\mu(k_s)}^{N_s}\text{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right)+\sigma_{k_s}^{sat2}}\right).\end{aligned} \tag{4.85}$$

Note that (4.85) is very tight and has been verified to be theoretically accurate in [103]. Specifically, $\mathbf{F}_{k_s}=\text{diag}\left(\widehat{\mathbf{f}}_{k_s}\right)\mathbf{x}_{k_s}\mathbf{x}_{k_s}^H\text{diag}\left(\widehat{\mathbf{f}}_{k_s}^H\right)$. $\overline{\mathbf{F}}_{k_s}=\mathbb{E}\left\{\mathbf{F}_{k_s}\right\}=\text{diag}\left(\widehat{\mathbf{f}}_{k_s}\right)\mathbf{X}_{k_s}\text{diag}\left(\widehat{\mathbf{f}}_{k_s}^H\right)$ is defined as the channel correlation matrix, which captures the expectation over the distribution of phase uncertainty. Based on the approximated rate expressions, $\mathcal{D}_1$ can be rewritten as $\mathcal{F}_1$.

$$\mathcal{F}_1:\quad\max_{W,P,\mathbf{c}^{sat},\mathbf{c}^{bs},q,\mathbf{r},\boldsymbol{\alpha},\eta,\xi}\ q \tag{4.86}$$

$$s.t.\quad\eta_{k_t}^{bs}-\xi_{k_t}^{bs}\geq\alpha_{k_t}\log 2,\ \forall k_t\in\mathcal{K}_t \tag{4.87}$$

$$\begin{aligned}&e^{\eta_{k_t}^{bs}}\leq\text{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_{k_t}\right)+\sum_{j=1,j\neq k_t}^{K_t}\text{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right)+\\&\text{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right)+\sum_{i=1}^{N_s}\text{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right)+\sigma_{k_t}^{bs2},\ \forall k_t\in\mathcal{K}_t\end{aligned} \tag{4.88}$$

$$e^{\xi_{k_t}^{bs}} \geq \sum_{j=1, j\neq k_t}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2}, \ \forall k_t \in \mathcal{K}_t \quad (4.89)$$

$$\eta_{k_s}^{sat} - \xi_{k_s}^{sat} \geq r_{k_s} \log 2, \ \forall k_s \in \mathcal{K}_s \quad (4.90)$$

$$e^{\eta_{k_s}^{sat}} \leq \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_{\mu(k_s)}\right) + \sum_{i=1, i\neq\mu(k_s)}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \ \forall k_s \in \mathcal{K}_s \quad (4.91)$$

$$e^{\xi_{k_s}^{sat}} \geq \sum_{i=1, i\neq\mu(k_s)}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \ \forall k_s \in \mathcal{K}_s \quad (4.92)$$

$$\eta_{c,k_t}^{bs} - \xi_{c,k_t}^{bs} \geq \sum_{j=1}^{K_t} C_j^{bs} \log 2, \ \forall k_t \in \mathcal{K}_t \quad (4.93)$$

$$e^{\eta_{c,k_t}^{bs}} \leq \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_c\right) + \sum_{j=1}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) +$$

$$\mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2}, \ \forall k_t \in \mathcal{K}_t \quad (4.94)$$

$$e^{\xi_{c,k_t}^{bs}} \geq \sum_{j=1}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2}, \ \forall k_t \in \mathcal{K}_t \quad (4.95)$$

$$\eta_{c,k_s}^{sat} - \xi_{c,k_s}^{sat} \geq \sum_{j=1}^{N_s} C_j^{sat} \log 2, \ \forall k_s \in \mathcal{K}_s \quad (4.96)$$

$$e^{\eta_{c,k_s}^{sat}} \leq \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \ \forall k_s \in \mathcal{K}_s \quad (4.97)$$

$$e^{\xi_{c,k_s}^{sat}} \geq \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \ \forall k_s \in \mathcal{K}_s \quad (4.98)$$

(4.30), (4.33), (4.36), (4.38), (4.79) − (4.84)

where η and ξ are the sets of introduced slack variables. The constraints (4.87)-(4.89), (4.90)-(4.92), (4.93)-(4.95), and (4.96)-(4.98) are respectively the expansions of the rate constraints (4.37), (4.39), (4.29) and (4.32). Note that (4.89), (4.92), (4.95) and (4.98) are non-convex with convex left-hand sides, which can be approximated by the first-order Taylor approximation. Hence, we obtain these approximated linear

constraints

$$\sum_{j=1,j\neq k_t}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2} \leq e^{\xi_{k_t}^{bs[n]}}\left(\xi_{k_t}^{bs} - \xi_{k_t}^{bs[n]} + 1\right), \tag{4.99}$$

$$\sum_{i=1,i\neq\mu(k_s)}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2} \leq e^{\xi_{k_s}^{sat[n]}}\left(\xi_{k_s}^{sat} - \xi_{k_s}^{sat[n]} + 1\right), \tag{4.100}$$

$$\sum_{j=1}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2} \leq e^{\xi_{c,k_t}^{bs[n]}}\left(\xi_{c,k_t}^{bs} - \xi_{c,k_t}^{bs[n]} + 1\right), \tag{4.101}$$

$$\sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2} \leq e^{\xi_{c,k_s}^{sat[n]}}\left(\xi_{c,k_s}^{sat} - \xi_{c,k_s}^{sat[n]} + 1\right). \tag{4.102}$$

where n represents the n-th SCA iteration. The constraints (4.89), (4.92), (4.95) and (4.98) belong to generalized nonlinear convex program with high computational complexity. Following the same method introduced in the previous Section, they can be represented in linear and SOC forms given by

$$t_{k_t}^{bs} \leq \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_{k_t}\right) + \sum_{j=1,j\neq k_t}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2}, \tag{4.103}$$

$$\left\| t_{k_t}^{bs} + \eta_{k_t}^{bs} - \left(\log(t_{k_t}^{bs[n]}) + 1\right) \quad 2\sqrt{t_{k_t}^{bs[n]}} \right\|_2 \leq t_{k_t}^{bs} - \eta_{k_t}^{bs} + \left(\log(t_{k_t}^{bs[n]}) + 1\right), \tag{4.104}$$

$$t_{k_s}^{sat} \leq \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_{\mu(k_s)}\right) + \sum_{i=1,i\neq\mu(k_s)}^{N_s} \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \tag{4.105}$$

$$\left\| t_{k_s}^{sat} + \eta_{k_s}^{sat} - \left(\log(t_{k_s}^{sat[n]}) + 1\right) \quad 2\sqrt{t_{k_s}^{sat[n]}} \right\|_2 \leq t_{k_s}^{sat} - \eta_{k_s}^{sat} + \left(\log(t_{k_s}^{sat[n]}) + 1\right), \tag{4.106}$$

$$t_{c,k_t}^{bs} \leq \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_c\right) + \sum_{j=1}^{K_t} \mathrm{tr}\left(\mathbf{H}_{k_t}\mathbf{P}_j\right) + \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_c\right) + \sum_{i=1}^{N_s} \mathrm{tr}\left(\overline{\mathbf{Z}}_{k_t}\mathbf{W}_i\right) + \sigma_{k_t}^{bs2}, \tag{4.107}$$

$$\left\| t_{c,k_t}^{bs} + \eta_{c,k_t}^{bs} - \left(\log(t_{c,k_t}^{bs[n]}) + 1\right) \quad 2\sqrt{t_{c,k_t}^{bs[n]}} \right\|_2 \leq t_{c,k_t}^{bs} - \eta_{c,k_t}^{bs} + \left(\log(t_{c,k_t}^{bs[n]}) + 1\right), \tag{4.108}$$

$$t_{c,k_s}^{sat} \leq \mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_c\right) + \sum_{i=1}^{N_s}\mathrm{tr}\left(\overline{\mathbf{F}}_{k_s}\mathbf{W}_i\right) + \sigma_{k_s}^{sat2}, \tag{4.109}$$

$$\left\| t_{c,k_s}^{sat} + \eta_{c,k_s}^{sat} - \left(\log(t_{c,k_s}^{sat[n]}) + 1\right) \quad 2\sqrt{t_{c,k_s}^{sat[n]}} \right\|_2 \leq t_{c,k_s}^{sat} - \eta_{c,k_s}^{sat} + \left(\log(t_{c,k_s}^{sat[n]}) + 1\right). \tag{4.110}$$

Since rank-one implies only one nonzero eigenvalue, the non-convex constraints (4.83) and (4.84) can be rewritten by

$$\mathrm{tr}\left(\mathbf{W}_c\right) - \lambda_{\max}\left(\mathbf{W}_c\right) = 0,\ \mathrm{tr}\left(\mathbf{W}_{n_s}\right) - \lambda_{\max}\left(\mathbf{W}_{n_s}\right) = 0,\ \forall n_s \in \mathcal{N}_s, \tag{4.111}$$

$$\mathrm{tr}\left(\mathbf{P}_c\right) - \lambda_{\max}\left(\mathbf{P}_c\right) = 0,\ \mathrm{tr}\left(\mathbf{P}_{k_t}\right) - \lambda_{\max}\left(\mathbf{P}_{k_t}\right) = 0,\ \forall k_t \in \mathcal{K}_t, \tag{4.112}$$

where $\lambda_{\max}\left(\mathbf{X}\right)$ denotes the maximum eigenvalue of $\mathbf{X} \succeq 0$. Then, we build a penalty function to insert these constraints into the objective function (4.86) and obtain

$$\begin{aligned}\max_{W,P,\mathbf{c}^{sat},\mathbf{c}^{bs},q,\mathbf{r},\alpha,\eta,\xi} q - \beta\Big(&\left[\mathrm{tr}\left(\mathbf{W}_c\right) - \lambda_{\max}\left(\mathbf{W}_c\right)\right] + \sum_{n_s=1}^{N_s}\left[\mathrm{tr}\left(\mathbf{W}_{n_s}\right) - \lambda_{\max}\left(\mathbf{W}_{n_s}\right)\right] \\ &+ \left[\mathrm{tr}\left(\mathbf{P}_c\right) - \lambda_{\max}\left(\mathbf{P}_c\right)\right] + \sum_{k_t=1}^{K_t}\left[\mathrm{tr}\left(\mathbf{P}_{k_t}\right) - \lambda_{\max}\left(\mathbf{P}_{k_t}\right)\right]\Big).\end{aligned} \tag{4.113}$$

β is a proper penalty factor to guarantee the penalty function as small as possible. (4.113) is nonconcave due to the existence of the penalty function. To tackle this issue, we adopt an iterative method [63]. By taking $\mathrm{tr}\left(\mathbf{W}_c\right) - \lambda_{\max}\left(\mathbf{W}_c\right)$ as an example, we have the following inequality

$$\mathrm{tr}(\mathbf{W}_c) - (\mathbf{v}_{c,\max}^{[n]})^H\mathbf{W}_c\mathbf{v}_{c,\max}^{[n]} \geq \mathrm{tr}(\mathbf{W}_c) - \lambda_{\max}(\mathbf{W}_c) \geq 0, \tag{4.114}$$

where $\mathbf{v}_{c,\max}$ is the normalized eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}\left(\mathbf{W}_c\right)$. Furthermore, we define $\mathbf{v}_{n_s,\max}$ as the corresponding eigenvector of $\lambda_{\max}\left(\mathbf{W}_{n_s}\right)$, and so does $\mathbf{b}_{c,\max}$ for $\lambda_{\max}\left(\mathbf{P}_c\right)$ and $\mathbf{b}_{k_t,\max}$ for $\lambda_{\max}\left(\mathbf{P}_{k_t}\right)$. Let PF

denote the iterative penalty function

$$\mathrm{PF} = \beta\Big(\left[\mathrm{tr}\left(\mathbf{W}_c\right) - \left(\mathbf{v}_{c,\max}^{[n]}\right)^H \mathbf{W}_c \mathbf{v}_{c,\max}^{[n]}\right] + \sum_{n_s=1}^{N_s}\left[\mathrm{tr}\left(\mathbf{W}_{n_s}\right) - \left(\mathbf{v}_{n_s,\max}^{[n]}\right)^H \mathbf{W}_{n_s} \mathbf{v}_{n_s,\max}^{[n]}\right]$$
$$+ \left[\mathrm{tr}\left(\mathbf{P}_c\right) - \left(\mathbf{b}_{c,\max}^{[n]}\right)^H \mathbf{P}_c \mathbf{b}_{c,\max}^{[n]}\right] + \sum_{k_t=1}^{K_t}\left[\mathrm{tr}\left(\mathbf{P}_{k_t}\right) - \left(\mathbf{b}_{k_t,\max}^{[n]}\right)^H \mathbf{P}_c \mathbf{b}_{k_t,\max}^{[n]}\right]\Big). \tag{4.115}$$

Eventually, the approximate problem at iteration-n is given by

$$\mathcal{G}_1: \max_{W,P,\mathbf{c}^{sat},\mathbf{c}^{bs},q,\mathbf{r},\alpha,\eta,\xi,t} q - \mathrm{PF} \tag{4.116}$$
$$s.t.\ (4.30), (4.33), (4.36), (4.38), (4.79) - (4.82), (4.87), (4.90), (4.93), (4.96)$$
$$(4.99), (4.101), (4.103), (4.104), (4.107), (4.108),\ \forall k_t \in \mathcal{K}_t$$
$$(4.100), (4.102), (4.105), (4.106), (4.109), (4.110),\ \forall k_s \in \mathcal{K}_s$$

The problem is convex involving only linear matrix inequality (LMI) and SOC constraints, and can be effectively solved by CVX. In each iteration, the problem defined around the solution of the previous iteration is solved. We summarize the procedure of this robust joint beamforming scheme in Algorithm 3. Finally, eigenvalue decomposition can be used to obtain the optimized beamforming vectors. The optimal solution $(W^{[n]}, P^{[n]}, \eta^{[n]}, \xi^{[n]}, t^{[n]})$ of the n-th iteration is a feasible solution of the $(n+1)$-th iteration. Thus, this algorithm generates a non-decreasing sequence of objective values, which are bounded above by the transmit power constraints. Moreover, the objective function is guaranteed to converge by the existence of lower bounds, i.e., (4.114). In other words, the rank-one constraints can be satisfied [63]. The obtained solution satisfies the KKT optimality conditions of $\mathcal{G}_1$, which are indeed identical to those of $\mathcal{D}_1$ at convergence [102]. However, the global optimality of the achieved solution can not be guaranteed. The MMF optimization problem of RSMA-based *cooperative STIN* considering satellite phase

uncertainty remains the same as $\mathcal{P}_2$. Here, we still omit the detailed optimization framework. The process keeps the same as that for the *coordinated STIN*.

