9-30-2016 12:00 AM

Highly Efficient Resource Allocation Techniques in 5G for NOMA-based Massive MIMO and Relaying Systems

Xin Liu, *The University of Western Ontario*

Supervisor: Xianbin Wang, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Electrical and Computer Engineering

Abstract

The explosive proliferation of smart devices in the 5-th generation (5G) network expects 1,000-fold capacity enhancement, leading to the urgent need of highly resource-efficient technologies. Non-orthogonal multiple access (NOMA), a promising spectral efficient technology for 5G to serve multiple users concurrently, can be combined with massive multiple input multiple output (MIMO) and relaying technology, to achieve highly efficient communications. Hence, this thesis studies the design and resource allocation of NOMA-based massive MIMO and relaying systems.

Due to hardware constraints and channel condition variation, the first topic of the thesis develops efficient antenna selection and user scheduling algorithms for sum rate maximization in two MIMO-NOMA scenarios. In the single-band scenario, the proposed algorithm improves antenna search efficiency by limiting the candidate antennas to those are beneficial to the relevant users. In the multi-band scenario, the proposed algorithm selects the antennas and users with the highest contribution total channel gain. Numerical results show that our proposed algorithms achieve similar performance to other algorithms with reduced complexity.

The second part of the thesis proposes the relaying and power allocation scheme for the NOMA-assisted relaying system to serve multiple cell-edge users. The relay node decodes its own message from the source NOMA signal and transmits the remaining part of signal to cell-edge users. The power allocation scheme is developed by minimizing the system outage probability. To further evaluate the system performance, the ergodic capacity is approximated by analyzing the interference at cell-edge users. Numerical results proves the performance improvement of the proposed system over conventional orthogonal multiple access mechanism.

Keywords: 5G; Massive MIMO; NOMA; Relaying; Resource Allocation

Acknowlegements

The completion of this thesis involves the contribution and supports by a great number of people, thanks to whom my graduate study is a valuable and unforgettable experience in my life.

I would like to express my deep gratitude for Prof. Xianbin Wang for offering me the opportunity to study in the University of Western Ontario. His enlightening supervision and foresight motivated me to identify the cutting-edge research topics, develop in-depth ideas and achieve efficient research progress. His kind patience also helps my professional development through exploration of different research methods. The research and communication ability I learned from him will significantly benefit my future work and life.

I would also thank my colleagues, Dr. Aydin Behnad and Dr. Guanghui Song, the post doctoral fellows in the research group for their warm-hearted help with some critical mathematical and technical details in my research. I gained lots of technical knowledge by discussing with them frequently.

Additionally, many thanks to my research group, friends from other groups, faculty and staff members of University of Western Ontario who have helped me with my study or life so that I was able to overcome various difficulties and finally completed this thesis.

Eventually I would like to demonstrate my special thanks to my beloved father and mother. They will always support me with any physical and spiritual emotional support they can provide.

Contents

List of Figures

List of Tables

List of Appendices

Abbreviations

5G	*The 5-th Generation Network*
AF	*Amplify-and-Forward*
AWGN	*Additive White Gaussian Noise*
BS	*Base Station*
CNAR	*Collaborative NOMA Assisted Relaying System*
D2D	*Device-to-Device*
DF	*Decode-and-Forward*
IoT	*Internet of Things*
MIMO	*Multi-Input Multi-Output*
mmWave	*millimeter-wave*
MRC	*Maximal Ratio Combing*
MT	*Mobile Terminal*
MU-MIMO	*Multi-User Multi-Input Multi-Output*
NOMA	*Non-orthogonal Multiple Access*
OFDMA	*Orthogonal Frequency Division Multiple Access*
OMA	*Orthogonal Multiple Access*
PSNR	*Post-processing Signal-to-Noise Ratio*
R-D	*Relay-Destination*
RF	*Radio Frequency*
S-R	*Source-Relay*
SIC	*Successive Interference Cancellation*

SNR *Signal-to-Noise Ratio*

SU-MIMO *Single-User Multi-Input Multi-Output*

TDMA *Time Division Multiple Access*

Chapter 1

Introduction

1.1 Background of 5G

The future 5-th generation (5G) network is expected to provide high-performance communications for rapidly increased devices at anytime and anywhere [1]. On one hand, a large number of smart devices, including smart phones, tablets and laptops for purposes of daily work and entertainments, require remarkably higher data rate, along with enhanced cell-edge rate [2]. The reason is according to [3], it is predicted that in 2020 the data traffic from software downloading, social networking, web browsing, file sharing and multi-media streaming will be significantly enhanced where the high-definition video traffic, with intensive requirements for data rate and real-time playing, is going to be 13 times over that in 2014. On the other hand, the Internet of Things (IoT) [4] would make various sensors, wearable devices, household appliance and even vehicles connected to the core network in order to establish the ”smart life” for personal health, economy and convenience. For industrial consideration, the monitor center should connect to multiple machines and sensors with very low latency to col-

lect real-time information and control pipes, valves and grids [5]. The data volume associated with each device in IoT don't have to be very large since there are only simple data and control message exchange. But a significant number of devices need to be connected simultaneously. Consequently, the expected overall network capacity enhancement will be 1,000 folds over current situation [6][7]. Compared to the extremely high requirements, the currently available spectrum resource becomes quite limited.

The newly developed millimeter-wave (mmWave) frequency bands may potentially help support the network traffic enhancement. Firstly, the short-distance line-of-sight (LOS) directional communications can be well supported by the 60 GHz mmWave band for very high data rate [8]. Furthermore, it is possible to obtain larger channel bandwidth, around 500 MHz per channel in mmWave compared to 5-20 MHz in mainstream microwave bands. Additionally, the small wavelength facilitates installing massive antennas into the mmWave transceivers [9]. However, the drawbacks from mmWave remarkably undermine the potential performance improvement. The most critical point is that the signal in mmWave bands is highly sensitive to blockages which lead to severe penetration [10]. This creates severe signal strength loss in non-line-of-sight (NLOS) environments, very common in large-scale communication systems. Moreover, practical hardware overhead further limits the application of mmWave technology. For instance, the mixed signal components in mmWave transceivers bring higher costs and energy consumption than microwave transceivers [9]. Therefore, it is hard for mmWave spectral resource to contribute significantly to serving massive users and increasing network capacity.

All the frequency bands in the currently available spectrum have almost been developed for telecommunications in mainstream protocols. The bands for LTE-based communication [11] occupy a large portion of available spectrum. The CDMA and WCDMA protocols [12] are

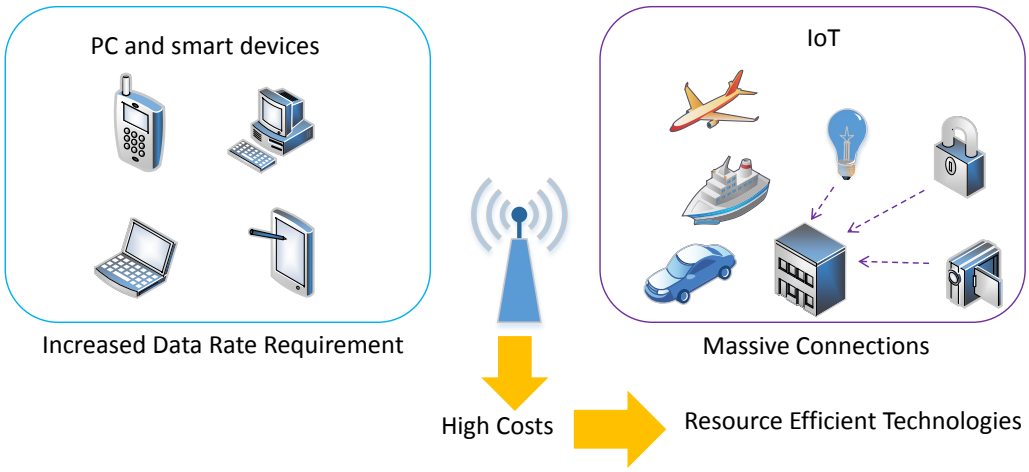

Figure 1.1: 5G use cases and corresponding requirements.

still using some bands for transitions from 3G to 4G in some undeveloped areas. Moreover, the 2.4 GHz frequency band has been widely used by Wi-Fi and Bluetooth protocols. As a result, novel highly efficient communication techniques are in need of the hour for exploiting the current resources.

1.2 Research Motivations

1.2.1 Advantages of Spectral and Power Efficient Technologies

Non-orthogonal multiple access (NOMA) is a 5G promising spectral efficiency technology applicable in the current spectrum resource [13]. Conventional orthogonal multiple access (OMA), e.g., time division multiple access (TDMA) [14] and orthogonal frequency division multiple access (OFDMA) [15], allocates orthogonal resource blocks for different users to avoid inter-user interference. Instead, NOMA introduces inter-user interference at the network side. To be specific, messages intended for several users are superimposed as a signal to be

transmitted in one resource block where messages are allocated with different power levels. When receiving the NOMA signal, certain user applies successive interference cancellation (SIC) to remove the interference from other messages in order to obtain the message for its own. Note that the power-domain multiplexing in NOMA allows a great number of messages superposed into one signal. Hence, NOMA can increase the number of served users per cell. And it is specially suitable for IoT where real-time connections between the control center and massive various devices should be established without specific data rate requirements.

Massive Multi-input Multi-output (MIMO) is another spectral efficient technology for 5G [16]. It can transmit multiple signals at the same time and frequency band by different spatial patterns to multiple receivers. These signals can carry different data for a large group of users or carry the same data for several specific users for performance gain. Additionally, with beamforming technology to preprocess the signal at the transmitter side, the focus of the signal can be narrowed to the targeted receiver so that the quality of the received signal is improved. In this way, each cell can potentially support more users and increase the received signal strength. Hence, the overall system capacity can be greatly enhanced within existing resources.

Relaying technology enables efficient utilization of power [17] to serve the increasing cell-edge users. The better relay-destination (R-D) channel condition, which results from the short R-D distance, leads to higher cell-edge user receive signal-to-noise ratio (SNR) than the direct source-destination (S-D) transmission with the same transmit power. Moreover, device-to-device (D2D) communication ability at smart devices enables some devices to be applied as the relays [18], which saves the cost of deploying the relays. The relaying technology can also enable the cooperation among cells if the relaying device is shared by several cells. Hence, relaying devices can take advantage of high power efficiency to improve communication quality

for all users [19].

1.2.2 Challenges for Resource Allocation

There are challenges to utilize NOMA, massive MIMO and relaying techniques and develop relevant resource allocation strategies for overwhelming 5G system performance.