Algorithm 3 Robust Joint Beamforming Scheme

Initialize: $n \leftarrow 0, W^{[n]}, P^{[n]}, \xi^{[n]}, t^{[n]}$;
repeat
Solve the problem $\mathcal{G}_1$ at $(W^{[n]}, P^{[n]}, \xi^{[n]}, t^{[n]})$ to get the optimal solution $(\breve{W}, \breve{P}, \breve{\xi}, \breve{t}, \breve{\text{objective}})$;
$n \leftarrow n + 1$;
Update $W^{[n]} \leftarrow \breve{W}, P^{[n]} \leftarrow \breve{P}, \xi^{[n]} \leftarrow \breve{\xi}, t^{[n]} \leftarrow \breve{t}, \text{objective}^{[n]} \leftarrow \breve{\text{objective}}$;
until $\left|\text{objective}^{[n]} - \text{objective}^{[n-1]}\right| < \varepsilon$;

Remark 4.2: Recall that the problem formulations in Algorithm 2 and Algorithm 3 involve only SOC and LMI constraints. They both can be efficiently solved by the standard interior-point method. It suggests that the worst-case runtime can be used to compare the computational complexities of different problems [104]. Hence, the worst-case computational complexity of the proposed joint beamforming scheme in Algorithm 2 and the robust joint beamforming scheme in Algorithm 3 are respectively $\mathcal{O}\left([N_s^2 + N_t K_t]^{3.5} \log\left(\varepsilon^{-1}\right)\right)$ *and* $\mathcal{O}\left([N_s^3 + N_t^2 K_t]^{3.5} \log\left(\varepsilon^{-1}\right)\right)$ *[98, 105], where* ε *is the convergence tolerance. Similarly, the complexity of the cooperative STIN scenarios of Algorithm 2 and Algorithm 3 are respectively* $\mathcal{O}\Big([N_s\left(N_s + K_t\right) + N_t\left(N_s + K_t\right)]^{3.5} \log\left(\varepsilon^{-1}\right)\Big)$ *and* $\mathcal{O}\Big([N_s^2\left(N_s + K_t\right) + N_t^2\left(N_s + K_t\right)]^{3.5} \log\left(\varepsilon^{-1}\right)\Big)$*, which are higher than the coordinated STIN scenarios because of the larger number of variables in precoder design.*

4.5 Simulation Results and Analysis

In this section, simulation results are provided to evaluate the performance of the proposed joint beamforming algorithms. Both perfect CSIT and imperfect CSIT with satellite channel phase uncertainties are considered. The tolerance of accuracy is set to be $\varepsilon = 10^{-4}$. Channel models have been introduced in Section 4.2.1, and

Table 4.1: Simulation parameters [Chapter 4]

Parameter	Value
Frequency band (carrier frequency)	Ka (28 GHz)
Satellite height	35786 km (GEO)
Bandwidth	500 MHz
3 dB angle	0.4°
Maximum beam gain	52 dBi
User terminal antenna gain	42.7 dBi
Rain fading parameters	$(\mu, \sigma) = (-3.125, 1.591)$
UPA inter-element spacing	$d_1 = d_2 = \frac{\lambda}{2}$
Number of NLoS paths	3

the simulation parameters are listed in Table 4.1 [68, 106]. The satellite is equipped with N_s antennas. ρ multicasting SUs locate uniformly in each beam coverage area. According to the architecture of single feed per beam, which is popular in modern satellites such as Eutelsat Ka-Sat, the number of SUs is $K_s = \rho N_s$. Meanwhile, the BS is deployed with UPA with N_t antennas. We assume K_t CUs are uniformly distributed within the BS coverage. In the satellite channel model, since we normalize the noise power by $\kappa T_{sys} B_w$, we can claim $\sigma_{k_s}^{sat2} = \sigma_{k_t}^{bs2} = 1,\ \forall k_s \in \mathcal{K}_s,\ \forall k_t \in \mathcal{K}_t$ in the simulations. The transmit SNRs [4] can be read from the transmit power P_s and P_t. All MMF rate curves throughout the simulations are calculated by averaging 100 channel realizations.

At first, we assume that perfect CSI is available at the GW. Fig. 4.3 compares the MMF rate performance of RSMA-based *coordinated* and *cooperative scheme*. The label "coordinated rsma" means RSMA is adopted at both the satellite and BS, while "cooperative rsma" means the satellite and BS work cooperatively as a super transmitter while RSMA is adopted. As P_t grows, we can see that the MMF rates of both schemes increase and tend to saturate at large P_t region. The *cooperative scheme* outperforms the *coordinated scheme* apparently at low P_t region. The gap between the two schemes decreases gradually as P_t grows and finally converges to

[4] According to the parameters given in Table 4.1 and the satellite channel model, the long-term received SNR is calculated to be around 0.67 times the transmit SNR.

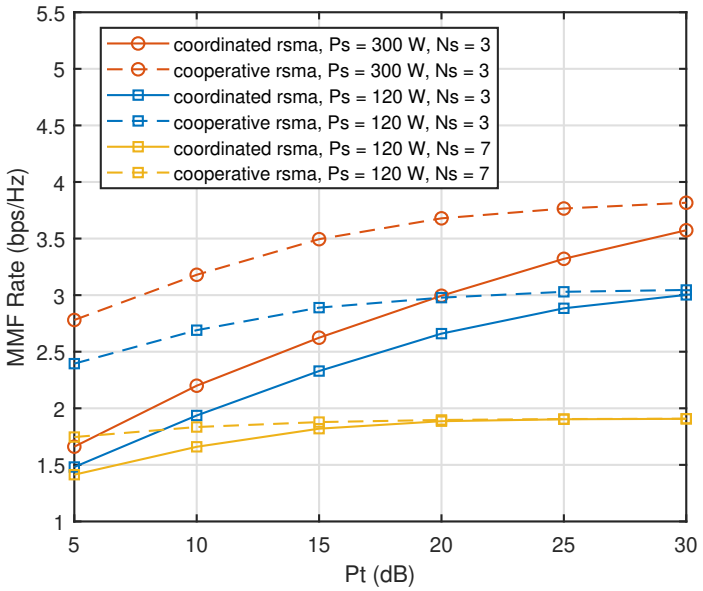

Figure 4.3: MMF rate versus P_t with different P_s and N_s. $N_t = 16$, $K_t = 4$, $K_s = \rho N_s$, $\rho = 2$.

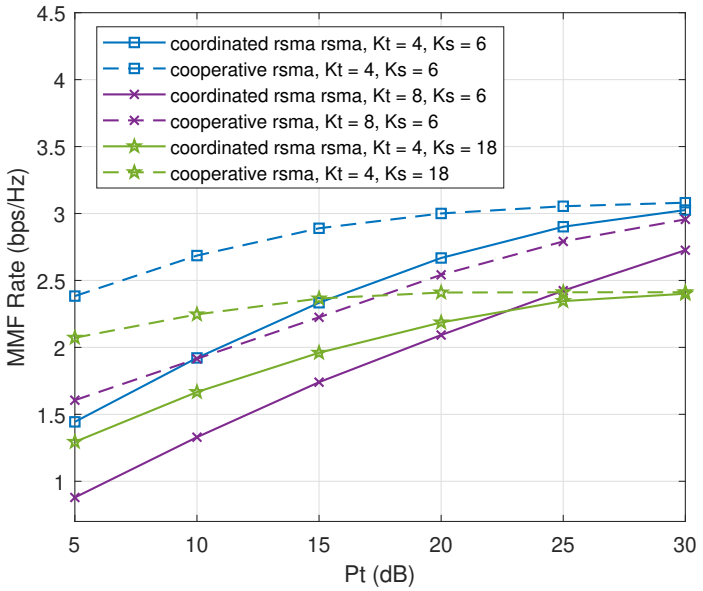

Figure 4.4: MMF rate versus P_t with different K_s and K_t. $N_t = 16$, $N_s = 3$, , $P_s = 120$W.

the same value when P_t is sufficiently large. The reasons are as follows. When P_t is relatively small, the STIN's performance is restricted in the *coordinated scheme* because the SINRs of CUs are much lower than the SINRs of SUs. Joint beamforming is designed to achieve optimal MMF rates. However, in the *cooperative scheme*, data exchange is assumed and the satellite can complement the services of BS to serve CUs, thereby remaining the optimized MMF rate at a higher level than that in the *coordinated scheme*. As P_t grows, the benefits of the *cooperative scheme* compared with the *coordinated scheme* decreases. When P_t is sufficiently large, the MMF rates of both schemes will finally converge to the same value due to the fixed satellite transmit power budget P_s. We also investigate the influence of different P_s and N_s setups. Apparently, the larger P_s is, the better MMF rate performance can be achieved. When N_s is increased from 3 to 7, by keeping $\rho = 2$, there will be $K_s = 14$ SUs. We can see that larger N_s leads to lower saturation MMF rates at high P_t region. The larger N_s is, the less transmit power is allocated to each satellite beam. Moreover, each SU will see more inter-beam interference due to the existence of more beams, thus resulting in performance degradation.

Fig. 4.4 depicts the MMF rates versus P_t with different number of SUs and CUs. When K_t is increased from 4 to 8, the performance will become worse in both *coordinated* and *cooperative scheme* especially at low P_t region, where the CUs take a dominant position of the system's MMF rate. On the other hand, when increasing the number of users per beam ρ from 2 to 6, i.e., from $K_s = 6$ to $K_s = 18$, we can still see the performance degradation in both *coordinated* and *cooperative scheme*. The performance degrades much at high P_t region, where the MMF rate is dominated by the satellite sub-system.

In Fig. 4.5, to investigate the influence of different transmission strategies in STIN, we compare the proposed RSMA-assisted beamforming with baseline strategies including SDMA, NOMA, a two-step beamforming, and fractional frequency reuse.

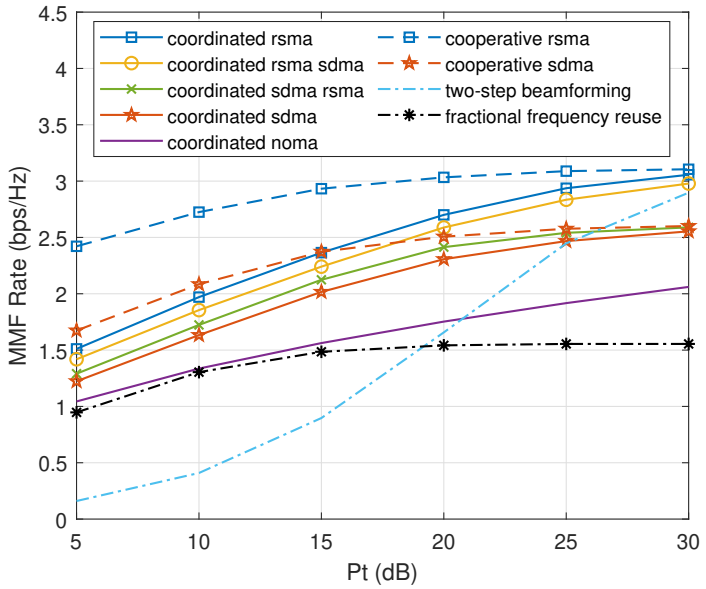

Figure 4.5: MMF rate versus P_t with different transmission strategies. $N_t = 16$, $K_t = 4$, $N_s = 3$, $K_s = 6$, $P_s = 120$W.

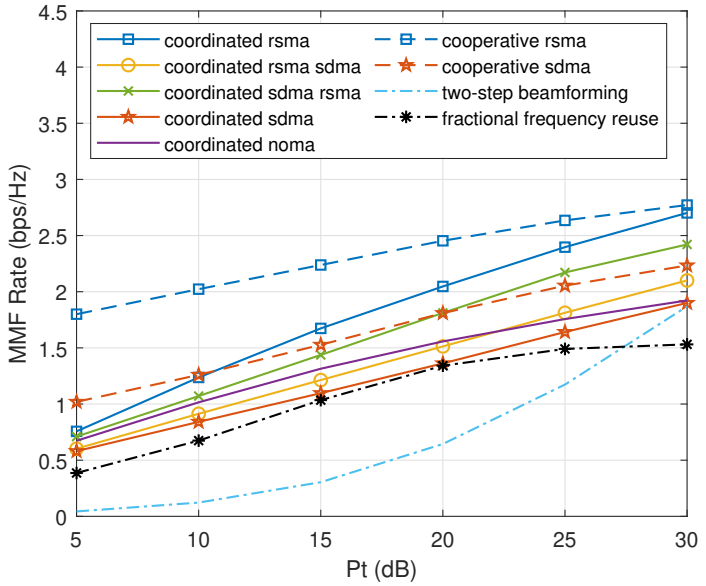

Figure 4.6: MMF rate versus P_t for different transmission strategies. $N_t = 4$, $K_t = 4$, $N_s = 3$, $K_s = 6$, $P_s = 120$W.

According to the analysis above, here we basically assume $N_s = 3$, $\rho = 2$, $K_s = 6$, $N_t = 16$, $K_t = 4$ for lower computational complexity. Different combinations of transmission strategies are considered in the *coordinated scheme*, e.g., the label "coordinated rsma sdma" means RSMA is used at the satellite while SDMA is used at the BS. It has been shown in [68] that adopting RSMA compared with SDMA in an overloaded system can provide more gains than in an underloaded system. Therefore, in this STIN where the satellite sub-system is always overloaded, and $N_t = 16$ is large enough to support an underloaded cellular sub-system, the performance improvement obtained by using RSMA compared with using SDMA at the satellite is more obvious than at the BS. As a consequence, the "coordinated rsma" successively outperforms "coordinated rsma sdma", "coordinated sdma rsma" and "coordinated sdma". For the "coordinated noma noma", SC-SIC is implemented at both the satellite and BS. The decoding order of NOMA at the satellite is decided by the ascending order of the weakest user's channel strength in each beam. We can observe that the MMF rate achieved by NOMA is the worst compared with RSMA and SDMA. The low performance of NOMA in multi-antenna settings is inline with the observations in [8] and the references therein. As discussed in Fig. 4.3, *cooperative schemes* can provide higher MMF rates than the corresponding *coordinated schemes*. Thus, the "cooperative rsma" outperforms "coordinated rsma" in Fig. 4.5, and finally, they tend to reach the same MMF rate restricted by the fixed P_s at very large P_t region. Similarly, the "cooperative sdma" outperforms "coordinated sdma". As P_t increases, they converge to the same value which is lower than that of RSMA. For the two-step beamforming, both CSI and data are not exchanged between the satellite and BS. The beamforming for the satellite is at first optimized. Then, the beamforming for the BS is optimized. Since the satellite beamfoming vectors are not jointly designed with the BS beamforming vectors, CUs will see serious interference from the satellite. As P_t grows, the value of minimum rate tends to reach the saturation MMF rate of RSMA-based *coordinated* and *cooperative schemes*. For the scheme

of fractional frequency reuse, the satellite and BS operate on different frequency bands. The spectrum cannot be effectively used, therefore resulting in poor MMF rate performance. In [68], it has been demonstrated that the conventional four-color frequency reuse of multibeam satellite systems performs the worst compared with full frequency reuse strategies. Thus, we do not compare with the four-color frequency reuse in this work.

In Fig. 4.6, the number of BS antennas is reduced to $N_t = 2 \times 2 = 4$, which is not enough to support effective beamforming at the BS so as to eliminate the intra-cell interference and the satellite interference. Compared with Fig. 4.5 with $N_t = 4 \times 4 = 16$ antennas, the MMF rates of all strategies are suppressed. Specifically, the performance of "coordinated sdma rsma" becomes better than the "coordinated rsma sdma". It implies that when N_t is not sufficient to suppress the intra-cell interference, the gains obtained by using RSMA compared with using SDMA at the BS can become more obvious than at the satellite. We can conclude that the larger N_t is, the better MMF rate performance can be achieved. In other words, as N_t increases, less P_t is required to reach the same MMF rate performance.

Furthermore, we assume imperfect CSI at the GW considering satellite phase uncertainties. Fig. 4.7 shows the MMF rate performance of the proposed robust joint beamforming in both RSMA-based *coordinated STIN* and *cooperative STIN*. As the variance of phase uncertainty δ^2 increases, the MMF rates of both schemes decrease gradually. From perfect CSIT to imperfect CSIT when $\delta^2 = 5°$, $\delta^2 = 15°$, and the phase-blind scenario, the corresponding MMF rates decrease gradually. The *cooperative STIN* still outperforms *coordinated STIN*. For comparison, we consider the conventional SDMA which performs well amongst the other baseline strategies.

From Fig. 4.8, we can observe that the gaps between perfect CSIT curves and imperfect CSIT curves become larger compared with the RSMA results in Fig. 4.7. RSMA is more robust to the channel phase uncertainty than SDMA due to its

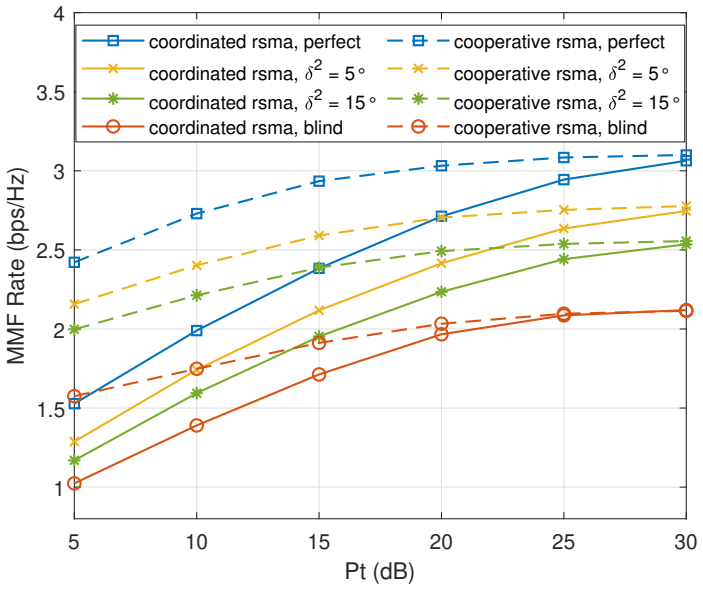

Figure 4.7: MMF rate versus P_t with different satellite phase uncertainties. RSMA is adopted at the transmitters. $N_t = 16$, $K_t = 4$, $N_s = 3$, $K_s = 6$, $P_s = 120$W.

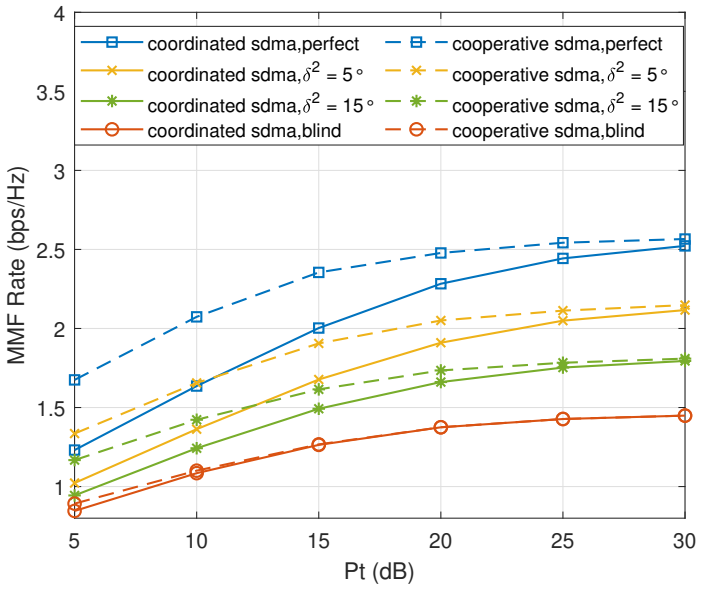

Figure 4.8: MMF rate versus P_t with different satellite phase uncertainties. SDMA is adopted at the transmitters. $N_t = 16$, $K_t = 4$, $N_s = 3$, $K_s = 6$, $P_s = 120$W.

more flexible architecture to partially decode the interference and partially treat the interference as noise.

4.6 Summary

In this chapter, we investigate the application of RSMA to STIN considering either perfect CSI or imperfect CSI with satellite channel phase uncertainties at the GW. Two RSMA-based STIN schemes are presented, namely the *coordinated scheme* relying on CSI exchange and the *cooperative scheme* relying on both CSI and data exchange at the GW. MMF optimization problems are formulated while satisfying transmit power budgets. To tackle the optimization, two iterative algorithms are respectively proposed. Through simulation results, the superiority of the proposed RSMA-based schemes for STIN is demonstrated compared with various baseline strategies. The robustness of RSMA is verified. In conclusion, RSMA is shown very promising for STIN to manage the interference in and between the satellite and terrestrial sub-systems.

Chapter 5

RSMA for Integrated Sensing and Communication Systems

This chapter introduces a general RSMA-assisted ISAC architecture, where the ISAC platform has a dual capability to simultaneously communicate with downlink users and probe detection signals to a moving target. To design an appropriate ISAC waveform, we investigate the RSMA-assisted ISAC beamfoming which jointly minimizes the CRB of target estimation and maximizes the minimum fairness rate (MFR) amongst communication users subject to the per-antenna power constraint. The superiority of RSMA-assisted ISAC is verified through simulation results in both terrestrial and satellite scenarios. RSMA is demonstrated to be a powerful multiple access and interference management strategy for ISAC, and provides a better communication-sensing trade-off compared with the conventional baseline strategies.

5.1 Introduction

As the growing number of communication equipments and various types of radars are placed on satellites, using ISAC waveform design to support simultaneous satellite communications and sensing becomes very necessary to explore. As introduced in the previous chapters, RSMA is a flexible and robust interference management strategy for multi-antenna systems, which relies on linearly precoded rate-splitting at the transmitter and SIC at the receivers, and has been proven to be promising for multibeam satellite systems in Chapter 3 and STIN in Chapter 4.

In this chapter, we present an overview of the interplay between RSMA and ISAC. RSMA-assisted ISAC which facilitates the integration of communications and moving target sensing is investigated to make better use of the RF spectrum and infrastructure. Rather than using the MSE of transmit beampattern approximation as the radar metric, explicit optimization of estimation performance at the radar receiver is studied. RSMA-assisted ISAC waveform optimization is for the first time studied to jointly minimize the CRB of the target estimation and maximize the MFR amongst all communication users subject to transmit power constraints. To solve the formulated non-convex problem efficiently, we propose an iterative algorithm based on SCA to solve the optimization. Simulation results show that RSMA is very effective for both terrestrial and satellite ISAC systems to manage the multiuser/inter-beam interference as well as performing the radar functionality.

5.2 System Model

We consider a general RSMA-assisted ISAC, where the antenna array is shared by a co-located monostatic MIMO radar system and a multiuser communication system as depicted in Fig. 5.1. The ISAC platform equipped with N_t transmit antennas and N_r

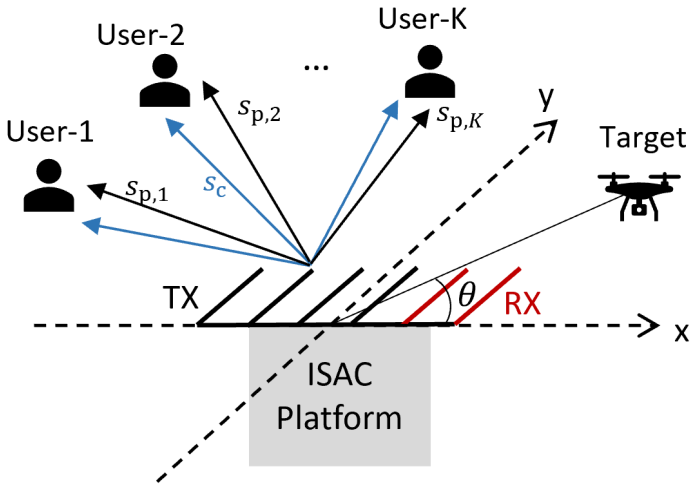

Figure 5.1: Model of an RSMA-assisted ISAC system.

receive antennas simultaneously senses a moving target and serves K downlink single-antenna users indexed by the set $\mathcal{K} = \{1, \cdots, K\}$. Since RSMA[1] is adopted, the messages $W_1, \cdots, W_K$ intended for the communication users are split into common parts and private parts. All common part messages $\{W_{c,1}, \cdots, W_{c,K}\}$ are jointly encoded into a common stream s_c, while all private part messages $\{W_{p,1}, \cdots, W_{p,K}\}$ are respectively encoded into private streams $s_1, \cdots, s_K$. Thus, we can denote $\mathbf{s}[l] = [s_c[l], s_1[l], \cdots, s_K[l]]^T$ as a $(K+1) \times 1$ vector of unit-power signal streams, where $l \in \mathcal{L} = \{1, \cdots, L\}$ is the discrete-time index within one coherent processing interval (CPI), and the transmit signal at time index l writes as

$$\mathbf{x}[l] = \mathbf{P}\mathbf{s}[l] = \mathbf{p}_c s_c[l] + \sum_{k \in \mathcal{K}} \mathbf{p}_k s_k[l]. \tag{5.1}$$

where $\mathbf{P} = [\mathbf{p}_c, \mathbf{p}_1, \cdots, \mathbf{p}_K] \in \mathbb{C}^{N_t \times (K+1)}$ is the beamforming matrix, which is fixed within one CPI. If L is sufficiently large, and the data streams are assumed to be independent of each other, satisfying $\frac{1}{L}\sum_{l=1}^{L} \mathbf{s}[l]\mathbf{s}[l]^H = \mathbf{I}_K$, the covariance matrix

[1]One-layer RSMA is considered here for brevity and ease of illustration.

of the transmit signal can be written as

$$\mathbf{R}_X = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}\left[l\right]\mathbf{x}\left[l\right]^H = \mathbf{P}\mathbf{P}^H. \tag{5.2}$$

5.2.1 Sensing Model and Metric

The $N_r \times 1$ reflected echo signal at the radar receiver writes as

$$\begin{aligned}\mathbf{y}_{\mathrm{r}}\left[l\right] &= \mathbf{H}_{\mathrm{r}}\mathbf{x}\left[l\right] + \mathbf{m}\left[l\right] \\ &= \alpha e^{j2\pi\mathcal{F}_D lT}\mathbf{b}\left(\theta\right)\mathbf{a}^H\left(\theta\right)\mathbf{x}\left[l\right] + \mathbf{m}\left[l\right],\end{aligned} \tag{5.3}$$

where $\mathbf{H}_{\mathrm{r}} \in \mathbb{C}^{N_r \times N_t}$ is the effective radar sensing channel. α stands for the complex reflection coefficient which is related to the radar cross-section (RCS) of the target. $\mathcal{F}_D = \frac{2vf_c}{c}$ denotes the Doppler frequency, with f_c and c respectively representing the carrier frequency and the speed of the light. v is the relative radar target velocity. T denotes the symbol period.