It is important to consider user scheduling and power allocation to cost-efficiently take advantage of NOMA protocol. Firstly, the dependence on SIC causes NOMA to be significantly sensitive to channel quality. Due to channel condition variation, the non-orthogonality becomes dynamic. Higher-level non-orthogonality leads to more contribution to improving system performance, but with higher costs, i.e., higher requirements on SIC and worse stability. So to control the level non-orthogonality, users should be scheduled into appropriate channels and frequency bands. In particular, to seek a tradeoff between resource sharing and single-user communication quality, the number of users sharing the same resource block and the relevant user features should be decided. Note that NOMA protocol is not well applicable to certain users associated with long distances, very poor channel conditions and very high target rates. The reason is if allocating too much resources to satisfy these users' requirements can cause resource insufficiency for other users, which will downgrade the overall system performance. Additionally, power level allocated to each message for the relevant NOMA user is supposed to be well considered. This is because NOMA is a power-domain multiplexing strategy; the power allocation influences the successful execution of SIC and whether single-user data rate requirements will be met.

Though the system capacity is greatly enhanced by massive MIMO, it brings about high

hardware and system complexity costs as well. The very large antenna array is partially responsible for hardware costs. Moreover, for signal processing, each antenna needs to be connected to a radio frequency (RF) chain, which introduces complicated hardware installation and processing overhead. Furthermore, in the transitional stage from the 4-th generation network to 5G, the number of RF chains cannot match massive candidate antenna in the large array, making it impractical to use all antennas in each time slot. Hence, selecting "good" antennas to match the existing RF chains exactly is a promising solution [20]. Moreover, each large-scale antenna array has the upper-bound of capability to serve users; in certain time slot, serving some users with extremely poor channel conditions will cost very much resource but obtain only limited performance. So similarly, it is important to schedule "good" users for communication in each time slot for cost-efficient resource utilization.

The relaying scheme should also be investigated to be adaptive to 5G in order to serve multiple cell-interior and cell-edge users. Currently, a large number of smart devices have been equipped with double antennas, and some of them can be applied as relaying devices to enable D2D communications. Moreover, unlicensed band is going to be exploited in 5G [21]. The resource efficiency in conventional relaying systems are not high enough since the relaying link requires one orthogonal resource block to avoid the interference. With extra unlicensed band resource provided, the communication and resource efficiency can be possibly promoted. Following these conditions, suitable relaying scheme for 5G with the idea of data offloading [22] is worth considering. This is to improve the system throughput in order to guarantee the quality of service for both the relaying devices (cell-interior users) and cell-edge users.

It is also inevitable to consider the integration of several 5G promising technologies. Firstly, the evolution of the communication system is usually comprehensive. For instance, CDMA

communication protocol [12] and rake receiver to counter the multipath fading are revolutionary technologies for highlighting the advancement of the 3G system over 2G. Then the 4G breakthrough is featured by even more techniques, involving OFDMA protocol [15] to improve the data rate, Voice over LTE (VoLTE) protocol [23] to transmit voice in pure data and the true packet switched architecture to replace the circuit switched one. So 5G is expected to be a revolution on communication system highlighted by multiple technologies. Furthermore, potentially compatible promising techniques for 5G are worth considering to improve the system performance as much as possible. Moreover, some techniques are complementary to each other. Massive MIMO is able to support multiple receivers or enhance the receive SNR at certain receivers. NOMA can be applied to establish connections to a great number of users simultaneously. The relaying technology can assist in transmitting signals to some users which cannot have been supported by the BS. The features of these techniques create opportunities for them to improve the performance for each other, contributing to better overall performance in the 5G communication system.

Possible integration can be based on performance improvement of certain technology. The drawback of NOMA is the communication quality for single user. In contrast to conventional orthogonal multiple access (OMA) where the power in certain time slot [14] or frequency band [15] is allocated to only one message, NOMA should allocate the same amount of power to several messages, which undermines the single-user data rate. Hence, it is necessary for NOMA to cooperate with other techniques for achieving better communication quality. Massive MIMO is able to provide received SNR gain by large-scale transmit antenna array. The spectral efficiency of NOMA can be further developed by the relaying technology to support multiple cell-edge users. In this way, it is natural to consider the design of NOMA-based mas-

sive MIMO and relaying systems. Furthermore, to maximize the performance of these systems within limited resources, it is important to consider relevant resource allocation strategies.

1.3 Research Objectives

Each research paper in the literature investigates only one promising technology in the future 5G on the preliminary implementation and performance. The objectives of this thesis are to conduct integration of potential efficient techniques for 5G, i.e., massive MIMO, NOMA protocol, relaying technology, and, on the basis of systems formed by integrated techniques, to propose resource allocation mechanisms in order to serve a great number of users simultaneously with high-level communication quality.

To exploit extremely high spectral efficiency for serving multiple users with good communication quality for the 5G networks, the first objective is to integrate massive MIMO and NOMA techniques into the 5G communication system. To take advantage of NOMA protocol, which is featured by power-domain multiplexing, to better the system performance, the power allocation among NOMA users should also be specified in this designed massive-MIMO NOMA system.

Next, in massive MIMO, limited RF chains make it impractical to use antenna all elements in the MIMO antenna array; Additionally, the performance of NOMA protocol significantly depends on channel conditions. Consequently, the following objective is to figure out antenna selection and NOMA user scheduling algorithms to maximize user sum rate and control the non-orthogonality between NOMA users prior to signal processing. In the massive MIMO-NOMA system, specific algorithms are proposed to solve the problem in single-subband and

multiple-subband scenarios, respectively. Moreover, to make the algorithms executable in the system, the trade-off between system performance and the computational complexity of the algorithms should be achieved.

Furthermore, as the proliferation of 5G smart devices, a large number of which have the relaying ability and double antennas, causes more cell-edge users, it is important to use NOMA to assist relaying systems to serve multiple cell-edge users. Thus, the next objective is to propose the NOMA assisted relaying scheme based on the new features of smart devices and newly developed unlicensed band resource to guarantee the single-user data rate. For achieving the minimal outage probability for the best effect of data rate guarantee, the NOMA-related power allocation scheme should be determined. For further system performance characterization and evaluation for the power allocation scheme, the system ergodic capacity based on this power allocation scheme needs to be analyzed.

1.4 Contributions

The main contributions of this thesis are listed as follows:

- A general review of advantages and relevant problems brought by NOMA, massive MIMO and relaying techniques in 5G are provided. A literature survey is done for: power allocation, user pairing, relaying scheme under NOMA protocol; antenna selection, user scheduling for MIMO system; performance analysis in NOMA-based relaying system.

- A massive MIMO-NOMA system is designed to improve the communication quality for NOMA users. Then focused on the single-subband situation, the power allocation

among NOMA users to maximize the sum rate is provided as the foundation for further investigation.

- To achieve efficient resource allocation in massive MIMO-NOMA system, the antenna selection and user scheduling mechanisms are investigated. According to different scenarios, the efficient antenna selection and user scheduling algorithms are proposed. In single-band scenario, the antenna selection problem is solved by efficient search algorithm, which achieves the search efficiency by limiting the candidate antennas into ones beneficial to relevant users. For joint antenna selection and user scheduling in multi-band multi-user scenario, joint AU contribution algorithm are raised by selecting the antennas and users with the highest contribution to the total channel gain. Simulation results demonstrate that proposed antenna selection algorithm achieves near-optimal performance, and joint AU contribution algorithm achieves similar performance to existing methods with reduced complexity. The proposed algorithms control the orthogonality among NOMA users in a high-level without losing stability.

- A Collaborative NOMA Assisted Relaying (CNAR) system is proposed with the collaboration of S-R NOMA link as macro-cell communication and R-D NOMA link as small-cell communication. The relay is executed in full-duplex way to improve the system throughput. Moreover, the S-R and R-D phases are executed in licensed and unlicensed band, respectively, to avoid the interference. Then with outage probability derived, the power allocation ratios are obtained by minimizing the outage probability. Then capacity analysis in high SNR regime is also provided to further characterize the system performance. It is achieved by analyzing the interference at cell-edge users based on

NOMA protocol. Simulation results validate our mathematical analysis, and show that the relaying system assisted by NOMA achieves lower outage probability and higher sum capacity than orthogonal multiple access (OMA). The proposed system develops the spectral efficiency of NOMA protocol to serve multiple cell-edge users concurrently with high data rate.

1.5 Thesis Outline

This outline of this thesis is as follows:

Chapter 2 investigates certain resource efficient technologies in 5G, i.e., NOMA, massive MIMO, as well as relaying, their relevant problems and literature. Firstly, the critical points of these technologies to achieve high resource efficiency are given. Then the problems brought by these technologies are explained. The problem involves user scheduling, antenna selection, power allocation and relaying mechanism design. Next, a study on the existing methods to solve the problems is discussed.

Chapter 3 demonstrates the proposed methods to solve antenna selection and user scheduling problems. The optimal power allocation scheme is provided for clarifying the standard of antenna selection and user scheduling. Then efficient search algorithm is proposed to solve antenna selection problem in single-band two user scenario. Next, Joint AU contribution algorithm is proposed for joint antenna selection and user scheduling in multi-band multi-user scenario. Simulation results are provided for validating the tradeoff between performance and complexity of our proposed algorithms.

Chapter 4 studies the solution to serving multiple cell-edge users concurrently. In the sys-

tem model, the relaying mechanism assisted by NOMA is proposed. In the next section, to characterize the system, outage probability analysis is given with mathematical insights, with the help of which the optimal power allocation method is also proposed. Following this, the sum capacity approximation is provided by analyzing the ergodic capacity for each user. Simulation results of the proposed mechanism are compared with the relaying scheme assisted by conventional OMA.

Chapter 5 summarizes the ideas, analysis and results from this thesis and discusses the future potential research work.

Chapter 2

Technologies for Efficient Utilization of Resources in 5G

In this chapter, the technical aspect of the future 5G networks, following problems and the relevant literature survey for existing solutions will be provided. The first section introduces resource efficient technologies for the future 5G, i.e., NOMA, massive MIMO, relaying mechanism, and describes their principles to achieve resource efficiency. Following this, the problems brought by these technologies are discussed, which involves antenna selection, user scheduling, power allocation and relaying scheme design. Lastly, the chapter provides currently available algorithms and approaches for these problems.

2.1 Principles of Resource-Efficient technologies in 5G

The future 5G are confronted with an explosive device proliferation and greatly enhanced data traffic. To satisfy these demands within relatively constrained resources and tolerably increased

costs, resource efficient strategies, involving NOMA, massive MIMO and the relaying technology, is worth exploring.

2.1.1 NOMA

Drawbacks of Conventional OMA

Conventional OMA is able to service multiple users but some drawbacks downgrade the system performance, one of which is serious interference management overhead at the network side. To serve multiple users at the same time, the network side uses OMA protocol by dividing the entire resource into orthogonal resource blocks in time-domain (TDMA) [14] or frequency-domain (OFDMA) [15] to avoid inter-user interference. But if the constrained overall resource is divided into multiple blocks for a great number of users, the resource volume per block will be even more limited. The granularity of single resource block also has its lower bound. Moreover, it is impractical if the network side undertakes all the interference management tasks since the backhaul and feedback issues cost high overhead [24]. Thus, it would be better to transfer some interference management tasks to the receiver side.