Note that for a monostatic radar, the direction of arrival (DoA) and the direction of departure (DoD) are the same, and can be denoted by θ, which is the azimuth angle. $\mathbf{a}(\theta) \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{b}(\theta) \in \mathbb{C}^{N_r \times 1}$ are the transmit and receive steering vectors, respectively. $\mathbf{m}\left[l\right]$ is the AWGN distributed by $\mathbf{m}\left[l\right] \sim \mathcal{CN}(\mathbf{0}_{N_r}, \sigma_m^2\mathbf{I}_{N_r})$, with σ_m^2 denoting the variance of each entry.

The steering vectors $\mathbf{a}(\theta)$ and $\mathbf{b}(\theta)$ can be expressed as a function of the Cartesian coordinates of the transmit and receive arrays as follows

$$\mathbf{a}(\theta) = e^{\left(j\frac{2\pi}{\lambda}\left[\bar{\mathbf{r}}_1,\cdots,\bar{\mathbf{r}}_{N_t}\right]^T[\cos\theta,\sin\theta,0]^T\right)}, \tag{5.4}$$

$$\mathbf{b}(\theta) = e^{\left(-j\frac{2\pi}{\lambda}[\mathbf{r}_1,\cdots,\mathbf{r}_{N_r}]^T[\cos\theta,\sin\theta,0]^T\right)}. \tag{5.5}$$

The matrices $[\bar{\mathbf{r}}_1,\cdots,\bar{\mathbf{r}}_{N_t}] \in \mathcal{R}^{3\times N_t}$ and $[\mathbf{r}_1,\cdots,\mathbf{r}_{N_r}] \in \mathcal{R}^{3\times N_r}$ have columns rep-

resenting the Cartesian coordinates of the transmit and receive array elements, respectively.

It is well-known that the CRB serves as a theoretical lower-bound of the variance of unbiased estimators for parameter estimation [107].

In general, the CRB is inversely proportional to the square root of the product of the SNR times L, and is valid only (by definition) for high SNR. In this Chapter, we consider the CRB as the radar sensing performance metric for target estimation [72, 108]. The CRB matrix can be calculated as CRB $= \mathbf{F}^{-1}$, where $\mathbf{F}$ is the Fisher information matrix (FIM) for estimating the real-valued target parameters $\boldsymbol{\xi} = [\theta, \alpha^{\Re}, \alpha^{\Im}, \mathcal{F}_D]^T$ given by

$$\mathbf{F} = \begin{bmatrix} F_{\theta\theta} & F_{\theta\alpha^{\Re}} & F_{\theta\alpha^{\Im}} & F_{\theta\mathcal{F}_D} \\ F_{\theta\alpha^{\Re}}^T & F_{\alpha^{\Re}\alpha^{\Re}} & F_{\alpha^{\Re}\alpha^{\Im}} & F_{\alpha^{\Re}\mathcal{F}_D} \\ F_{\theta\alpha^{\Im}}^T & F_{\alpha^{\Re}\alpha^{\Im}}^T & F_{\alpha^{\Im}\alpha^{\Im}} & F_{\alpha^{\Im}\mathcal{F}_D} \\ F_{\theta\mathcal{F}_D}^T & F_{\alpha^{\Re}\mathcal{F}_D}^T & F_{\alpha^{\Im}\mathcal{F}_D}^T & F_{\mathcal{F}_D\mathcal{F}_D} \end{bmatrix}. \tag{5.6}$$

From [107], by denoting $\boldsymbol{\mu}[l] = \mathbf{y}_{\mathrm{r}}[l] - \mathbf{m}[l]$, the elements of FIM are expressed by

$$[\mathbf{F}]_{i,j} = \frac{2}{\sigma_m^2} \mathrm{Re}\Big\{ \sum_{l=1}^{L} \frac{\partial \boldsymbol{\mu}[l]^H}{\partial \xi_i} \frac{\partial \boldsymbol{\mu}[l]}{\partial \xi_j} \Big\}, \ i, j \in \{1, \cdots, 4\}, \tag{5.7}$$

where ξ_i, ξ_j are the elements of $\boldsymbol{\xi}$. By denoting $\mathbf{A} = \mathbf{b}(\theta)\mathbf{a}^H(\theta)$, the derivatives in (5.7) are expressed as follows

$$\frac{\partial \boldsymbol{\mu}[l]}{\partial \theta} = \alpha e^{j2\pi\mathcal{F}_D l T} \frac{\partial \mathbf{A}}{\partial \theta} \mathbf{x}[l], \tag{5.8}$$

$$\frac{\partial \boldsymbol{\mu}[l]}{\partial \alpha^{\Re}} = e^{j2\pi\mathcal{F}_D l T} \mathbf{A}\mathbf{x}[l], \tag{5.9}$$

$$\frac{\partial \boldsymbol{\mu}[l]}{\partial \alpha^{\Im}} = j e^{j2\pi\mathcal{F}_D l T} \mathbf{A}\mathbf{x}[l], \tag{5.10}$$

$$\frac{\partial \boldsymbol{\mu}[l]}{\partial \mathcal{F}_D} = \alpha (j2\pi l T) e^{j2\pi\mathcal{F}_D l T} \mathbf{A}\mathbf{x}[l]. \tag{5.11}$$

By substituting (5.8) - (5.11) into (5.7), the elements of the FIM are given by

$$F_{\theta,\theta} = \frac{2\left|\alpha\right|^2 L}{\sigma_m^2}\mathrm{Re}\Big\{\mathrm{tr}\big(\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{R}_X\frac{\partial \mathbf{A}}{\partial \theta}^H\big)\Big\}, \tag{5.12}$$

$$F_{\alpha^{\Re},\alpha^{\Re}} = F_{\alpha^{\Im},\alpha^{\Im}} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\mathbf{A}^H\big)\Big\}, \tag{5.13}$$

$$F_{\mathcal{F}_D,\mathcal{F}_D} = \frac{2\left|\alpha\right|^2 L}{\sigma_m^2}\mathrm{Re}\Big\{\big(\sum_{l=1}^{L}\left(2\pi lT\right)^2\big)\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\mathbf{A}^H\big)\Big\}, \tag{5.14}$$

$$F_{\alpha^{\Re},\alpha^{\Im}} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{j\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\mathbf{A}^H\big)\Big\} = 0, \tag{5.15}$$

$$F_{\theta,\alpha^{\Re}} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{\alpha^*\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\frac{\partial \mathbf{A}}{\partial \theta}^H\big)\Big\}, \tag{5.16}$$

$$F_{\theta,\alpha^{\Im}} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{\alpha^* j\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\frac{\partial \mathbf{A}}{\partial \theta}^H\big)\Big\}, \tag{5.17}$$

$$F_{\theta,\mathcal{F}_D} = \frac{2\left|\alpha\right|^2 L}{\sigma_m^2}\mathrm{Re}\Big\{j\big(\sum_{l=1}^{L}2\pi lT\big)\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\frac{\partial \mathbf{A}}{\partial \theta}^H\big)\Big\}, \tag{5.18}$$

$$F_{\alpha^{\Re},\mathcal{F}_D} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{\alpha j\big(\sum_{l=1}^{L}2\pi lT\big)\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\mathbf{A}^H\big)\Big\}, \tag{5.19}$$

$$F_{\alpha^{\Im},\mathcal{F}_D} = \frac{2L}{\sigma_m^2}\mathrm{Re}\Big\{\alpha\big(\sum_{l=1}^{L}2\pi lT\big)\mathrm{tr}\big(\mathbf{A}\mathbf{R}_X\mathbf{A}^H\big)\Big\}. \tag{5.20}$$

Note that $[\mathbf{F}]_{i,j}$ are all dependent of $\mathbf{R}_X$. As discussed in [109], $\mathbf{R}_X$ can be designed appropriately to improve the estimation capability of a MIMO radar by minimizing the trace, determinant or largest eigenvalue of the CRB matrix.

5.2.2 Communication Model and Metric

At each user side, the received signal is given by

$$\begin{aligned} y_k\left[l\right] &= \mathbf{h}_k^H\mathbf{x}\left[l\right] + n_k\left[l\right] \\ &= \mathbf{h}_k^H\mathbf{p}_c s_c\left[l\right] + \mathbf{h}_k^H\sum_{k\in\mathcal{K}}\mathbf{p}_k s_k\left[l\right] + n_k\left[l\right],\ \forall k\in\mathcal{K}. \end{aligned} \tag{5.21}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_\mathrm{t}\times 1}$ denotes the channel between the ISAC transmitter and user-k. $n_k[l] \sim \mathcal{CN}(0, \sigma_{n,k}^2)$ represents the AWGN with zero mean. We assume the noise variance $\sigma_{n,k}^2 = \sigma_n^2,\ \forall k \in \mathcal{K}$.

Following the decoding order of RSMA, each user first decodes the common stream by treating all private streams as noise. The SINR of decoding s_c at user-k is expressed by

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \mathbf{p}_c\right|^2}{\sum_{i\in\mathcal{K}} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_n^2},\ \forall k \in \mathcal{K}. \tag{5.22}$$

$R_{c,k} = \log_2(1+\gamma_{c,k})$ is the corresponding achievable rate when assuming Gaussian signalling. To guarantee that each user is capable of decoding the common stream, we define the common rate as $R_c = \min_{k\in\mathcal{K}}\{R_{c,k}\} = \sum_{k\in\mathcal{K}} C_k$, where C_k is the rate of the common part of the k-th user's message. After the common stream is re-encoded, precoded and subtracted from the received signal through SIC, each user then decodes its desired private stream. The SINR of decoding s_k at user-k is given by

$$\gamma_k = \frac{\left|\mathbf{h}_k^H \mathbf{p}_k\right|^2}{\sum_{i\in\mathcal{K}, i\neq k} \left|\mathbf{h}_k^H \mathbf{p}_i\right|^2 + \sigma_n^2},\ \forall k \in \mathcal{K}. \tag{5.23}$$

The achievable rate of the private stream is $R_k = \log_2(1+\gamma_k)$, and the total achievable rate of user-k, assuming Gaussian signalling, writes as $R_{k,\mathrm{tot}} = C_k + R_k, \forall k \in \mathcal{K}$.

To mitigate multiuser interference, the precoders can be designed to maximize the MFR, which is defined by

$$\mathrm{MFR}(\mathbf{P}) = \min_{k\in\mathcal{K}} (C_k + R_k). \tag{5.24}$$

For the baseline strategies, SDMA-assisted ISAC is enabled by turning off the

common stream in (5.1). NOMA-assisted ISAC relies on superposition coding at the transmitter and SIC at each user. The precoders and decoding orders are typically jointly optimized. By taking a two-user system as an example, and considering the specific decoding order, where the message of user-1 is decoded before the message of user-2, user-2 is able to decode messages of both users, while user-1 only decodes its desired message. Therefore, RSMA boils down to NOMA by encoding W_1 into the common stream s_{c}, encoding W_2 into s_2 and turning off s_1.

5.3 ISAC Beamforming Optimization

The RSMA-assisted ISAC beamforming matrix can be designed by investigating the trade-off between communication and radar performance. In this chapter, we employ the CRB as the radar performance metric, which represents a lower bound on the variance of unbiased estimators, and employ the MFR as the communication performance metric to ensure the quality of service.

The optimization problem is formulated to maximize the communication MFR while minimizing the largest eigenvalue of the CRB matrix, which is equivalent to maximizing the smallest eigenvalue of FIM. Assuming perfect CSIT, the optimization problem is written as

$$\max_{\mathbf{P},\mathbf{c},t_{\mathrm{FIM}}} \left[\min_{k\in\mathcal{K}} \left(C_k + R_k\right)\right] + \lambda t_{\mathrm{FIM}} \tag{5.25}$$

$$s.t. \quad \mathbf{F} \succeq t_{\mathrm{FIM}}\mathbf{I}_4 \tag{5.26}$$

$$\mathrm{diag}(\mathbf{P}\mathbf{P}^H) = \frac{P\mathbf{1}^{N_t\times 1}}{N_t} \tag{5.27}$$

$$R_{c,k} \geq \sum_{i=1}^{K} C_i, \ \forall k \in \mathcal{K} \tag{5.28}$$

$$C_k \geq 0, \ \forall k \in \mathcal{K}, \tag{5.29}$$

where $\mathbf{c} = [C_1, \cdots, C_K]^T$ is the vector of common rate portions. t_{FIM} is the variable representing the smallest eigenvalue of FIM according to (5.26). $\mathbf{I}$ is an identity matrix (which is of the same dimension as $\mathbf{F}$). λ is the regularization parameter to prioritize either communications or radar sensing. P denotes the sum transmit power budget. The constraint (5.27) ensures the transmit power of each antenna to be the same, which is commonly used for MIMO radar to avoid saturation of transmit power amplifiers in practical systems. The constraint (5.28) ensures that the common stream can be successfully decoded by all communication users, and (5.29) guarantees the non-negativity of all common rate portions.

By defining $\mathbf{P}_c = \mathbf{p}_c\mathbf{p}_c^H, \mathbf{P}_k = \mathbf{p}_k\mathbf{p}_k^H, \mathbf{H}_k = \mathbf{h}_k\mathbf{h}_k^H$, the original problem (5.25) - (5.29) can be equivalently transformed into SDP form with rank-one constraints, which is given by

$$\max_{\mathbf{P}_c, \{\mathbf{P}_k\}_{k=1}^K, \mathbf{c}, t_{\text{FIM}}, \mathbf{r}, q} \quad q + \lambda t_{\text{FIM}} \tag{5.30}$$

$$s.t. \quad \mathbf{F} \succeq t_{\text{FIM}}\mathbf{I}_4 \tag{5.31}$$

$$\text{diag}\left(\mathbf{P}_c + \sum_{k=1}^{K}\mathbf{P}_k\right) = \frac{P\mathbf{1}^{N_t\times 1}}{N_t} \tag{5.32}$$

$$\mathbf{P}_c \succeq 0, \ \mathbf{P}_k \succeq 0, \ \forall k \in \mathcal{K} \tag{5.33}$$

$$\text{rank}\left(\mathbf{P}_c\right) = 1, \ \text{rank}\left(\mathbf{P}_k\right) = 1, \ \forall k \in \mathcal{K} \tag{5.34}$$

$$\log_2\left(1 + \frac{\text{tr}\left(\mathbf{H}_k\mathbf{P}_c\right)}{\sum_{j\in\mathcal{K}}\text{tr}\left(\mathbf{H}_k\mathbf{P}_j\right) + \sigma_n^2}\right) \geq \sum_{i=1}^{K} C_i, \tag{5.35}$$

$$C_k \geq 0, \ \forall k \in \mathcal{K} \tag{5.36}$$

$$\log_2\left(1 + \frac{\text{tr}\left(\mathbf{H}_k\mathbf{P}_k\right)}{\sum_{j\in\mathcal{K}, j\neq k}\text{tr}\left(\mathbf{H}_k\mathbf{P}_j\right) + \sigma_n^2}\right) \geq r_k, \tag{5.37}$$

$$C_k + r_k \geq q, \ \forall k \in \mathcal{K} \tag{5.38}$$

where $\mathbf{r} = [r_1, \cdots, r_K]^T$, q are auxiliary variables. The covariance matrix of the transmit signal is expressed by $\mathbf{R}_X = \mathbf{P}\mathbf{P}^H = \mathbf{P}_c + \sum_{k=1}^{K}\mathbf{P}_k$.

With respect to the equivalent problem (5.30) - (5.38), we can observe that the rank-one constraints (5.34) and the rate constraints (5.35) and (5.37) are non-convex. To deal with the non-convexity of rate constraints (5.35) and (5.37), we first rewrite them by introducing slack variables $\{\eta_{c,k}\}_{k=1}^{K}, \{\beta_{c,k}\}_{k=1}^{K}, \{\eta_k\}_{k=1}^{K}, \{\beta_k\}_{k=1}^{K}$ as

$$\eta_{c,k} - \beta_{c,k} \geq \sum_{i=1}^{K} C_i \log 2, \ \forall k \in \mathcal{K}, \tag{5.39}$$

$$e^{\eta_{c,k}} \leq \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_c\right) + \sum_{j \in \mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}, \tag{5.40}$$

$$e^{\beta_{c,k}} \geq \sum_{j \in \mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}, \tag{5.41}$$

$$\eta_k - \beta_k \geq r_k \log 2, \ \forall k \in \mathcal{K}, \tag{5.42}$$

$$e^{\eta_k} \leq \sum_{j \in \mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}, \tag{5.43}$$

$$e^{\beta_k} \geq \sum_{j \in \mathcal{K}, j \neq k} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}. \tag{5.44}$$

Note that (5.41) and (5.44) are still non-convex with convex left-hand sides which can be approximated by the first-order Taylor approximation given as follows

$$\sum_{j \in \mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2 \leq e^{\beta_{c,k}^{[n]}} \left(\beta_{c,k} - \beta_{c,k}^{[n]} + 1\right), \ \forall k \in \mathcal{K}, \tag{5.45}$$

$$\sum_{j \in \mathcal{K}, j \neq k} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2 \leq e^{\beta_k^{[n]}} \left(\beta_k - \beta_k^{[n]} + 1\right), \ \forall k \in \mathcal{K}, \tag{5.46}$$

where n represents the n-th SCA iteration. (5.40) and (5.43) belong to generalized nonlinear convex program, which leads to high computational complexity. Aiming at more efficient implementation, we introduce $\{\tau_{c,k}\}_{k=1}^{K}, \{\tau_k\}_{k=1}^{K}$, and rewrite (5.40) and (5.43) as

$$\tau_{c,k} \leq \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_c\right) + \sum_{j \in \mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k \mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}, \tag{5.47}$$

$$\tau_{c,k} \log\left(\tau_{c,k}\right) \geq \tau_{c,k}\eta_{c,k}, \ \forall k \in \mathcal{K}, \tag{5.48}$$

$$\tau_k \leq \sum_{j\in\mathcal{K}} \mathrm{tr}\left(\mathbf{H}_k\mathbf{P}_j\right) + \sigma_n^2, \ \forall k \in \mathcal{K}, \tag{5.49}$$

$$\tau_k \log\left(\tau_k\right) \geq \tau_k\eta_k, \ \forall k \in \mathcal{K}. \tag{5.50}$$

The left-hand sides of (5.48) and (5.50) are non-convex, so we compute the first-order Taylor approximations, which are respectively

$$\tau_{c,k}^{[n]} \log\left(\tau_{c,k}^{[n]}\right) + \left(\tau_{c,k} - \tau_{c,k}^{[n]}\right)\left[\log\left(\tau_{c,k}^{[n]}\right) + 1\right] \geq \tau_{c,k}\eta_{c,k}, \ \forall k \in \mathcal{K}, \tag{5.51}$$

$$\tau_k^{[n]} \log\left(\tau_k^{[n]}\right) + \left(\tau_k - \tau_k^{[n]}\right)\left[\log\left(\tau_k^{[n]}\right) + 1\right] \geq \tau_k\eta_k \ \forall k \in \mathcal{K}. \tag{5.52}$$

The equivalent SOC forms are

$$\left\|\left[\tau_{c,k} + \eta_{c,k} - \left(\log\left(\tau_{c,k}^{[n]}\right) + 1\right), \ 2\sqrt{\tau_{c,k}^{[n]}}\right]\right\|_2 \leq \tau_{c,k} - \eta_{c,k} + \log\left(\tau_{c,k}^{[n]}\right) + 1, \ \forall k \in \mathcal{K}, \tag{5.53}$$

$$\left\|\left[\tau_k + \eta_k - \left(\log\left(\tau_k^{[n]}\right) + 1\right), \ 2\sqrt{\tau_k^{[n]}}\right]\right\|_2 \leq \tau_k - \eta_k + \log\left(\tau_k^{[n]}\right) + 1, \ \forall k \in \mathcal{K}. \tag{5.54}$$

For the rank-one constraints (5.34), we can build an iterative penalty function to insert these rank-one constraints into the objective function. By defining $\mathbf{v}_{c,\max}^{[n]}$ as the the normalized eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}\left(\mathbf{P}_c^{[n]}\right)$, and $\left\{\mathbf{v}_{k,\max}^{[n]}\right\}_{k=1}^K$ as the the normalized eigenvector corresponding to $\left\{\lambda_{\max}\left(\mathbf{P}_k^{[n]}\right)\right\}_{k=1}^K$, the problem (5.30) - (5.38) can be reformulated by

$$\begin{aligned}
\mathcal{Y}: \max_{\mathbf{P}_c,\{\mathbf{P}_k\}_{k=1}^K,\mathbf{c},t_{\mathrm{FIM}},\mathbf{r},q,\eta,\beta,\tau} \ & q + \lambda t_{\mathrm{FIM}} - \mathrm{PF} \qquad (5.55)\\
s.t. \ & (5.31) - (5.33), (5.36), (5.38)\\
& (5.39), (5.42), (5.45), (5.46), (5.47)\\
& (5.49), (5.53), (5.54)
\end{aligned}$$

where η, β, τ are defined as the sets of introduced slack variables. The iterative penalty function is expressed by

$$\begin{aligned} \mathrm{PF} = \lambda_{\mathrm{pf}} \Big(\Big[\mathrm{tr}\left(\mathbf{P}_c\right) - \left(\mathbf{v}_{c,\max}^{[n]}\right)^H \mathbf{P}_c \mathbf{v}_{c,\max}^{[n]} \Big] \\ + \sum_{k=1}^{K} \Big[\mathrm{tr}\left(\mathbf{P}_k\right) - \left(\mathbf{v}_{k,\max}^{[n]}\right)^H \mathbf{P}_k \mathbf{v}_{k,\max}^{[n]} \Big] \Big). \end{aligned} \tag{5.56}$$

λ_{pf} is a proper penalty factor to guarantee the penalty function as small as possible. Problem (5.55) is convex and can be effectively solved by the CVX toolbox. The results obtained from the n-th iteration are treated as constants while solving (5.55).

We summarize the procedure of this ISAC beamforming design in Algorithm 4. ε is the tolerance value. The convergence of Algorithm 4 is guaranteed since the solution of Problem (5.55) at iteration-n is a feasible solution to the problem at iteration-$n+1$. Finally, eigenvalue decomposition can be used to calculate the optimized beamforming vectors, and the optimized CRB is obtained accordingly. Note that Problem (5.55) involves only SOC and LMI constraints, it can be solved by using interior-point methods with the worst-case computational complexity $\mathcal{O}\big(\log(\varepsilon^{-1})[N_t^2\left(K+1\right)]^{3.5}\big)$ [17, 105, 110].

Algorithm 4 ISAC Beamforming Optimization

Initialize: $n \leftarrow 0, \mathbf{P}_c^{[n]}, \left\{\mathbf{P}_k^{[n]}\right\}_{k=1}^K, \beta^{[n]}, \tau^{[n]}$;
repeat
Solve the problem $\mathcal{Y}$ at $\mathbf{P}_c^{[n]}, \left\{\mathbf{P}_k^{[n]}\right\}_{k=1}^K, \beta^{[n]}, \tau^{[n]}$ to get the optimal $\breve{\mathbf{P}}_c, \left\{\breve{\mathbf{P}}_k\right\}_{k=1}^K, \breve{\beta}, \breve{\tau}, \breve{\text{objective}}$;
$n \leftarrow n+1$;
Update $\mathbf{P}_c^{[n]} \leftarrow \breve{\mathbf{P}}_c, \left\{\mathbf{P}_k^{[n]}\right\}_{k=1}^K \leftarrow \left\{\breve{\mathbf{P}}_k\right\}_{k=1}^K, \beta^{[n]} \leftarrow \breve{\beta}, \tau^{[n]} \leftarrow \breve{\tau}, \text{objective}^{[n]} \leftarrow \breve{\text{objective}}$;
until $\left|\text{objective}^{[n]} - \text{objective}^{[n-1]}\right| < \varepsilon$;

5.4 Simulation Results and Analysis

In this section, the performance of the proposed algorithm is evaluated using simulation results of both terrestrial and satellite ISAC systems. The performance of RSMA-assisted ISAC is evaluated in terms of the trade-off between MFR and Root CRB (RCRB).

First, we consider a terrestrial radar-communication system where the ISAC platform is equipped with $N_t = 8$ transmit antennas and $N_r = 9$ receive antennas. The system employs a uniform linear array (ULA) with half-wavelength adjacent antenna spacing. The sum transmit power budget is $P = 20$ dBm, and the noise power at each user is $\sigma_m^2 = 0$ dBm. The communication channel is set as Rayleigh fading with each entry drawn from $\mathcal{CN} \sim (0, 1)$. We assume $K = 4$ communication users, and the target is located at $\theta = 45°$. The relative target velocity is $v = 8$ m/s. The number of transmit symbols within one CPI is $L = 1024$. In Fig. 5.2, the curves of the trade-off between MFR and RCRB of different target parameters are plotted. All results are obtained by solving the formulated optimization problem and all results are averaged over 100 channel realizations. The radar SNR is defined as $\text{SNR}_{\text{radar}} = |\alpha|^2 P/\sigma_m^2 = -20$ dB. For the baseline strategies, SDMA-assisted ISAC can be simulated as a special case of RSMA by turning off the common stream. The decoding order of NOMA-assisted ISAC is the ascending order of channel strengths. No user grouping is considered. We can observe that when the priority is the communication functionality, both RSMA-assisted and SDMA-assisted ISAC achieve similar MFR. As the priority is shifted to sensing, the RSMA-assisted ISAC achieves a considerably better trade-off compared with SDMA. Similar trade-off performance can be observed in [82] where the ISAC beamforming was designed by optimizing the communication and radar metric, namely, WSR and beampattern MSE. From Fig. 5.2, the NOMA-assisted ISAC achieves the poorest trade-off due to the DoF loss in multi-antenna NOMA [8]. At the leftmost points which correspond to prioritizing

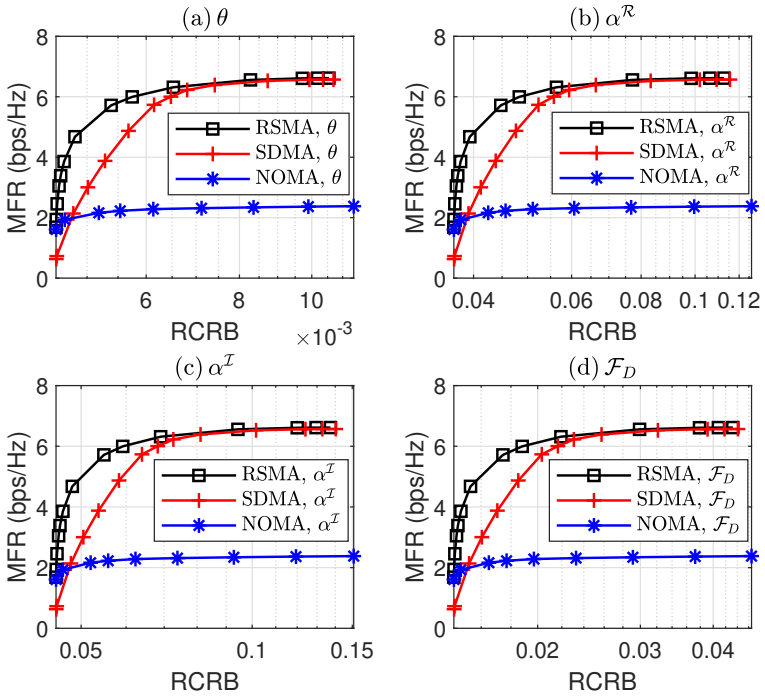

Figure 5.2: MFR versus RCRB in a terrestrial ISAC system, (a) θ (°), (b) $\alpha^{\mathfrak{R}}$, (c) $\alpha^{\mathfrak{I}}$, (d) $\mathcal{F}_D$. $N_t = 8,\ N_r = 9,\ K = 4,\ L = 1024, \mathrm{SNR}_{\mathrm{radar}} = -20$ dB.

the radar functionality, the optimized precoders are linearly dependent[2] on each other. Thus, the SDMA-assisted ISAC can no longer exploit spatial DoF provided by multiple antennas and leads to lower MFR compared with the RSMA-assisted and NOMA-assisted ISAC which employ SIC at user sides to manage the multiuser interference.