The other drawback for conventional OMA is the limitation on resource sharing. Whether certain user reaches its target rate can be a standard for communication quality. For some users with good channel conditions, the target rate can be easily reached so the resource can be wasted to some extent. For users with poor channel conditions, we may need to allocate an orthogonal block with large resource amount to guarantee the data rate [25][26], which results in relatively low resource efficiency [27]. The phenomenon above will cause the overall system performance degraded. Additionally, with the development of IoT, the BS or some

control center has to establish massive real-time connections to a variety of smart devices with different channel conditions [7][28]. Hence, it's a good strategy if certain users with good channel conditions can share one resource block with ones with poor channel conditions.

NOMA Advantages over OMA and NOMA Principle

NOMA protocol is an advantageous strategy for user side interference management and resource sharing. NOMA allows several messages multiplexed at the same time and frequency. At the transmitter side by superposition coding, several messages are superimposed with different allocated power level as a NOMA signal. Generally, more power is allocated to messages of users with poorer channel conditions as the compensation, which also guarantees these users' receive SNR in a comparable level to conventional OMA. At the receiver side, the users with poor channel conditions treat the messages for other users and the environmental noise as the whole noise for message decoding. The users with better channel conditions apply SIC for message decoding. To be specific, they decode the messages for users with poorer channel conditions in a similar manner. This decoding process is most likely successful due to better channel conditions. Then the decoded messages are removed so that these uses are confronted with less inter-user interference when they decode the messages for their own, which also compensates for the less allocated power to these message. In other words, the general power allocation idea achieves some user fairness, which can be further improved with accurate value.

Therefore, without considering the data rate with respect to each signal, we can support a large number of users using NOMA, theoretically. The supporting ability is not restricted by the number or the granularity of the resource blocks as it is in conventional OMA since NOMA is a power-domain multiplexing strategy. To elaborate the NOMA principle, we assume one

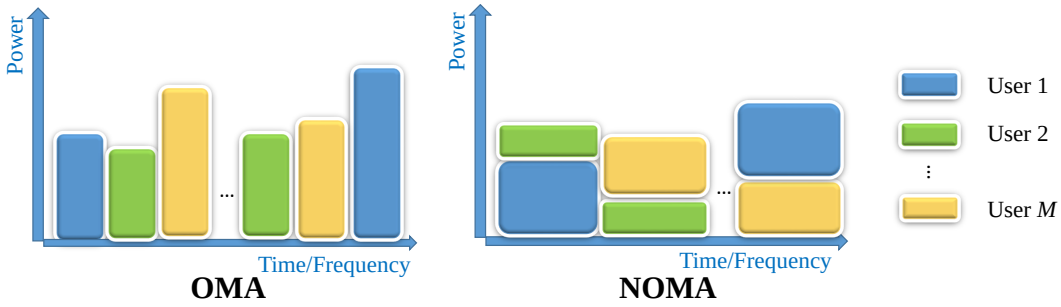

Figure 2.1: Protocol difference between OMA and NOMA.

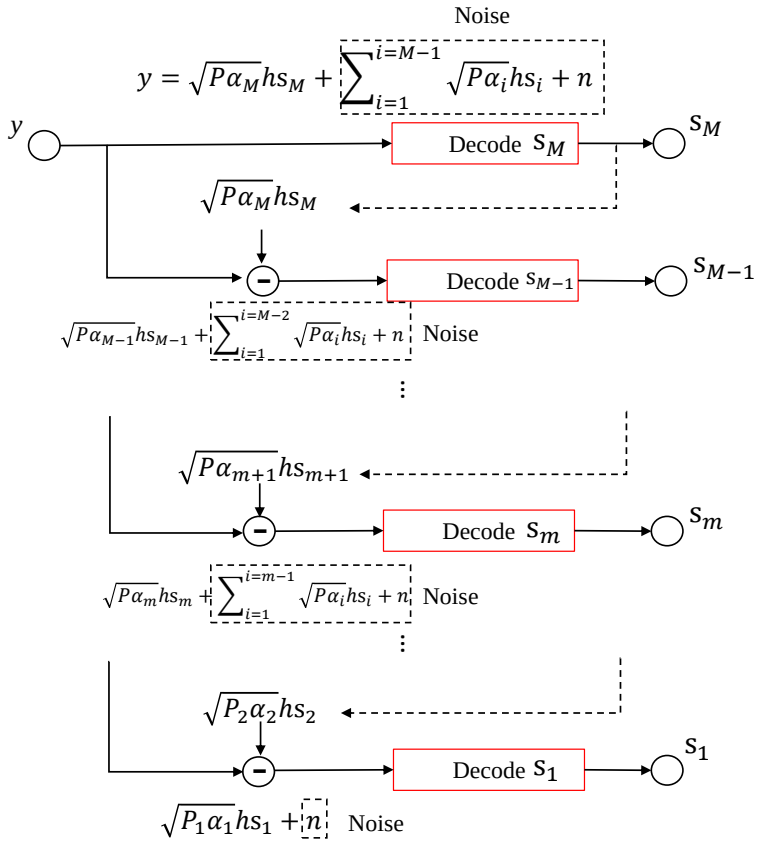

Figure 2.2: General case of SIC mechanism for NOMA.

transmitter sends a NOMA signal superimposed by messages to M users. The norms of channel coefficients from the transmitter to the 1-st, ..., m-th, ..., M-th receivers are ordered as $|h_1| > ... > |h_m| > ... > |h_M|$. The superimposed NOMA signal is given by

$$x = P(\sqrt{\alpha_1}s_1 + ... + \sqrt{\alpha_m}s_m + ... + \sqrt{\alpha_M}s_M) \tag{2.1}$$

where the power allocation ratios are ordered as $\alpha_1 < ... < \alpha_m < ... < \alpha_M$ to compensate for the poor channel conditions. The receive signal at the m-th user becomes

$$y = P|h_m| \sum_{i=1}^{M} \sqrt{\alpha_i} + n_m \tag{2.2}$$

The M-th user treat messages for other $(M - 1)$ users and the environmental noise as the equivalent noise to decode the message for itself. So its post-processing SNR (PSNR) is given by

$$SNR_M = \frac{\rho|h_M|^2\alpha_M}{\rho|h_M|^2 \sum_{i=1}^{M-1} \alpha_i + 1} \tag{2.3}$$

For the m-th user, due to the ordered channel gains, it can most likely decode messages from $(m + 1)$-th to M-th user successfully. Based on SIC, these messages are removed from the original NOMA signal received by the m-th user. Then this user decodes its own message by treating the remaining messages and environmental noise as the equivalent noise, through which the PSNR becomes

$$SNR_m = \frac{\rho|h_m|^2\alpha_m}{\rho|h_m|^2 \sum_{i=1}^{m-1} \alpha_i + 1} \tag{2.4}$$

Specially, the 1-st user only has to consider the environmental noise for its PSNR, resulting

in

$$SNR_1 = \rho|h_1|^2\alpha_1. \tag{2.5}$$

The above decoding process is illustrated in Fig. 2.2 for clearance.

NOMA Performance in Two-user Case

In many situations, there is minimum rate requirement for single user, for which only two user share one resource block under NOMA protocol. In this typical case, users with better channel conditions are denoted as near user (namely the 1-st user in the general case), while the other as far users (namely the 2-nd user when $M = 2$ in the general case). For message decoding, far user considers the near user message and environmental noise as the whole noise to decode the message for its own. According to the aforementioned power allocation idea, the power allocated to the message for far user is much higher than it is for near user to compensate for poor channel conditions. Next, near user uses SIC to decode and then remove the message for far user, based on which it decode its own message with only the environmental noise as the noise.

We firstly characterize NOMA performance in this simple two-user case. Fig. 2.3 demonstrates the achievable capacity as a function of transmit SNR. The power ratio allocated for far user is 0.75 to compensate for the poorer channel condition. One can observe that for far user, as transmit SNR increases, the achievable capacity gets close to a certain constant. This is because far user treats the message for near user as part of the noise. When the transmit power grows, the growth of noise power has similar level as that of useful signal power. In other words, in Eq. (2.3) when $M = 2$, $\rho \to \infty$ makes $SNR_M \to \frac{\alpha_2}{\alpha_1}$. For near user, its achievable

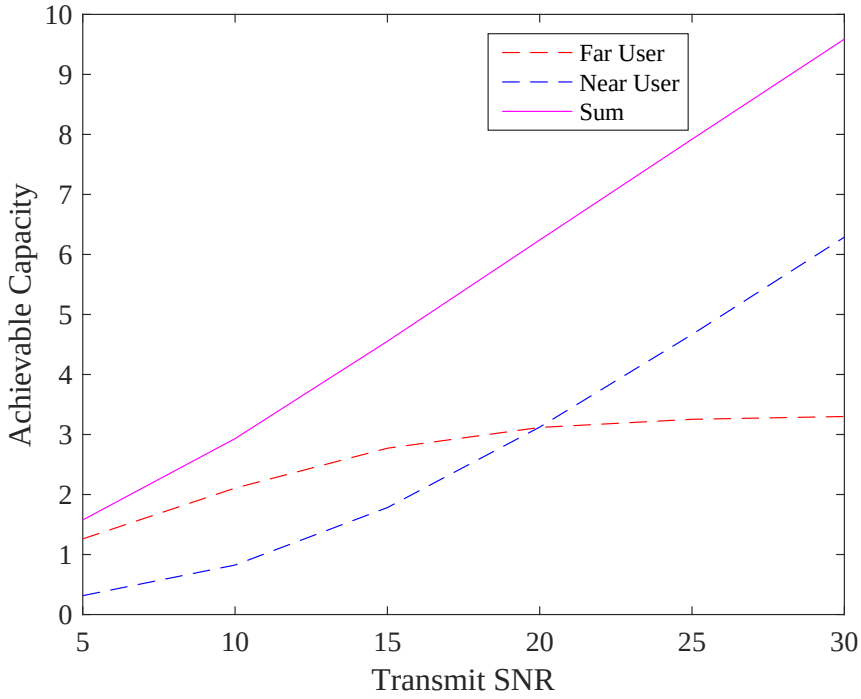

Figure 2.3: Achievable capacity under NOMA protocol.

capacity nearly grows linearly with transmit SNR. The reason is due to SIC, the noise for its message decoding is only the environmental noise, the power of which doesn't become larger as transmit power increases. If we consider it in Eq. (2.5), SNR_1 increase linearly with ρ. In this way, since the capacity for far user and near user both increase monotonically with transmit SNR, the sum achievable capacity follows the same tendency.