The sensing capability at the radar receiver is evaluated in Fig. 5.3 in terms of the target estimation root mean square error (RMSE). Radar subspace-based estimation algorithms, e.g., [111] can be used to estimate the Doppler frequency, the direction of the target and the reflection coefficient from the radar received signal. Throughout the simulations, communication symbols $\mathbf{s}[l]$ in (5.1) are generated as random quadrature-phase-shift-keying (QPSK) modulated sequences, and the precoders are obtained by solving the formulated ISAC beamforming optimization problem. Fig.

[2]From the simulation results, we can observe that the optimized precoders are linearly dependent on each other. Intuitively, when mostly prioritizing the radar functionality, the optimized precoders are designed to radiate the highest power towards the target angle.

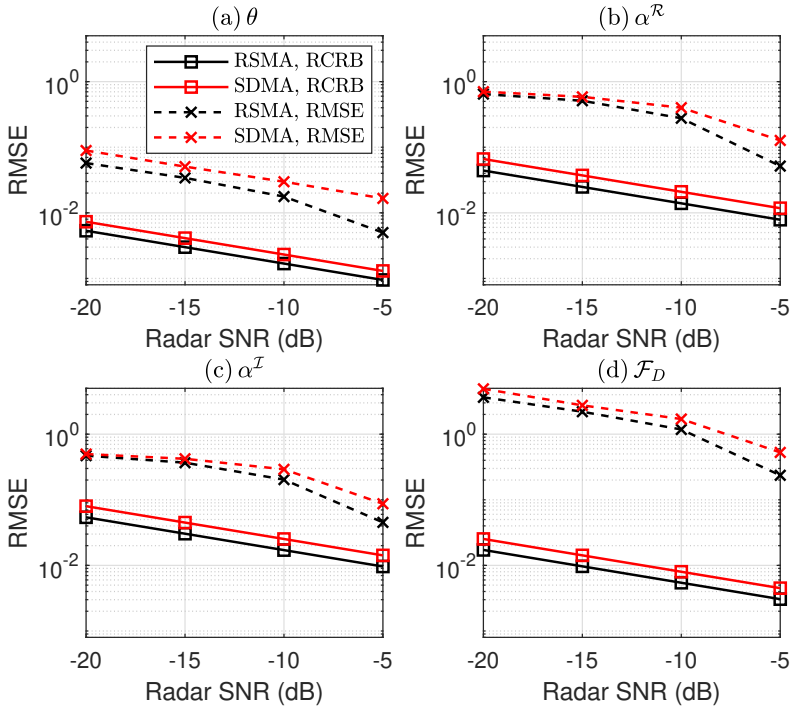

Figure 5.3: Target estimation performance in a terrestrial ISAC system, (a) θ (°), (b) $\alpha^{\Re}$, (c) $\alpha^{\Im}$, (d) $\mathcal{F}_D$. $N_t = 8,\ N_r = 9,\ K = 4,\ L = 1024$.

5.3 depicts the RMSE and RCRB with the increase of radar SNR while setting the MFR of RSMA-assisted and SDMA-assisted ISAC to be 6 bps/Hz. NOMA-assisted ISAC is not evaluated due to its poor MFR performance and 6 bps/Hz cannot be satisfied. We can observe that the RMSEs of different target parameters are lower-bounded by the corresponding RCRBs, and are expected to approach the RCRBs at high radar SNR regimes. As expected, the RSMA-assisted ISAC always outperforms SDMA-assisted ISAC in terms of the target parameter estimation performance.

Next, a satellite radar-communication system is considered, where the ISAC satellite could be a multibeam LEO satellite simultaneously serving single-antenna satellite users and sensing a moving target within the satellite coverage area. Considering a single feed per beam architecture, which is popular in modern satellites such as Eutelsat Ka-Sat, where one antenna feed is required to generate one beam. We can simply assume $\rho = 2$ uniformly distributed satellite users in each beam, and

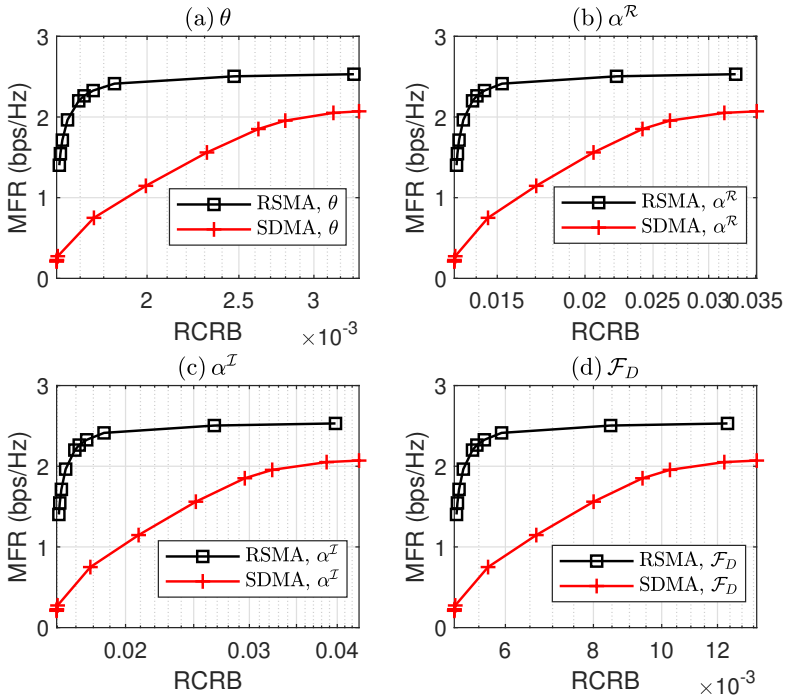

Figure 5.4: MFR versus RCRB in a satellite ISAC system, (a) θ (°), (b) $\alpha^{\Re}$, (c) $\alpha^{\Im}$, (d) $\mathcal{F}_D$. $N_t = 8$, $N_r = 9$, $K = 16$, $L = 1024, \text{SNR}_{\text{radar}} = -20$ dB.

the multibeam satellite channel model has been discussed in the previous chapters. $K = \rho N_t = 16$ satellite users follow multibeam multicast transmission. Fig. 5.4 shows the trade-off curves between MFR and RCRB in a multibeam satellite ISAC system. From Fig. 5.4, the trade-off performance gain provided by RSMA-assisted design is more obvious than the terrestrial scenario given in Fig. 5.2. The gaps between RSMA-assisted and SDMA-assisted ISAC can be observed from the rightmost points which correspond to prioritizing the communication functionality. This is due to the superiority of using RSMA in an overloaded communication system [68]. Since NOMA leads to extremely high receiver complexity when the number of users is large and also a waste of spatial resources in multi-antenna settings, we do not compare with NOMA-assisted ISAC in this scenario. Above all, we can conclude that RSMA is a very effective and powerful strategy for both terrestrial and satellite ISAC systems to manage the multiuser/inter-beam interference as well as performing the radar functionality.

5.5 Summary

In this chapter, we provide an overview of the interplay between two promising technologies, namely, RSMA and ISAC. We start from a general RSMA-assisted ISAC model and introduced the performance metrics for both radar sensing and communications. Then, we introduce a design example which jointly minimizes the CRB of target estimation and maximizes MFR amongst communication users subject to the per-antenna power constraint. Through simulation results, RSMA is demonstrated to be a very powerful and promising technique for ISAC systems in both terrestrial and satellite scenarios.

Chapter 6

Conclusion

6.1 Summary of Thesis Achievements

In this thesis, we considered the problems of applying RSMA to non-terrestrial communication and sensing networks, and addressed a number of optimization problems and algorithms in various scenarios, namely the multigroup multicast and multibeam satellite systems, STIN and ISAC systems, which are envisioned to play key roles in next-generation wireless networks. Simulation results and analysis are presented to evaluate the performance of all the proposed algorithms.

In Chapter 3, we explored the benefits of adopting RSMA for multigroup/multibeam multicast in the presence of imperfect CSIT. We considered both underloaded and overloaded regimes and addressed the problem to achieve max-min fairness. Through MMF-DoF analysis, RSMA was shown to provide gains in both underloaded and overloaded regimes compared with the conventional scheme. Then, we formulated a generic MMF optimization problem and developed a WMMSE algorithm based on SAA to solve the optimization. Through simulation results, the DoF gains of RSMA over the conventional scheme were shown to translate into rate benefits. The

effectiveness of using RSMA for multigroup multicast and multibeam satellite systems was demonstrated taking into account CSIT uncertainty and practical challenges in multibeam satellite systems, such as per-feed transmit power constraints, hotspots, uneven user distribution per beam and overloaded regimes. The RSMA transmitter and receiver architecture, PHY layer design and LLS platform were also investigated, including finite length polar coding, finite alphabet modulation, AMC algorithm, etc. LLS results showed that RSMA is a very promising MA scheme for practical implementation in numerous application areas.

In Chapter 4, motivated by the benefits of RSMA presented in Chapter 3, we further investigated the application of RSMA to STIN considering either perfect CSI or imperfect CSI with satellite channel phase uncertainties at the GW to manage the interference within and between both sub-networks. A multiuser downlink framework was presented for the integrated network where the satellite exploits multibeam multicast communication to serve SUs, while the terrestrial BS employs UPA and serves CUs in a densely populated area. RSMA can be used at both the satellite and the BS to mitigate the interference including inter-beam interference, intra-cell interference and interference between the two sub-systems. Two RSMA-based STIN schemes were presented, namely the coordinated scheme and the cooperative scheme. For the coordinated scheme, the satellite and BS exchanged CSI of both direct and interfering links at the GW, and coordinated beamforming to manage the interference. For the cooperative scheme, the satellite and BS exchanged both CSI and data at the GW. All propagation links (including interfering ones) were exploited to carry useful data upon appropriate beamforming. MMF optimization problems were formulated. To tackle the optimization, an iterative algorithm was proposed based on the SCA approach to reformulating the original problem into an equivalent convex one, which belongs to a SOCP. Then, we considered imperfect CSIT with satellite channel phase uncertainty. An expectation-based robust beamforming optimization algorithm was developed using SCA together with a penalty function. Simulation

results demonstrated the superiority and robustness of the proposed RSMA-based cooperative scheme and coordinated scheme compared with the baseline strategies. Therefore, RSMA was shown very promising for STIN to manage the interference in and between the satellite and terrestrial sub-systems.

In Chapter 5, RSMA was extended to the ISAC setup to make better use of the RF spectrum and infrastructure. We investigated a general RSMA-assisted ISAC system, where the antenna array is shared by a co-located monostatic MIMO radar system and a multiuser communication system. The problem addressed the trade-off between serving multiple downlink communication users and sensing a moving target. Explicit optimization of estimation performance at the radar receiver was concerned with. We formulated an RSMA-assisted ISAC beamforming optimization problem to jointly minimize the CRB of the target estimation and maximize the minimum fairness rate amongst all communication users subject to transmit power constraints. An iterative algorithm based on SCA was developed to solve the optimization. Simulation results demonstrated the benefits of RSMA for both terrestrial and satellite ISAC systems to manage the multiuser/inter-beam interference and simultaneously perform the radar functionality.

6.2 Future Work

In conclusion of this thesis, some potential future research directions are listed as follows:

1) RSMA for SAGIN:

The space-air-ground integrated network (SAGIN), which integrates spaceborne, airborne and terrestrial/marine networks has been envisioned to provide heterogeneous services and seamless network coverage. The spaceborne part consists of diverse

types of satellites and constellations, while the airborne network consists of balloons, aeroplanes, unmanned aerial vehicles (UAVs), etc. However, due to the spectrum sharing among these segments, interference becomes one of the major challenges and advanced interference management schemes are required. RSMA is envisioned to enhance the system performance as it leverages two extreme interference management strategies, namely fully treating interference as noise and fully decoding interference. In addition to the work addressed in Chapter 4 of this thesis, which focused on the integration of a GEO satellite and a single terrestrial BS, the integration between more platforms could be further explored. Compared with satellites and terrestrial BSs, UAVs enjoy much higher mobility, ease of deployment, coverage extension and low cost. The challenges are their high mobility and limited battery capacity to fly, hover and communicate. Facing these practical issues, RSMA has great potential to tackle these challenges because of its robustness towards CSIT imperfections, and capability to reduce communication energy consumption. UAVs may act as aerial BSs, relays or aerial receivers, which present great compatibility with SAGIN to enhance the network services. Moreover, the interplay of RSMA for SAGIN with other enablers such as machine learning (ML) is also worth studying to achieve ubiquitous intelligent connectivity.

2) RSMA-assisted ISAC with mmWave:

The explosive growth of data traffic and the scarcity of spectrum resources have motivated the investigation of millimeter wave (mmWave) communications. A number of ISAC scenarios involve mmWave frequencies. The frequency band from 30 GHz to 300 GHz requires massive antennas to overcome path losses. It shows potentials to achieve high data rates for communication and high resolution for radar operation due to the huge available bandwidth in the mmWave frequency bands and multiplexing gains achievable with massive antenna arrays. To reduce the transceiver hardware complexity and power consumption, hybrid analog-digital (HAD) structure

is typically used, which is able to reduce the number of required RF chains and achieve higher energy efficiency compared to fully digital precoding. HAD precoding design for ISAC systems at the mmWave band has been investigated in [112,113] to provide efficient trade-offs between downlink communications and radar performance. Inspired by the appealing advantages of RSMA in multi-antenna systems, the benefits of introducing RSMA and HAD to tackle the multiuser interference in the context of mmWave communications have been demonstrated in [31,96,114]. As a consequence, the interplay between RSMA-assisted ISAC and HAD for mmWave is becoming another interesting research topic.