Then we extend the above analysis to the general case where M users share one NOMA signal. For the 2-nd to the M-th user, as each of them need to consider other users' messages as the noise, the achievable capacity is upper-bounded by certain constant. To be specific, according to Eq. (2.4), as $\rho \to \infty$, SNR_m is asymptotically equivalent to $\frac{\alpha_m}{\sum_{i=1}^{m-1} \alpha_i}$, which is the upper-bound of the m-th user PSNR. This also indicates that at the user side, the power ratio among relevant NOMA users can be maintained, which is expressed as PSNR. For the

1-st user, corresponding to the near user in Fig. 2.3, its lower allocated power ratio can be compensated by the receive SNR which can increase unlimitedly. Therefore, NOMA protocol takes advantage of the user with the best channel condition to increase the sum rate and ensures good levels of achievable rate for other users.

2.1.2 Massive MIMO

Though the applicable spectrum is slight broaden by the unlicensed band in the future 5G, the existing resource is still far from being sufficient for the expected 1,000 fold capacity increase. The massive MIMO technology, developed from conventional MIMO, is a spectral efficient strategy to improve the 5G system performance. It is featured by multiple transmitter antennas and multiple receive antennas to enable multiple signal inputs and outputs so that the system throughput is significantly enhanced. At the receiver side, there can be different forms, i.e., a large receive antenna array corresponding to single-user multi-Input multi-Output (SU-MIMO) and multiple devices equipped with single antenna corresponding to Multi-User Multi-Input Multi-Output (MU-MIMO) [29].

For conventional MIMO, one reason to achieve spectral efficiency is the spatial domain multiplexing. To be concrete, by multiple antennas, different signals are transmitted and received in different spatial patterns at the same time and frequency band. The benefits from spatial multiplexing can be considered in a system level and at the receiver. In a system level, spatial multiplexing enables MIMO system to achieve a great degree of freedom gain with desired channel conditions. For instance, in high SNR regime, the system capacity is proportional to $\{\min(L_T, L_R)\log(SNR)\}$ [30]. The benefits from spatial multiplexing at the receiver

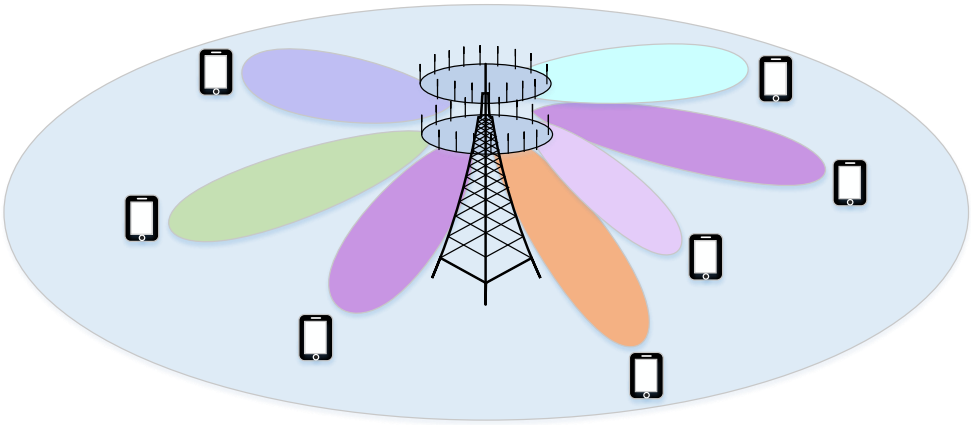

Figure 2.4: Depiction of massive MIMO system.

side can be the diversity gain [31], which results from the following situation. If L_R antennas at user side receive the same signal from one transmit antenna, we receive L_R copies of the same signal corresponding to L_R different SNR levels. There are some mechanisms at the receiver side to obtain the diversity gain. According to the Selection Combining [32], it's easy to select the maximum SNR from all the copies as the output SNR. However, the output can be further improved if we make full use of all the L_R copies. Hence, if we use receive-maximal ratio combining (MRC) mechanism [33], the output SNR will be the summation of receive SNR at all L_R receive antennas. Similarly, from multiple transmit antennas sending the same signal, the user device with single receive antenna can also take advantage all different transmission paths to improve receive SNR if transmit-MRC mechanism is applied [34].

Another reason for high spectral efficiency in MIMO system is beamforming that signals intended for a group of closely positioned users are modulated and transmitted in a specific angel targeting at the user group to improve the transmission accuracy [35]. In other words, the allocated power can be almost concentrated at that angle associated with targeted user group so that little leaked power cannot cause seriously interference to other group of users.

Massive MIMO is also named as very large MIMO, full-dimension MIMO, hyper MIMO and the large-scale antenna system. It reaps all the advantages from conventional MIMO but on a very large scale by massive candidate antennas at the BS side and lots of receive antennas at the user side [36]. Additional benefits can be obtained in massive MIMO. For instance, from a great number of candidate antennas, it is more possible to find certain antennas with extremely good channel conditions, with which the transmit power is saved and system efficiency is increased; the wireless network coverage can also be expanded to communicate with much more users. Additionally, in MU-MIMO cases, massive transmit antennas at the BS contribute to inter-user interference diminishment, which is achieved by the asymptotical orthogonality among users if linear matched filter downlink precoding is applied at the BS side [37].

2.1.3 Relaying Technology

The relaying mechanism is widely used to improve the receive signal strength at cell-edge users, which can be further exploited in 5G. Firstly, because of D2D connection ability in smart devices, they can execute the relaying functions, which lower the load of the ISP without having to establish additional infrastructure [7]. Furthermore, it is consistent with one of the intentions by 5G, i.e., to shrink the cell size. Specifically, the size of macro cell can be shrunk to the extent which exactly guarantees transmissions to relaying devices; one relaying device is in charge of a small cell to improve the communication quality at users far from the BS. Smaller cell size can increase the energy efficiency by short-range transmissions and easier power concentration. Lastly, the newly developed high-frequency unlicensed band resource in 5G well matches WiFi protocol for short range D2D communications and data offloading [11].

2.2 Challenges of Utilizing NOMA, MIMO and Relays

2.2.1 User Scheduling and Power Allocation in NOMA

In NOMA protocol, to guarantee the quality of service to each user, generally two users are paired to share the same NOMA signal [38]. For both user to receive good-level service, it's worth considering which two users among all the users should be paired. Moreover, similar to antenna side, in each time slot some users, if involved in communications, will get low quality of service and cause low energy efficiency due to extremely poor channel conditions. If they are forced paired with some users to execute NOMA protocol, they will affect normal communications of other users; what's more, the system performance, e.g., the sum rate, will be even more downgraded. In this extreme case, those users will be given up temporarily for cost-effective usage of system resources. This is fair since they can be scheduled in other time slots when channel conditions are better [39].

With NOMA integrated into massive MIMO system, to improve the system performance, it is important to select antennas and schedule users jointly as it is in conventional MIMO system [40][41]. The reason is if we select antennas and users separately, e.g., select the antenna subset first, some of the selected antenna may not be the best matches for users to be selected, which may also influence user pairing strategy and performance. Hence, the joint selection of antennas and users can lead to the best matched selected antennas and users.

Apart from user scheduling, power allocation is another step for NOMA system performance improvement since NOMA is a power-domain multiplexing strategy. The objective of power allocation can be maximization of sum rate or achieving the fairness with regard to data rate guarantee among NOMA users [42].

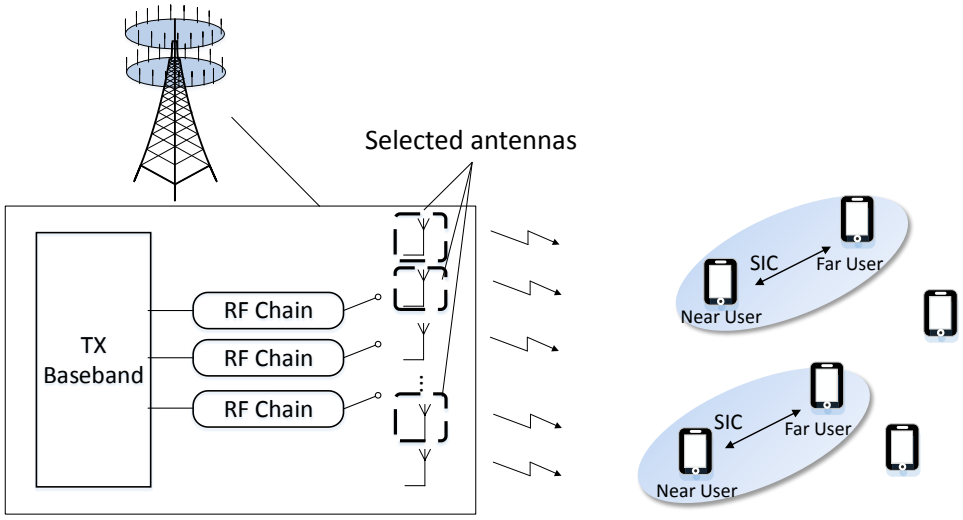

Figure 2.5: Illustration of antenna selection and user scheduling.

2.2.2 Antenna Selection and User Scheduling in Massive MIMO

The efficiency and diversity gain in massive MIMO is obtained at the expense of very high computational complexity and hardware cost. In particular, each antenna used for data transmission needs to be connected to one RF chain for signal processing. Each RF chain in transmitter side consists of up-converter, power amplifier, filters and a digital-to-analog converter. At the receiver side, each RF chain contains down-convertor, a low noise amplifier and analog-to-digital convertor. Typical massive MIMO antenna array size can be 8×8, 16×16; it would be even larger in the future [16]. Confronted with large number of users, if massive antennas are all taken into use, the hardware costs will increase significantly. Potential problems will be extended to system configuration and peripheral maintenance. Additionally, the increased complexity of encoding and decoding spatial-time codes cannot be ignored. The above problems will impact real-time message transmission.

Antenna selection is a cost-effective solution for the tradeoff between system performance

and costs. In each time slot, there are some antennas corresponding to better performance than others. So the basic idea is to select L_T "best" antennas from all M_T candidate antennas based on some criteria so that the number of required RF chains is kept as L_T. Those unselected antennas contribute moderately to system performance while cost extra complexity. Hence, performance can still be maintained in high-level and hardware complexity can be reduced.

2.2.3 NOMA Assisted Relaying System Design

NOMA protocol can increase the efficiency of relaying mechanism. Conventional relay only accepts single message input and output single message to cell-edge users. Instead, by superposition coding, NOMA provides multiple inputs and multiple outputs for the relay. Moreover, in the further 5G, smart device with D2D communication ability can be applied as a relay. With NOMA assisted, the relay can decode the message from source NOMA signal for itself and transmit the remaining part to multiple cell-edge users. This strategy saves the the deployment of relays and improve the system spectral efficiency.