3) RSMA-assisted ISAC with V2X:

For the coming generation of vehicle-to-everything (V2X) networks, ISAC serves as a particularly suitable technology aiming to jointly provide high throughput vehicular communication service and remote sensing service for vehicle localization and anti-collision detection [115]. The characteristics of vehicular networks include high mobility, rigorous requirements on transmission latency and reliability, etc. Recent studies have shown that RSMA is robust against CSIT imperfections resulting from user mobility and feedback delay in multiuser (Massive) MIMO [116], and it outperforms existing MA schemes in finite block length regimes [117,118]. Therefore, RSMA-assisted ISAC has great potential to become a promising research topic for future V2X networks.

4) RSMA-assisted ISAC with OFDM:

OFDM has been widely investigated as one of the key techniques in wireless networks. It was also found to be useful in radar sensing [74]. Due to the promising application to radar sensing, and the key role in 4G and 5G wireless communication standards, OFDM waveforms for ISAC systems have been explored in [119,120]. The benefits of implementing RSMA in a multicarrier communication system have

been demonstrated in [121, 122]. Thus, implementing RSMA in an OFDM-based ISAC system is worth being investigated as a future direction.

Appendix A

Transceiver modules

The transmitter and receiver architecture for RSMA multigroup multicast is depicted in Fig. 3.9. Detailed explanations of each module are described as follows:

1) Encoder:

From Fig. 3.9, $\mathbf{w}_{c,1}, \cdots, \mathbf{w}_{c,M}$ represent all common parts of the group messages, which are bit vectors of length $K_{c,1}, \cdots, K_{c,M}$. All private parts of the group messages are denoted by $\mathbf{w}_{p,1}, \cdots, \mathbf{w}_{p,M}$, which are bit vectors of length $K_{p,1}, \cdots, K_{p,M}$. Through the encoder, all common parts $\mathbf{w}_{c,1}, \cdots, \mathbf{w}_{c,M}$ are jointly encoded into a common codeword ν_c of code block length N_c, while the private parts $\mathbf{w}_{p,1}, \cdots, \mathbf{w}_{p,M}$ are encoded individually into private codewords $\nu_{p,1}, \cdots, \nu_{p,M}$. The code block lengths are respectively $N_{p,1}, \cdots, N_{p,M}$. We consider polar coding for the channel coding process. The block length of a conventional polar code is expressed as $N = 2^n$, where n is a positive integer. The polar encoding operation can be written as $\nu = \mathbf{u}\mathbf{G}_N$, where $\mathbf{G}_N = \mathbf{B}_N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}$. $\mathbf{B}_N$ is the bit-reversal matrix and $\otimes n$ represents the n-fold Kronecker product. $\mathbf{u}$ denotes the length-N uncoded bit vector input to the encoder which consists of K information bits and $N-K$ frozen bits. Let $\mathcal{A} \in \{1, \cdots, N\}$ be the set of positions of the information bits, and $\mathcal{A}^c$ be the set of

positions of the frozen bits. Therefore, we have $\mathcal{A} \cap \mathcal{A}^c = \phi$ and $\mathcal{A} \cup \mathcal{A}^c = \{1, \cdots, N\}$. Specifically, we can construct the private uncoded bit vectors $\mathbf{u}_{p,1}, \cdots, \mathbf{u}_{p,M}$ by setting $\mathbf{u}_{p,m,\mathcal{A}_m} = \mathbf{w}_{p,m}, \forall m \in \mathcal{M}$. The sets $\mathcal{A}_{p,1}, \cdots, \mathcal{A}_{p,M}$ contain information bit indices of the private messages. To jointly encode the common information bit vectors, $\mathbf{w}_{c,1}, \cdots, \mathbf{w}_{c,M}$ are at first appended into $\mathbf{w}_c = [\mathbf{w}_{c,1}, \cdots, \mathbf{w}_{c,M}]$. Then, the common uncoded bit vector $\mathbf{u}_c$ is constructed by setting $\mathbf{u}_{c,\mathcal{A}_c} = \mathbf{w}_c$, where the set $\mathcal{A}_c$ collects information bit indices of the common message. Values of all frozen bits are fixed and known by both the encoder and the decoder. After obtaining the codewords ν_c and $\nu_1, \cdots, \nu_M$, interleavers are adopted before modulation.

2) Modulator:

The interleavered bit vectors $\nu'_c, \nu'_1, \cdots, \nu'_M$ are respectively modulated into a common stream $\mathbf{s}_c$ and multiple private streams $\mathbf{s}_1, \cdots, \mathbf{s}_M$. For a given modulation scheme with alphabet $\mathcal{M}$ and modulation order $|\mathcal{M}| = 2^m$, the interleavered bits $(\nu'_{mi+1}, \cdots, \nu'_{mi+m})$, $i \in \left\{0, 1, \cdots, \frac{N}{m} - 1\right\}$ are mapped to a constellation signal $s \in \mathcal{M}$ according to the Gray labeling. If a stream $\mathbf{s}$ is of length S, its corresponding code block length is $N = mS$.

3) AMC Algorithm:

Appropriate modulation schemes and coding parameters are determined by the AMC algorithm to maximize the system throughput depending on the channel characteristics. The algorithm uses the Average rates $\overline{R}_c$ and $\overline{r}_1, \cdots, \overline{r}_M$ obtained from the MMF optimization problems with assumptions of Gaussian signalling and infinite block length. The Average rates of the common and private streams are actually calculated based on the optimized precoders by taking an average of 1000 channel realizations due to the effects of imperfect CSIT. According to each given Average rate, we first determine a corresponding modulation scheme from a modulation alphabet set $\mathcal{Q}$. Here, we consider quadrature amplitude modulation

(QAM) schemes including 4-QAM, 16-QAM, 64-QAM and 256-QAM. The set of feasible modulation schemes for a given rate $\overline{R}_l \in \left\{\overline{R}_c, \overline{r}_1, \cdots, \overline{r}_M\right\}$ is given by

$$\mathcal{Q}\left(\overline{R}_l, \beta\right) = \left\{\mathcal{M} : \log_2|\mathcal{M}| \geq \min\left(\frac{\overline{R}_l}{\beta}, m'\right), \mathcal{M} \in \mathcal{Q}\right\}. \tag{A.1}$$

where β is the maximum code rate indicating the proportion of information. m' is the logarithm of the highest modulation order, i.e., $m' = 8$ for 256-QAM in this work. For all $\overline{R}_l \in \left\{\overline{R}_c, \overline{r}_1, \cdots, \overline{r}_M\right\}$, the modulation alphabets of the common and private streams are determined by

$$\mathcal{M}_l = \text{argmin}_{\mathcal{M}\in\mathcal{Q}\left(\overline{R}_l,\beta\right)}\left|\mathcal{M}\right|, \ \forall l \in \{c, 1, \cdots, M\}. \tag{A.2}$$

Thus, when all the streams are of length S, the code block lengths and code rates are respectively calculated as

$$N_l = S\log_2\left(\left|\mathcal{M}_l\right|\right), \ \forall l \in \{c, 1, \cdots, M\}, \tag{A.3}$$

$$r_l = \frac{\left\lceil N_l \text{min}\left(\frac{\overline{R}_l}{\log_2|\mathcal{M}_l|}, \beta\right)\right\rceil}{N_l}, \ \forall l \in \{c, 1, \cdots, M\}. \tag{A.4}$$

4) Equalizer:

For each user-$k \in \mathcal{K}$, MMSE equalizers are used to detect the common and private streams. The common stream equalizer $g_{c,k}^{MMSE}$ is calculated by minimising the MSE $\varepsilon_{c,k} = \mathbb{E}\left\{\left|g_{c,k}y_k - s_c\right|^2\right\} = \left|g_{c,k}\right|^2 T_{c,k} - 2\mathcal{R}\left\{g_{c,k}\mathbf{h}_k^H\mathbf{p}_c\right\} + 1$, where $T_{c,k} = \left|\mathbf{h}_k^H\mathbf{p}_c\right|^2 + \left|\mathbf{h}_k^H\mathbf{p}_{\mu(k)}\right|^2 + \sum_{j=1,j\neq\mu(k)}^{M}\left|\mathbf{h}_k^H\mathbf{p}_j\right|^2 + \sigma_n^2$. To minimize the MSEs, we let $\frac{\partial\varepsilon_{c,k}}{\partial g_{c,k}} = 0$ and obtain

$$g_{c,k}^{MMSE} = \mathbf{p}_c^H\mathbf{h}_k T_{c,k}^{-1} = \frac{\mathbf{p}_c^H\mathbf{h}_k}{\left|\mathbf{h}_k^H\mathbf{p}_c\right|^2 + \sum_{j=1}^{M}\left|\mathbf{h}_k^H\mathbf{p}_j\right|^2 + \sigma_n^2}. \tag{A.5}$$

After the common stream is reconstructed and subtracted, the private stream equal-

izer g_k^{MMSE} is calculated by minimising the MSE $\varepsilon_k = \mathbb{E}\{ \left| g_k \left(y_k - \mathbf{h}_k^H \mathbf{p}_c s_c \right) - s_k \right|^2 \} = |g_k|^2 T_k - 2\mathcal{R}\left\{ g_k \mathbf{h}_k^H \mathbf{p}_{\mu(k)} \right\} + 1$, where $T_k = T_{c,k} - \left| \mathbf{h}_k^H \mathbf{p}_c \right|^2$. By letting $\frac{\partial \varepsilon_k}{\partial g_k} = 0$, the MMSE equalizers for private streams writes as

$$g_k^{MMSE} = \mathbf{p}_{\mu(k)}^H \mathbf{h}_k T_k^{-1} = \frac{\mathbf{p}_{\mu(k)}^H \mathbf{h}_k}{\sum_{j=1}^{M} \left| \mathbf{h}_k^H \mathbf{p}_j \right|^2 + \sigma_n^2}. \tag{A.6}$$

5) Demodulator and Decoder:

We use the log-likelihood ratio (LLR) method [32, 123], which is an efficient demodulator in bit-interleaved coded modulation (BICM) systems and is calculated from the equalized signal for Soft Decision (SD) decoding of polar codes. A conventional polar decoder is then employed [124]. From Fig. 3.9, it should be noted that signal reconstruction is performed at the output of the polar decoder. The reconstruction module is the same as the process at the transmitter to reconstruct a precoded signal for SIC.

References

[1] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *arXiv preprint arXiv:2209.00491*, 2022.

[2] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Communications Surveys & Tutorials*, 2022.

[3] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Transactions on Signal processing*, vol. 63, no. 9, pp. 2239–2252, 2015.

[4] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Transactions on Signal processing*, vol. 63, no. 14, pp. 3628–3640, 2015.

[5] Y. Mao, B. Clerckx, and V. O. 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, p. 133, 2018.

[6] 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, 2016.

[7] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[8] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1310–1343, 2021.

[9] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Communications Letters*, 2019.

[10] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Network*, vol. 35, no. 2, pp. 244–251, 2020.

[11] M. Á. Vázquez, A. Perez-Neira, D. Christopoulos, S. Chatzinotas, B. Ottersten, P.-D. Arapoglou, A. Ginesi, and G. Taricco, "Precoding in multibeam satellite communications: Present and future challenges," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 88–95, 2016.

[12] A. I. Perez-Neira, M. A. Vazquez, M. B. Shankar, S. Maleki, and S. Chatzinotas, "Signal processing for high-throughput satellites: Challenges in new interference-limited scenarios," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 112–131, 2019.

[13] X. Zhu and C. Jiang, "Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 437–461, 2021.

[14] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *IEEE journal on selected areas in communications*, 2022.

[15] B. Lee and W. Shin, "Max-min fairness precoder design for rate-splitting multiple access: Impact of imperfect channel knowledge," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 1355–1359, 2023.

[16] 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, 2016.

[17] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8754–8770, 2019.

[18] A. A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, and M.-S. Alouini, "Interference mitigation via rate-splitting and common message decoding in cloud radio access networks," *IEEE Access*, vol. 7, pp. 80350–80365, 2019.

[19] J. Zhang, B. Clerckx, J. Ge, and Y. Mao, "Cooperative rate splitting for MISO broadcast channel with user relaying, and performance benefits over cooperative NOMA," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1678–1682, 2019.

[20] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3232–3246, 2015.

[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, 2016.

[22] 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, 2017.

[23] 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, 2016.

[24] Y. Mao and B. Clerckx, "Beyond dirty paper coding for multi-antenna broadcast channel with partial CSIT: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6775–6791, 2020.

[25] M. Caus, A. Pastore, M. Navarro, T. Ramírez, C. Mosquera, N. Noels, N. Alagha, and A. I. Perez-Neira, "Exploratory analysis of superposition coding and rate splitting for multibeam satellite systems," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–5, IEEE, 2018.

[26] 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, 2017.

[27] H. Chen, D. Mi, B. Clerckx, Z. Chu, J. Shi, and P. Xiao, "Joint power and subcarrier allocation optimization for multigroup multicast systems with rate splitting," *IEEE Transactions on Vehicular Technology*, 2019.

[28] A. Z. Yalcin, M. Yuksel, and B. Clerckx, "Rate splitting for multi-group multicasting with a common message," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12281–12285, 2020.

[29] O. Tervo, L.-N. Trant, S. Chatzinotas, B. Ottersten, and M. Juntti, "Multigroup multicast beamforming and antenna selection with rate-splitting in multicell systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2018.

[30] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8196–8211, 2017.

[31] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7025–7038, 2017.

[32] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, "Rate-splitting multiple access for downlink multi-antenna communications: Physical layer design and link-level simulations," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–6, IEEE, 2020.

[33] L. Yin, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for multi-group multicast cellular and satellite communications: PHY layer design and link-level simulations," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2021.

[34] T. De Cola, A. Ginesi, G. Giambene, G. C. Polyzos, V. A. Siris, N. Fotiou, Y. Thomas, *et al.*, "Network and protocol architectures for future satellite systems," *Foundations and Trends® in Networking*, vol. 12, no. 1-2, pp. 1–161, 2017.

[35] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4695–4707, 2015.

[36] J. Wang, L. Zhou, K. Yang, X. Wang, and Y. Liu, "Multicast precoding for multigateway multibeam satellite systems with feeder link interference," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1637–1650, 2019.

[37] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Processing*, vol. 54, no. 6-1, pp. 2239–2251, 2006.

[38] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, 2008.

[39] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," in *Advances in neural information processing systems*, pp. 1759–1767, 2009.

[40] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2685–2698, 2017.

[41] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5132–5142, 2014.

[42] B. Hu, C. Hua, C. Chen, and X. Guan, "Multicast beamforming for wireless backhaul with user-centric clustering in cloud-RANs," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2016.

[43] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 12–21, 2012.

[44] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6118–6131, 2016.

[45] M. Sadeghi, E. Björnson, E. G. Larsson, C. Yuen, and T. L. Marzetta, "Max–min fair transmit precoding for multi-group multicasting in massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1358–1373, 2017.

[46] G. Zheng, S. Chatzinotas, and B. Ottersten, "Generic optimization of linear precoding in multibeam satellite systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2308–2320, 2012.

[47] V. Joroughi, M. Á. Vázquez, and A. I. Pérez-Neira, "Generalized multicast multibeam precoding for satellite communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 952–966, 2016.

[48] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[49] Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama, N. Kato, and M. Toyoshima, "Prospects and challenges of context-aware multimedia content delivery in cooperative satellite and terrestrial networks," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 55–61, 2014.

[50] H. Zhang, C. Jiang, J. Wang, L. Wang, Y. Ren, and L. Hanzo, "Multicast beamforming optimization in cloud-based heterogeneous terrestrial and satellite networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1766–1776, 2019.

[51] N. U. Hassan, C. Huang, C. Yuen, A. Ahmad, and Y. Zhang, "Dense small satellite networks for modern terrestrial communication systems: Benefits, infrastructure, and technologies," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 96–103, 2020.

[52] M. Lin, Z. Lin, W.-P. Zhu, and J.-B. Wang, "Joint beamforming for secure communication in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 1017–1029, 2018.

[53] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Transmit beamforming for spectral coexistence of satellite and terrestrial networks," in *8th Int. Conf. Cogn. Radio Oriented Wireless Netw.*, pp. 275–281, 2013.

[54] X. Zhu, C. Jiang, L. Yin, L. Kuang, N. Ge, and J. Lu, "Cooperative multigroup multicast transmission in integrated terrestrial-satellite networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 5, pp. 981–992, 2018.

[55] Y. Zhang, L. Yin, C. Jiang, and Y. Qian, "Joint beamforming design and resource allocation for terrestrial-satellite cooperation system," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 778–791, 2019.

[56] Z. Lin, M. Lin, J.-B. Wang, X. Wu, and W.-P. Zhu, "Joint optimization for secure WIPT in satellite-terrestrial integrated networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, 2018.

[57] J. Li, K. Xue, D. S. Wei, J. Liu, and Y. Zhang, "Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2367–2381, 2020.

[58] W. Lu, K. An, and T. Liang, "Robust beamforming design for sum secrecy rate maximization in multibeam satellite systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1568–1572, 2019.

[59] X. Guo, D. Yang, Z. Luo, H. Wang, and J. Kuang, "Robust THP design for energy efficiency of multibeam satellite systems with imperfect CSI," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 428–432, 2019.

[60] Z. Lin, M. Lin, J. Ouyang, W.-P. Zhu, A. D. Panagopoulos, and M.-S. Alouini, "Robust secure beamforming for multibeam satellite communication systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6202–6206, 2019.

[61] A. Gharanjik, M. B. Shankar, P.-D. Arapoglou, M. Bengtsson, and B. Ottersten, "Robust precoding design for multibeam downlink satellite channel with phase uncertainty," in *Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3083–3087, 2015.

[62] W. Wang, L. Gao, R. Ding, J. Lei, L. You, C. A. Chan, and X. Gao, "Resource efficiency optimization for robust beamforming in multi-beam satellite communications," *IEEE Trans. Veh. Technol.*, 2021.

[63] J. Chu, X. Chen, C. Zhong, and Z. Zhang, "Robust design for noma-based multibeam LEO satellite internet of things," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1959–1970, 2020.

[64] L. Yin, B. Clerckx, and Y. Mao, "Rate-splitting multiple access for multi-antenna broadcast channels with statistical CSIT," in *Wireless Commun. Netw. Conf. Workshops (WCNCW)*, pp. 1–6, 2021.

[65] "Second generation framing structure, channel coding and modulation systems for broadcasting, interative services, news gathering and other broadband satellite applications; Part 2: DVB-S2 Extensions (DVB-S2X)," standard, European Broadcasting Union (EBU), document ETSI EN 302-307-2 V1.1.1, Oct. 2014.

[66] M. Vazquez, M. Caus, and A. Perez-Neira, "Rate-splitting for MIMO multi-beam satellite systems," in *WSA 2018; 22nd Int. ITG Workshop Smart Antennas*, pp. 1–6, 2018.

[67] L. Yin and B. Clerckx, "Rate-splitting multiple access for multibeam satellite communications," in *IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, pp. 1–6, 2020.

[68] L. Yin and B. Clerckx, "Rate-splitting multiple access for multigroup multicast and multibeam satellite systems," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 976–990, 2020.

[69] Z. W. Si, L. Yin, and B. Clerckx, "Rate-splitting multiple access for multigateway multibeam satellite systems with feeder link interference," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 2147–2162, 2022.

[70] Z. Lin, M. Lin, T. de Cola, J.-B. Wang, W.-P. Zhu, and J. Cheng, "Supporting IoT with rate-splitting multiple access in satellite and aerial integrated networks," *IEEE Internet Things J.*, 2021.

[71] Z. Lin, M. Lin, B. Champagne, W.-P. Zhu, and N. Al-Dhahir, "Secure and energy efficient transmission for RSMA-based cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 251–255, 2020.

[72] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-rao bound optimization for joint radar-communication beamforming," *IEEE Transactions on Signal Processing*, vol. 70, pp. 240–253, 2021.

[73] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Dual-function radar-communications: Information embedding using sidelobe control and waveform diversity," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2168–2181, 2015.

[74] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.

[75] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, 2020.

[76] H. Wymeersch, G. Seco-Granados, G. Destino, D. Dardari, and F. Tufvesson, "5G mmWave positioning for vehicular networks," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 80–86, 2017.

[77] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150–157, 2015.

[78] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2755–2770, 2018.

[79] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4264–4279, 2018.

[80] C. Xu, B. Clerckx, S. Chen, Y. Mao, and J. Zhang, "Rate-splitting multiple access for multi-antenna joint communication and radar transmissions," in *IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, pp. 1–6, 2020.

[81] C. Xu, B. Clerckx, S. Chen, Y. Mao, and J. Zhang, "Rate-splitting multiple access for multi-antenna joint radar and communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1332–1347, 2021.

[82] R. C. Loli, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for multi-antenna joint radar and communications with partial CSIT: Precoder optimization and link-level simulations," *arXiv preprint arXiv:2201.10621*, 2022.

[83] O. Dizdar, A. Kaushik, B. Clerckx, and C. Masouros, "Energy efficient dual-functional radar-communication: Rate-splitting multiple access, low-resolution DACs, and RF chain selection," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 986–1006, 2022.

[84] M. Caus, A. Pastore, M. Navarro, T. Ramirez, C. Mosquera, N. Noels, N. Alagha, and A. I. Perez-Neira, "Exploratory analysis of superposition coding and rate splitting for multibeam satellite systems," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–5, 2018.

[85] M. Vazquez, M. Caus, and A. Perez-Neira, "Rate splitting for MIMO multibeam satellite systems," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, pp. 1–6, March 2018.

[86] B. Lee, W. Shin, and H. V. Poor, "Weighted sum-rate maximization for rate-splitting multiple access with imperfect channel knowledge," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 218–220, 2021.

[87] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on information theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[88] G. Caire, N. Jindal, and S. Shamai, "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 287–291, IEEE, 2007.

[89] H. Chen and C. Qi, "User grouping for sum-rate maximization in multiuser multibeam satellite communications," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2019.

[90] H. Chen, D. Mi, Z. Chu, P. Xiao, Y. Xu, and D. He, "Link-level performance of rate-splitting based downlink multiuser MISO systems," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, IEEE, 2020.

[91] S. Vassaki, M. I. Poulakis, A. D. Panagopoulos, and P. Constantinou, "Power allocation in cognitive satellite terrestrial networks with QoS constraints," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1344–1347, 2013.

[92] M. Jia, X. Gu, Q. Guo, W. Xiang, and N. Zhang, "Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 96–106, 2016.

[93] J. P. Choi and C. Joo, "Challenges for efficient and seamless space-terrestrial heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 156–162, 2015.

[94] Y. Zhang, Y. Wu, A. Liu, X. Xia, T. Pan, and X. Liu, "Deep learning-based channel prediction for LEO satellite massive MIMO communication system," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1835–1839, 2021.

[95] Y. Zhang, A. Liu, P. Li, and S. Jiang, "Deep learning (DL)-based channel prediction and hybrid beamforming for LEO satellite massive MIMO system," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 23705–23715, 2022.

[96] Z. Li, S. Yang, and T. Clessienne, "A general rate splitting scheme for hybrid precoding in mmWave systems," in *IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, 2019.

[97] M. Aloisio and P. Angeletti, "Multi-amplifiers architectures for power reconfigurability," in *IEEE Int. Vac. Electron. Conf.*, pp. 1–2, 2007.

[98] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, 2019.

[99] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, 2015.

[100] K.-G. Nguyen, Q.-D. Vu, M. Juntti, and L.-N. Tran, "Distributed solutions for energy efficiency fairness in multicell MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6232–6247, 2017.

[101] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear algebra and its applications*, vol. 284, no. 1-3, pp. 193–228, 1998.

[102] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations research*, vol. 26, no. 4, pp. 681–683, 1978.

[103] M. Shao and W.-K. Ma, "A simple way to approximate average robust multiuser MISO transmit optimization under covariance-based CSIT," in *Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3504–3508, 2017.

[104] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, 2014.

[105] Y. Ye, *Interior point algorithms: theory and analysis*, vol. 44. John Wiley & Sons, 2011.

[106] Z. Lin, M. Lin, J.-B. Wang, T. De Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 657–670, 2019.

[107] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory.* Prentice-Hall, Inc., 1993.

[108] F. Liu, Y.-F. Liu, C. Masouros, A. Li, and Y. C. Eldar, "A joint radar-communication precoding design based on Cramér-rao bound optimization," in *2022 IEEE Radar Conference (RadarConf22)*, pp. 1–6, 2022.

[109] J. Li, L. Xu, P. Stoica, K. W. Forsythe, and D. W. Bliss, "Range compression and waveform optimization for MIMO radar: A Cramér–Rao bound based study," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 218–232, 2007.

[110] Y. Mao, B. Clerckx, J. Zhang, V. O. Li, and M. A. Arafah, "Max-min fairness of K-user cooperative rate-splitting in MISO broadcast channel with user relaying," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6362–6376, 2020.

[111] H. Ren and A. Manikas, "MIMO radar with array manifold extenders," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 1942–1954, 2020.

[112] F. Liu and C. Masouros, "Hybrid beamforming with sub-arrayed MIMO radar: Enabling joint sensing and communication at mmWave band," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7770–7774, 2019.

[113] A. Kaushik, C. Masouros, and F. Liu, "Hardware efficient joint radar-communications with hybrid precoding and RF chain optimization," in *ICC 2021 - IEEE International Conference on Communications*, pp. 1–6, 2021.

[114] O. Kolawole, A. Panazafeironoulos, and T. Ratnarajah, "A rate-splitting strategy for multi-user millimeter-wave systems with imperfect CSI," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2018.

[115] F. Liu and C. Masouros, "A tutorial on joint radar and communication transmission for vehicular networks—Part I: Background and fundamentals," *IEEE Communications Letters*, vol. 25, no. 2, pp. 322–326, 2020.

[116] O. Dizdar, Y. Mao, and B. Clerckx, "Rate-splitting multiple access to mitigate the curse of mobility in (massive) MIMO networks," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6765–6780, 2021.

[117] Y. Xu, Y. Mao, O. Dizdar, and B. Clerckx, "Max-min fairness of rate-splitting multiple access with finite blocklength communications," *IEEE Transactions on Vehicular Technology*, pp. 1–6, 2022.

[118] Y. Xu, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access with finite blocklength for short-packet and low-latency downlink communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12333–12337, 2022.

[119] M. F. Keskin, H. Wymeersch, and V. Koivunen, "MIMO-OFDM joint radar-communications: Is ICI friend or foe?," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1393–1408, 2021.

[120] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint com-

munication and radar sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1295–1315, 2021.

[121] H. Chen, D. Mi, T. Wang, Z. Chu, Y. Xu, D. He, and P. Xiao, "Rate-splitting for multicarrier multigroup multicast: Precoder design and error performance," *IEEE Transactions on Broadcasting*, vol. 67, no. 3, pp. 619–630, 2021.

[122] O. Dizdar and B. Clerckx, "Rate-splitting multiple access for communications and jamming in multi-antenna multi-carrier cognitive radio systems," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 628–643, 2022.

[123] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient mmse-based demodulator for MIMO bit-interleaved coded modulation," in *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04.*, vol. 4, pp. 2455–2459 Vol.4, 2004.

[124] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.