However, some technical issues are supposed to be specified for the NOMA assisted relaying system. To provide high throughput to support multiple users simultaneously, the relaying scheme of the system need further investigation. Moreover, the power allocation method among NOMA-based messages should be figured out to provide basic data rate guarantee for single user.

2.3 Considerations on NOMA User Pairing

NOMA user pairing strategy can be investigated into two scenarios, based on fixed allocated power and target rate guarantee, respectively. The scenario with fixed allocated power is less complex since power ratios don't need dynamically adjusting. For another scenario, user pairing and real-time power allocation are jointly considered to satisfy the user target rate [43].

2.3.1 User Pairing with Fixed Allocated Power

In this case, we focus on the user pairing issue with predefined power ratios. Following the assumptions in Eq. (2.1), we assume the m-th and the n-th user conditioned on $m > n$ form a user pair with $\alpha_m^2 + \alpha_n^2 = 1$. Due to $|h_m|^2 < |h_n|^2$, based on NOMA principle, it holds that $\alpha_m > \alpha_n$. So the achievable rate associated with these two users are

$$R_m = \log(1 + \frac{\alpha_m |h_m|^2}{\alpha_n |h_m|^2 + \frac{1}{\rho}}) \tag{2.6}$$

and

$$R_n = \log(1 + \rho \alpha_n |h_n|^2). \tag{2.7}$$

For the conventional OMA scheme, the achievable data rate associated with i-th user becomes

$$\overline{R}_i = \log(1 + \rho |h_i|^2) \tag{2.8}$$

where $i \in \{m, n\}$. So the sum rate gain of NOMA over OMA scheme is given by

$$\begin{aligned} R_m + R_n - \overline{R}_m - \overline{R}_n &\overset{\rho\to\infty}{\to} \log(\frac{1}{\alpha_n}) + \log(\rho\alpha_n|h_n|^2) - \log(\rho|h_m||h_n|) \\ &= \log|h_m| - \log|h_n|, \end{aligned} \tag{2.9}$$

in very high SNR area. So more performance gain will be obtained if there is a lot difference between the paired users' channel gains.

2.3.2 User Pairing with Consideration of Target Rate

With the consideration of target rate, NOMA is executed under some condition. For some user with weak connections to the transmitter in certain subband, satisfying the target rate is the priority. If the transmit power is more than enough for that user to reach the target rate, then the relevant subband can admit another user with strong connections to share the power. Otherwise, the whole subband will be allocated to the weak user to compensate for the poor channel quality as much as possible or we can schedule it in other time slot when its corresponding channel conditions are better.

Following the assumption in Eq. (2.1) again, we focus on the m-th user as the one with poor connection in the relevant subband and we regard the n-th user as the potential user to be admitted into the same subband. Assume the target rate is $R_0 = \log(1 + \gamma_0)$ for the m-th user, then to satisfy the target rate, the following inequation should be met

$$\frac{\alpha_m|h_m|^2}{\alpha_n|h_m|^2 + \frac{1}{\rho}} \geqslant \gamma_0. \tag{2.10}$$

This indicates that the maximal power allocation ratio to n-th user becomes

$$\alpha_n = \begin{cases} 0, & \text{for } |h_m|^2 \leqslant \frac{\gamma_0}{\rho} \\ \frac{|h_m|^2 - \frac{\gamma_0}{\rho}}{|h_m|^2(1+\gamma_0)}, & \text{for } |h_m|^2 > \frac{\gamma_0}{\rho} \end{cases} \tag{2.11}$$

When $|h_m|^2 \leqslant \frac{\gamma_0}{\rho}$, it means the target rate for the m-th user cannot even be reached. So a general pairing strategy is when $|h_m|^2 > \frac{\gamma_0}{\rho}$, the m-th user can be paired with the n-th user. Otherwise, we can consider allocate the entire power of relevant subband to the m-th user or not schedule it for communication in the time slot since weak connection will reduce the energy efficiency.

Conditioned on the situation where the target rate for the m-th user can be reached, the performance gain of NOMA over allocating all power to the m-th user is given by

$$\begin{aligned} R_m + R_n - \widetilde{R}_m &= \log(1 + \frac{\alpha_m |h_m|^2}{\alpha_n |h_m|^2 + \frac{1}{\rho}}) + \log(1 + \rho\alpha_n |h_n|^2) - \log(1 + \rho |h_m|^2) \\ &= \log \frac{1 + \rho\alpha_n |h_n|^2}{1 + \rho\alpha_n |h_m|^2} \end{aligned} \tag{2.12}$$

2.4 Considerations on NOMA Power Allocation

Since NOMA is a power-domain multiplexing strategy, power allocation strategy is extremely important for the efficiency of using NOMA and user experience. User sum rate maximization and fairness are mainstream objectives in optimization problems for resource allocation [44][45][46]. The problem formulations of these two objectives are provided as follows where we assume that data rate can reach the achievable rate.

2.4.1 Sum Rate Maximization

Generally in sum rate maximization problems, one critical constraint is the data rate associated with each user should meet the target rate as the communication quality guarantee. So the sum rate maximization problem can be formulated as

$$\begin{aligned} &\max_{\alpha_1,...,\alpha_M} \sum_{i=1}^{M} \log(1 + SNR_i) \\ &s.t. \quad \begin{cases} \sum_{i=1}^{M} \alpha_i = 1 \\ 0 \leqslant \alpha_i \leqslant 1 \quad \forall i, i \in \{1, ..., M\} \\ SNR_i \geqslant \gamma_0 \quad \forall i, i \in \{1, ..., M\} \end{cases} \end{aligned} \tag{2.13}$$

where SNR_i is the PSNR at the i-th user.

2.4.2 Fairness Consideration

Although sum rate maximization leads to system performance maximization when the total power is enough. However, due to higher data rate requirements or user proliferation, each user can experience outage where the data rate doesn't reach the target rate. Thus, user fairness is worth considering for each user to experience as few outage events as possible, which can be formulated as an outage probability minimization problem. Particularly, we define Event A_j happens when

$$\frac{\rho|h_m|^2\alpha_j}{\rho|h_m|^2 \sum_{i=1}^{j-1} \alpha_i + 1} < \gamma_j \quad \forall j, j \in \{m, ..., M\} \tag{2.14}$$

is satisfied [42]. To optimization problem to achieve user fairness can be formulated as

$$\min_{\alpha_1,\ldots,\alpha_M} \max_j \Pr[A_j] \quad s.t. \quad \begin{cases} \sum_{i=1}^{M} \alpha_i = 1 \\ 0 \leqslant \alpha_i \leqslant 1 \end{cases} \tag{2.15}$$

2.5 Antenna Selection Algorithms

2.5.1 Antenna Selection and Use Scheduling Based on Exhaustive Search

Generally, higher channel norm leads to better massive MIMO system performance. To apply channel norm as the criterion for antenna selection and user scheduling, one direct strategy is to exhaustively search possible sub-matrix from the original complete matrix and select the one corresponding to the maximal channel norm. The size of the sub-matrix is the number of required antennas and users. But the computational complexity is extremely high due to direct search of all possible combinations. The complexity can be reduced slightly in NBJTRAS algorithm by transforming matrix computation to vector computation [47]. NBJTRAS involves two stages, i.e., the operation on row and column dimension, of which the steps are given as follows.

Suppose the complete channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$. Note that each receive antenna in the following algorithm can be considered as a single-antenna user device.

Stage 1: The Operation on Row Dimension Define $i_r \in \{1, 2, .., C_{N_R}^{L_R}\}$ as the row combination index, which corresponds to the i_r-th sub-matrix $\mathbf{H}_{i_r} \in \mathbb{C}^{L_R \times N_T}$ by $\mathbf{l}_{i_r} = [l_{i_r}^1\ l_{i_r}^2\ ...\ l_{i_r}^{L_R}]^T$. In

this way, we obtain $\mathbf{H}_{i_r}$ as

$$\mathbf{H}_{i_r} = \begin{bmatrix} \mathbf{h}^T_{l^1_{i_r}} \\ \mathbf{h}^T_{l^2_{i_r}} \\ \vdots \\ \mathbf{h}^T_{l^{L_R}_{i_r}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{i_r}\langle 1,1\rangle & \dots & \mathbf{H}_{i_r}\langle 1,N_T\rangle \\ \mathbf{H}_{i_r}\langle 2,1\rangle & \dots & \mathbf{H}_{i_r}\langle 2,N_T\rangle \\ \vdots & \dots & \vdots \\ \mathbf{H}_{i_r}\langle L_R,1\rangle & \dots & \mathbf{H}_{i_r}\langle L_R,N_T\rangle \end{bmatrix} \tag{2.16}$$

where $\mathbf{h}^T_x$ is the x-th row of $\mathbf{H}$. Then we define $m^x_{i_r}$ to express the magnitude of x-th column in $\mathbf{H}_{i_r}$, which is given by

$$m^x_{i_r} = \sum_{j=1}^{L_R} |\mathbf{H}_{i_r}\langle j,x\rangle|^2, 1 \leqslant x \leqslant N_T. \tag{2.17}$$

Then we obtain the norm vector $\mathbf{m}^T_{i_r}$ as

$$\mathbf{m}^T_{i_r} = [m^1_{i_r}\ m^2_{i_r}\ ...\ m^{N_T}_{i_r}]. \tag{2.18}$$

If we apply the evaluation $\mathbf{m}^T_{i_r}$ to all the $C^{L_R}_{N_R}$ possible combinations, we obtain the norm matrix $\mathbf{M} \in \mathbb{C}^{C^{L_R}_{N_R} \times N_T}$ expressed by

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}^T_1 \\ \mathbf{m}^T_2 \\ \vdots \\ \mathbf{m}^T_{C^{L_R}_{N_R}} \end{bmatrix} = \begin{bmatrix} m^1_1 & m^2_1 & ... & m^{N_T}_1 \\ m^1_2 & m^2_2 & ... & m^{N_T}_2 \\ \vdots & \vdots & \dots & \vdots \\ m^1_{C^{L_R}_{N_R}} & m^2_{C^{L_R}_{N_R}} & ... & m^{N_T}_{C^{L_R}_{N_R}} \end{bmatrix} \tag{2.19}$$

Stage 2: The Operation on Column Dimension We find the largest L_T elements in the r-th row of $\mathbf{M}$ and attain the summation of them, denoted as $m_{\max}^{i_r}$. At the same time, we record the column indices of these L_T elements in the defined index vector $\mathbf{l}_{i_c}(i_r) = [l_{i_c}^1(i_r)\, l_{i_c}^2(i_r) \ldots l_{i_c}^{L_T}(i_r)]^T$. This then yields the max-norm vector

$$\mathbf{m}_{\max}^T = [m_{\max}^1\ m_{\max}^2 \ldots m_{\max}^{C_{N_R}^{L_R}}]. \tag{2.20}$$

Then we find

$$\bar{i}_r = \arg \max_{1 \leqslant i_r \leqslant C_{N_R}^{L_R}} m_{\max}^{i_r} \tag{2.21}$$

Finally, the indices of selected receive and transmit antennas are recorded in $\mathbf{l}_{\bar{i}_r}$ and $\mathbf{l}_{i_c}(\bar{i}_r)$, respectively, according to which we obtain the subset channel matrix $\mathbf{H}_{\text{sub}}$ as the optimal solution.

The complexity of NBJTRAS algorithm is analyzed for further comparison to other algorithms. For Stage 1, in each $\mathbf{H}_{i_r}$, the summation of elements in each column requires L_R computational operations. To deal with N_T columns for each $\mathbf{H}_{i_r}$ matrix and $C_{N_R}^{L_R}$ matrices in total, it needs $L_R N_T C_{N_R}^{L_R}$ operations. For Stage 2, there are $C_{N_R}^{L_R}$ rows in matrix $\mathbf{M}$ and we need to find the largest L_T elements from N_T elements per row with summing these elements up. Then the maximal element should be found from $C_{N_R}^{L_R}$ in vector $\mathbf{m}_{\max}^T$. So $C_{N_R}^{L_R}(N_T + L_T) + C_{N_R}^{L_R}$ operations are needed for Stage 2. Eventually, by ignoring factors with low magnitudes, the complexity of NBJTRAS algorithm is $L_R N_T C_{N_R}^{L_R}$.

2.5.2 Antenna Selection and Use Scheduling Based on Successive Elimination

Another strategy of antenna selection and user scheduling is to successively remove the antennas and users who generate the least contribution to system performance. This strategy can be regarded as a sub-optimal algorithm. The complexity is gradually reduced as antennas and users are eliminated one after another. JASUS Algorithm [48] elaborates this strategy in this way. It successively deletes the antenna which undermines performance the most in each iteration. In the mean time, it chooses the group of users generating a great level of orthogonality during each iteration, which is elaborated by SUS Algorithm nested in JASUS Algorithm. The steps of JASUS and SUS are provided as follows. The target is to select N antennas out of M candidate antennas and schedule K users out of X candidate users in each time slot.

Algorithm 1 Steps of JASUS

initialize $\mathcal{A} \leftarrow \{1, ..., M\}$, $t \leftarrow 1$;
while $t \leqslant (M - N)$ **do** $maxRate \leftarrow 0$;
  for each m in $\mathcal{A}$ **do**
    $\mathcal{U}_t \leftarrow$ a set of K users using SUS($\mathcal{A}\backslash\{m\}, K$);
    $R_{-m} = R_{sum}(\mathcal{A}\backslash\{m\}, \mathcal{U}_t)$;
    if $R_{-m} > maxRate$ **then**
      $maxRate \leftarrow R_{-m}$;
      $m_{bad} \leftarrow m$;
      $\mathcal{U} \leftarrow \mathcal{U}_t$;
    end if
  end for
  $\mathcal{A} \leftarrow \mathcal{A}\backslash\{m_{bad}\}$;
  $t \leftarrow t + 1$;
end while
The set of antenna is given by $\mathcal{A}$ and the set of user is given by $\mathcal{U}$;

Here, we also analyze the complexity of JASUS algorithm. Note that due to multiple complicated loops in this algorithm, the results of complexity order estimation are not identical

Algorithm 2 Steps of the semi-orthogonal user selection function (SUS)

1: initialize $\mathcal{U} \leftarrow \{1, ..., X\}$, $i \leftarrow 1$, $\mathcal{S}_{\mathcal{U}} \leftarrow \varnothing$
2: **while** $i < (X - K)$ **do**
3: **for each** x in $\mathcal{U}$ **do**
4: $\mathbf{g}_{x,\mathcal{A}} = \mathbf{h}_{x,\mathcal{A}} - \sum_{j=1}^{i-1} \frac{\mathbf{h}_{x,\mathcal{A}}\widetilde{\mathbf{g}}_j^H \widetilde{\mathbf{g}}_j}{\|\widetilde{\mathbf{g}}_j\|^2}$;
5: **end for**
6: $i_{opt} = \mathbf{arg}\max_{x\in\mathcal{U}} \|\mathbf{g}_{x,\mathcal{A}}\|_2$;
7: $\mathcal{S}_{\mathcal{U}} \cup \mathcal{S}_{\mathcal{U}}\{i_{opt}\}$;
8: $\mathcal{U} \leftarrow \mathcal{U} \backslash \mathcal{S}_{\mathcal{U}}$;
9: $\widetilde{\mathbf{g}}_i = \mathbf{g}_{i_{opt}}$;
10: $i \leftarrow i + 1$;
11: **end while**
12: Output the set of user is given by $\mathcal{U}$;

from different perspectives. But the orders match the time consumption of the simulation executed in Chapter 3. Hence, we provide one perspective of estimating the complexity as the representative. For each t from JASUS algorithm, the main computational operations are on Step 3 5 where the operation times are $i(M - t)(X - i)$ for each i from the while loop in Step 2. Considering the number of loops, the key operation times in SUS algithm are $\sum_{i=1}^{K-1} i(M - t)(X-i) = (M-t)[\frac{(K-1)KX}{2} - \frac{(K-1)K(2K-1)}{6}]$. Back to JASUS algorithm, for each t within one while loop in Step 2, the main operation times are $(M-t)\sum_{i=1}^{K-1} i(M-t)(X-i)$. Since there are totally $(M - N)$ loops, the whole operations are $\sum_{t=1}^{M-N}(M - t)\sum_{i=1}^{K-1} i(M - t)(X - i)$. By concentrating on factors with high magnitude, the complexity is obtained as $K^2(M - N)MX$.

2.6 Current Designs for NOMA-based Relaying System

There are two existing ideas of deploying the relaying devices. One way is to apply a relay terminal for signal strength enhancement for single far user. The other way is to apply multiple devices, which share one NOMA signal, as relays to improve the receive SNR for each other.

They are described as follows.

2.6.1 System with Single Relay

The system is proposed by authors in [49]. In this system, there are two users in one cell where one user has weak connections to the BS due to long distance. So one relay is introduced to enhance the signal strength for far user. The transmission is executed in two phases:

First Phase: The BS transmits a NOMA signal superimposed by two messages. The relay only decodes the message for far user. The near user decodes the message for far user first, then it uses SIC to decode the message for its own.

Second Phase: The BS transmit the single message to near user. The relay transmits single message to far user. far user will receive its own message, but near user will receive two messages. Since based on SIC, near user removes the message for far user again and decodes the message for its own.

Therefore, the capacity for the near user is

$$C_1 = \frac{1}{2}\log(1 + \gamma_1(t_1)) + \frac{1}{2}\log(1 + \gamma_1(t_2)) \tag{2.22}$$

where $\gamma_1(t_1)$ is the PSNR obtained at near user in the first phase, and $\gamma_1(t_2)$ is the PSNR obtained in the second phase. And the capacity for the far user is given by

$$C_2 = \frac{1}{2}\log(1 + \min\{\gamma_0, \gamma_r, \gamma_2\}) \tag{2.23}$$

where γ_0 is the SNR for near user to decode the far user message in the first phase, γ_r is the

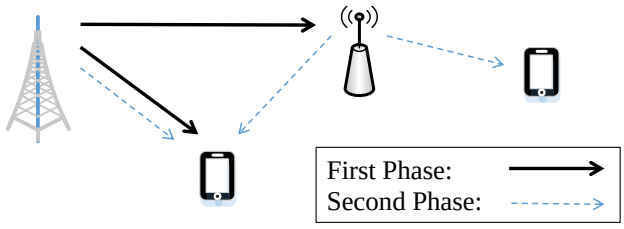

Figure 2.6: NOMA-based Relaying System with Single Relay.

SNR for the relay to decode the far user message in the first phase, and γ_2 is the SNR at far user in the second phase.

The achievable diversity and ergodic capacity in this model are also analyzed in [49]. The achievable diversity order is one for each message. And assuming ρ_b is the transmit SNR as the BS, the sum ergodic capacity of this system has the scaling $\log \rho_b$.

2.6.2 System with Multiple Relay Devices

The advantages of NOMA can be further demonstrated when every user served by the same BS can be applied as a relay. This idea is achieved by the authors' system design in [50]. To explain the design concretely, we follow the assumptions in Eq. (2.1). There are totally M phases, corresponding to M time slots, in this mechanism. In the 1-st Phase, the BS superimposes the messages for all users as a NOMA signal and transmits. Except the M-th user, each user decodes their own message using SIC. Then for the following phases, all users but the M-th one transmit the decoded message for the farther users to obtain the diversity gain. For example, for the $(m+1)$-th Phase, only the m-th user transmit the superposed $(M-m)$ messages to other users. In this way, the m-th user obtains m copies of its own message to improve the receive SNR. In other words, the user with poorer channel conditions will receive more message copies

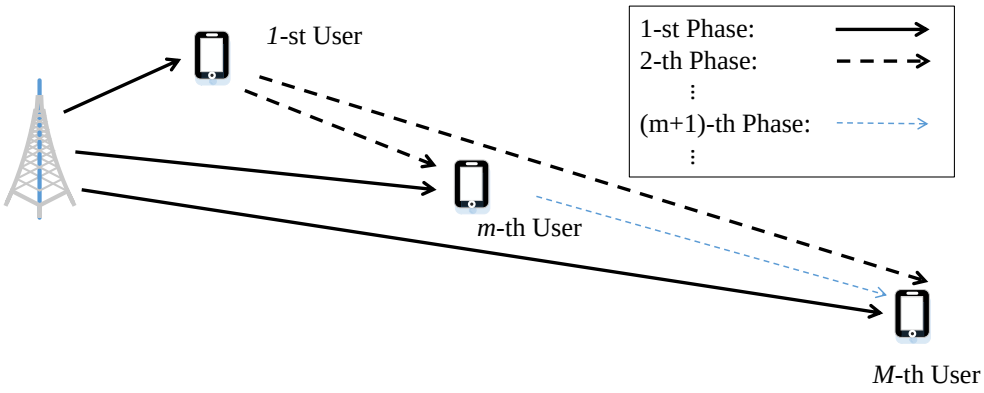

Figure 2.7: NOMA-assisted Relaying System with Multiple Relays.

for higher diversity gain as the compensation.

The diversity order is also obtained for the above scheme. With the assumption that $(m-1)$ best users detect relevant messages successfully, by combining the observations from multiple phases, the PSNR at the m-th user to decode its own message is

$$SNR_m^{Combine} = \frac{|h_m|^2\alpha_m}{|h_m|^2\sum_{i=1}^{m-1}\alpha_i+\frac{1}{\rho}} + \sum_{j=1}^{m-2}\frac{|h_{j\to m}|^2\alpha_{(j\to m),m}}{|h_{j\to m}|^2\sum_{i=j+1}^{m-1}\alpha_{(j\to m),i}+\frac{1}{\rho}} + \rho|h_{(m-1)\to m}|^2\alpha_{[(m-1)\to m],m} \tag{2.24}$$

where $h_{j\to m}$ is the channel coefficient from the j-th user to the m-th user and $\alpha_{(j\to m),m}$ is the power allocation ratio to the message for the m-th user in the NOMA signal from the j-th user to the m-th user. According to the proposition in [50], the scheme can guarantee that the m-th ordered user achieves a diversity order of M.

2.7 Chapter Summary

This chapter describes two 5G spectral efficient technologies and literature review about their applications so far. Firstly, the technical details of massive MIMO, NOMA and how they

achieve high spectral efficiency is introduced. Following this, the problems brought by these technologies are described, which are antennas selection, user scheduling in massive MIMO, and user pairing, power allocation, relaying mechanism under NOMA protocol. Next, as literature survey, some existing algorithms for solving the aforementioned problems are presented with mathematical demonstrations.

Chapter 3

Efficient Antenna Selection and User Scheduling in 5G Massive MIMO-NOMA System

3.1 Introduction

The remarkable growth of smart devices has led to 1,000-fold expected traffic enhancement for the future 5G network system [2], compared to which the existing spectrum resource is quite constrained. To support a great number of users as well as explosively increased network capacity with limited spectrum, 5G networks depend on critical transmission technologies to provide extremely high spectrum efficiency.

NOMA, which provides concurrent transmissions for multiple users, is recognized as one promising solution for high spectral efficiency in 5G [13]. In traditional OMA, messages intended for different users are transmitted in different time slots under TDMA protocol [14], or

in different frequency bands under OFDMA protocol [15]. Instead, NOMA allows several users to share the same time and spectral resources [50]. Specifically, NOMA serves different users by transmitting multiple messages at different power levels to achieve non-orthogonal reuse, which induces inter-user interference. At user side, the receivers apply successive interference cancellation (SIC) to remove the interference and separate these messages for corresponding users [51]. The performance of NOMA system can be improved if suitable users are clustered as a NOMA group for SIC.

Massive MIMO is also a technology which archives high spectral efficiency in 5G [52]. The very large antenna array is able to transmit massive data streams with different spatial patterns concurrently at the same frequency band to achieve spatial reuse. As a result, the throughput within the same spectrum is significantly improved. Another application is the massive antennas can transmit the same message to for the receiver side to achieve diversity gain, which increases the receive SNR. Moreover, beamforming technology in massive MIMO can potentially transmit messages in an specific angle, which is targeted at relevant NOMA user group. With the transmit power concentrated at that angle, the energy efficiency in increased and interference to other user groups is reduced.

Therefore, the integration of massive MIMO and NOMA technologies becomes a promising solution in obtaining extremely high spectral efficiency in 5G systems. However, massive MIMO-NOMA system brings certain major technical challenges at the same time. Firstly, MIMO RF chain elements, containing RF amplifier and analog-to-digital/digital-to-analog converter, increase the hardware cost along with system complexity, since each antenna should be match with one RF chain for signal processing. Note that during the transitional stage from 4G to 5G, RF chains are fewer than candidate antennas at the BS. So one critical scheme is to

operate MIMO-NOMA with limited RF chains is to select the best subset of antennas for these RF chains [53][54][55]. Different antennas correspond to different performance; even the same antenna subset has different performance over time due to random user distribution and user mobility. Hence, we should select the antenna subset with good performance in each time slot. Furthermore, since channel conditions vary over time and frequency, it is more cost-effective to schedule users into good channels at each time slot. Moreover, NOMA requires user pairing for SIC, where typically one near user (with strong channel gain) and one far user (with weak channel gain) are scheduled as a user pair. The inter-user channel gain different influences the SIC effect. Moreover, there will be very broad bandwidth in the future 5G, including licensed band and unlicensed band. Since channel gains with respect to different users are various in different subbands, we should assign user pairs into different subbands to improve the system performance. Hence, it is important to solve the antenna selection and user scheduling problem jointly in massive MIMO-NOMA system.

A few researchers have worked on the MIMO-NOMA system. In [43], user pairing algorithms are raised for two NOMA systems, but only single band scenarios are considered. A proportional-fairness scheduler is employed to decide user pairs in [56]; B. Kim et. al schedules two users with high correlation and large channel gain difference as a user pair [57]. But these two methods only focus on NOMA user pairing scheme. A norm-based joint transmit and receive antenna selection (NBJTRAS) scheme is presented in [47], which exhaustedly searches all candidate transmit and receive antennas for a subset with the largest channel norm. For [48], a joint antenna selection and user scheduling (JASUS) algorithm is proposed. It deletes the undesired antennas and users which generate a minimum contribution to the system performance on after another. However, the above two algorithms can only be applied to scenarios with

small candidate antenna and user sets due to high computational complexity. It is difficult to apply these algorithms to massive MIMO-NOMA system. Moreover, to the best of our knowledge, there exist no antenna selection algorithm based on NOMA communication protocol. In [58], M. F. Hanif et al. investigates the sum rate maximization problem of a NOMA-based MISO system, but the constraint of minimum data rate for each user is not considered.

In this chapter, we investigate the antenna selection and user scheduling problems in massive MIMO-NOMA system. We first consider a simple single-band two-user scenario, for which we figure out the power allocation strategy is to allocation the exact power for far user to reach the target rate. Then we raise an efficient search algorithm for single-band antenna selection. The desired antennas are searched from limited candidate antennas which are beneficial to the related users. Then we propose a joint AU contribution algorithm for joint antenna selection and user scheduling in multi-band multi-user scenario. It selects and schedules the antennas and users with the highest contribution to total channel gain.

The remainder of this chapter is organized as follows. Section II describes the system model and formulates the problem. In Section III, an efficient search algorithm is proposed for single-band two-user scenario. In Section IV, we present the joint AU contribution algorithm for multi-user multi-band scenario. Numerical results are presented and analyzed in Section V. Finally we conclude the paper in Section VI.

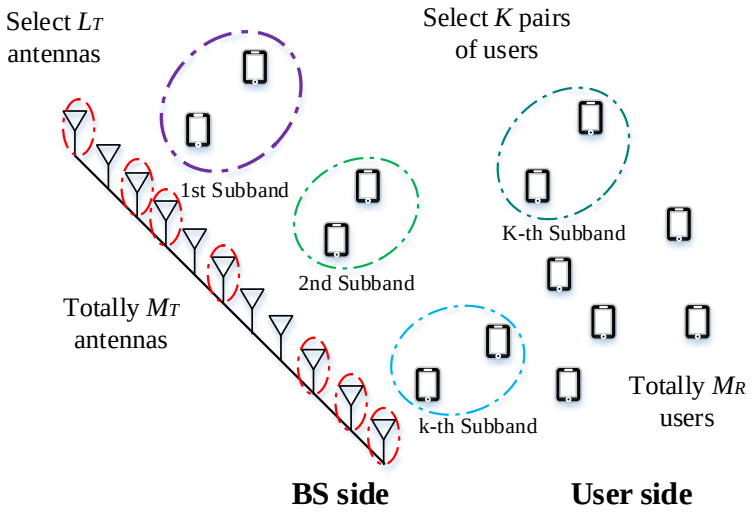

Figure 3.1: Massive MIMO-NOMA system with antenna selection and user scheduling.

3.2 System Model and Problem Formulation

3.2.1 Massive MIMO-NOMA System Model

We consider a massive MIMO-NOMA system in Fig. 4.1. The BS has M_T transmit antennas to communicate with M_R users. The total transmit bandwidth is divided into K orthogonal subbands. Since the BS has only L_T RF chains, L_T antennas are selected to serve users. For NOMA, two users are assigned to each subband to form a user pair. So $2K$ users in total are scheduled to be served. The candidate antenna set, candidate user set, desired selected antenna subset, desired scheduled user subset are denoted as $\mathcal{A}$, $\mathcal{U}$, $\mathcal{A}_S^o$, $\mathcal{U}_S^o$, respectively.

Provided antennas are selected and users are scheduled, a near user and a far user are served concurrently in the k-th subband, depicted in Fig. 3.2. Assume the same antenna corresponds to different channel coefficients in different subbands. Channel vectors from BS to near user and far user are denoted as $\mathbf{h}_k \in \mathbb{C}^{1\times L_T}$ and $\mathbf{g}_k \in \mathbb{C}^{1\times L_T}$, respectively, where $|\mathbf{h}_k|^2 > |\mathbf{g}_k|^2$. In each subband, total transmit power serving two users is P_0 and the power density of noise is N_0.

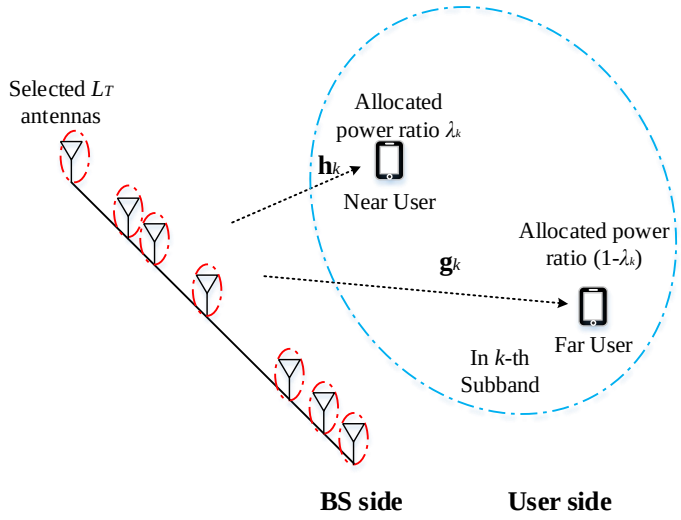

Figure 3.2: User service in k-th subband of massive MIMO-NOMA system.

According to NOMA protocol and relevant SIC mechanism [50], near user rate R_k^N and far user rate R_k^F should meet the following constraints

$$R_k^N \leqslant \log_2(1 + \rho\alpha_k|\mathbf{h}_k|^2), \tag{3.1}$$

$$R_k^F \leqslant \log_2(1 + \frac{\rho(1-\alpha_k)|\mathbf{g}_k|^2}{\rho\alpha_k|\mathbf{g}_k|^2+1}) \tag{3.2}$$

where $\alpha_k \in [0, 1]$ is the allocated power ratio for near user, $(1 - \alpha_k)$ is the ratio for far user and transmit SNR is defined as $\rho = \frac{P_0}{N_0}$ for each subband.

3.2.2 Problem Formulation

Suppose upper bound of data rate in Eq. (3.1)(3.2) can be obtained for each user. Based on selected antenna and user subsets $\mathcal{A}_S^o$ and $\mathcal{U}_S^o$, the sum rate in k-th subband is given by

$$R_{sum}^k(\mathcal{A}_S^o, \mathcal{U}_S^o) = \log_2[(1 + \rho\alpha_k|\mathbf{h}_k|^2)(1 + \frac{\rho(1-\alpha_k)|\mathbf{g}_k|^2}{\rho\alpha_k|\mathbf{g}_k|^2+1})]. \tag{3.3}$$

In NOMA, PSNR denotes the receive SNR obtained after SIC at the relevant users. To provide data rate guarantee, we set the minimum required PSNR as t for each user. Then the user sum rate maximization problem can be formulated as:

$$\begin{aligned} &\max_{\mathcal{A}_S^o \subset \mathcal{A}, \mathcal{U}_S^o \subset \mathcal{U}, \alpha_k \in [0,1]} R_{sum}(\mathcal{A}_S^o, \mathcal{U}_S^o) = \sum_{k=1}^{K} R_{sum}^k(\mathcal{A}_S^o, \mathcal{U}_S^o) \\ &s.t. \quad \begin{cases} \gamma_k^N(\mathcal{A}_S^o, \mathcal{U}_S^o), \gamma_k^F(\mathcal{A}_S^o, \mathcal{U}_S^o) \geqslant t, \qquad (k = 1, 2, \cdots, K) \\ L_T = |\mathcal{A}_S^o| < |\mathcal{A}| = M_T, \\ 2K = |\mathcal{U}_S^o| < |\mathcal{U}| = M_R \end{cases} \end{aligned} \tag{3.4}$$

where in k-th subband, PSNR at near user γ_k^N and PSNR at far user γ_k^F are expressed as

$$\gamma_k^N(\mathcal{A}_S^o, \mathcal{U}_S^o) = \rho\alpha_k|\mathbf{h}_k|^2; \quad \gamma_k^F(\mathcal{A}_S^o, \mathcal{U}_S^o) = \frac{\rho(1-\alpha_k)|\mathbf{g}_k|^2}{\rho\alpha_k|\mathbf{g}_k|^2+1}. \tag{3.5}$$

Although exhaustive search can be employed to find the optimal solution to the Problem (3.4), it requires high complexity. To solve the problem with low complexity, we give our algorithms in the following sections.

3.3 Antenna Selection in Single-band Two-user Scenario

To investigate the solution to our problem in multi-band multi-user scenario, we first consider simple single-band two-user antenna selection problem. For simplicity, we ignore script k for every related variable in this section.

3.3.1 Power Allocation Scheme

As the objective of our problem, user sum rate maximization is greatly affected by allocated power ratio to the message for each user. Hence, we raise the following power allocation scheme as the foundation for proposing further antenna selection and user scheduling algorithm.

We consider power allocation under the assumption that antennas have benn selected and users have been scheduled. In Eq. (3.3), we let $z = (1 + \rho\alpha|\mathbf{h}|^2)(1 + \frac{\rho(1-\alpha)|\mathbf{g}|^2}{\rho\alpha|\mathbf{g}|^2+1})$, so z derivative of α is given by

$$\begin{aligned}\frac{\partial z}{\partial \alpha} &= \rho|\mathbf{h}|^2 - \frac{\rho|\mathbf{g}|^2(1+\rho|\mathbf{g}|^2)}{(\rho\alpha|\mathbf{g}|^2+1)^2} + \frac{\rho^2|\mathbf{h}|^2|\mathbf{g}|^2(1-3\alpha-\alpha^2\rho|\mathbf{g}|^2+\alpha^2)}{(\rho\alpha|\mathbf{g}|^2+1)^2} \\ &= \frac{(\frac{1}{\rho|\mathbf{g}|^2}+1)(\frac{|\mathbf{h}|^2}{|\mathbf{g}|^2}-1)}{(\alpha+\frac{1}{\rho|\mathbf{g}|^2})^2} > 0,\end{aligned} \tag{3.6}$$

which proves that the sum rate in single-band scenario increases monotonically with α. Thus, the optimal power allocation method is to allocate the minimum power for far user PSNR to meet t, i.e., $\alpha = \frac{(1-\frac{t}{\rho|\mathbf{g}|^2})}{t+1}$. Based on Eq. (3.3), the maximum single-band two-user sum rate expression is derived as

$$R_p(\mathbf{h},\mathbf{g}) = \log_2[(1+\rho\frac{|\mathbf{h}|^2(1-\frac{t}{\rho|\mathbf{g}|^2})}{t+1})(1+t)]. \tag{3.7}$$

We validate whether PSNR at both user meet the constraints in (3.4) simultaneously by checking whether near user PSNR meets t, i.e., $\rho\alpha|\mathbf{h}|^2 \geqslant t$ where $\alpha = \frac{(1-\frac{t}{\rho|\mathbf{g}|^2})}{t+1}$. The reason is when near user PSNR meets t, it indicates far user PSNR has already met t by using the optimized power ratio. If near user PSNR cannot meet t by $\alpha = \frac{(1-\frac{t}{\rho|\mathbf{g}|^2})}{t+1}$, total power will be

allocated to near user message to maximize single-band user rate.

3.3.2 Efficient Search Algorithm for Antenna Selection

From Eq. (3.7), $R_p(\mathbf{h}, \mathbf{g})$ increases monotonically with channel gain from BS to near user $|\mathbf{h}|$ and channel gain associated with far user $|\mathbf{g}|$. Thus, we propose our efficient search algorithm targeting at selecting antennas with highest channel gain efficiently. The key for efficiency is that desired antennas are selected from limited candidate antennas with high channel gain for relevant users. However, it is not in an exhaustive way that the desired antenna subset is searched out of this limited scope.

To better describe our algorithm, we define $\mathcal{A}_S$ as an antenna subset with $|\mathcal{A}_S| = L_T$ and denote the single-band two users as User 1 and User 2. The channel gains of them are $|\mathbf{h}^{U_1}|$ and $|\mathbf{h}^{U_2}|$, respectively. We initialize $\mathcal{A}_S$ by selecting L_T antennas with highest channel gains for User 2. Then we successively swap the antenna with the lowest User 1 channel gain in $\mathcal{A}_S$ for the antenna with the highest User 1 channel gain in complement $\mathcal{A}_S^c$. Essentially, these successive operations are steps in a ordered searching process rather than exhaustive search in the limited scope. This is because within the selected antenna subset in each loop, a relatively ”bad” antenna for User 2 is replaced by a relatively ”good” antennas for User 1, which maintains the total channel gain of this antenna subset in a level as high as possible. Next, for each updated $\mathcal{A}_S$ after antenna swapping, we identify the user with higher channel gain as near user and the other as far user; Then we allocate power for both users using the scheme in Subsection 3.3.1. Finally, if the whole search process ends up without any $\mathcal{A}_S$ for both user PSNR to reach t concurrently, all power will be allocated to near user to maximize

single-band user rate.

Our efficient search algorithm is elaborated in Algorithm 3. a_i and a_i^c is the i-th item in $\mathcal{A}_S$ and $\mathcal{A}_S^c$. $\mathcal{A}_S \backslash \{a_i^*\} \cup \{a_i^{c*}\}$ denotes a set given by excluding a_i^* from $\mathcal{A}_S$ and adding a_i^{c*} in. $h_b^{U_1}$ ($b = 1, 2, ..., M_T$) is the channel coefficient from b-th antenna to User 1.

Algorithm 3 Efficient Search Algorithm

1: initialize $\mathcal{A}$, $R_{max} = 0$;
2: **while** $(\mathcal{A}_S = \varnothing) \| (\max\limits_{a_i^c \in \mathcal{A}_S^c} |h_{a_i^c}^{U_1}| > \min\limits_{a_i \in \mathcal{A}_S} |h_{a_i}^{U_1}|)$ **do**
3: **if** $\mathcal{A}_S = \varnothing$ **then**
4: select L_T antennas corresponding to the largest $|\mathbf{h}^{U_2}|$ from $\mathcal{A}$ into $\mathcal{A}_S$, and obtain $\mathcal{A}_S^c$, $|\mathbf{h}^{U_1}|$;
5: **else**
6: $a_i^{c*} = \arg\max\limits_{a_i^c \in \mathcal{A}_S^c} |h_{a_i^c}^{U_1}|$ and $a_i^* = \arg\min\limits_{a_i \in \mathcal{A}_S} |h_{a_i}^{U_1}|$;
7: $\mathcal{A}_S \leftarrow \mathcal{A}_S \backslash \{a_i^*\} \cup \{a_i^{c*}\}$, update $|\mathbf{h}^{U_1}|$, $|\mathbf{h}^{U_2}|$;
8: $\mathcal{A}_S^c \leftarrow \mathcal{A}_S^c \backslash \{a_i^{c*}\} \cup \{a_i^*\}$;
9: **end if**
10: $|\mathbf{h}| \leftarrow \max(|\mathbf{h}^{U_1}|, |\mathbf{h}^{U_2}|)$, $|\mathbf{g}| \leftarrow \min(|\mathbf{h}^{U_1}|, |\mathbf{h}^{U_2}|)$;
11: calculate $\alpha = \frac{1 - \frac{t}{\rho |\mathbf{g}|^2}}{t+1}$;
12: **if** $\rho \alpha |\mathbf{h}|^2 \geqslant t$ & $R_p(\mathbf{h}, \mathbf{g}) > R_{max}$ **then**
13: $R_{max} = R_p(\mathbf{h}, \mathbf{g})$, $\mathcal{A}_S^o \leftarrow \mathcal{A}_S$;
14: **end if**
15: **end while**
16: **if** $R_{max} = 0$ **then**
17: $R_{max} = \log_2(1 + \rho |\mathbf{h}|^2)$, $\mathcal{A}_S^o \leftarrow \mathcal{A}_S$;
18: **end if**
19: output R_{max}, $\mathcal{A}_S^o$;

The complexity of this algorithm is analyzed as follows for further comparison. Firstly, in Step 4, we rank the norms of channel coefficients for both User 1 and User 2, costing $(2M_T)$ computational operations. Next stage is the efficient search where loop operations are required. For each while loop in Step 2, we need to calculate the channel gain relevant to potential antenna group for both User 1 and User 2, which needs L_T operations. The searching times depend on the the relationship between L_T and $(M_T - L_T)$. If $L_T \leqslant M_T - L_T$, the extreme case

is there is no overlap between the antenna subset beneficial to User 1 and User 2 where all the beneficial antennas for User 2 should be replaced during swapping, resulting in L_T searching times. If $L_T > M_T - L_T$, the largest number of non-overlapping antennas is $(M_T - L_T)$, causing L_T searching times. Thereafter, the searching times will be $\min(L_T, M_T - L_T)$. In the above analysis, by considering variables with high magnitudes, the complexity of efficient search algorithm becomes $O(M_T + L_T \min(L_T, M_T - L_T))$.

3.4 Joint Antenna Selection and User Scheduling in Multi-band Multi-user Scenario

In this section, we investigate the problem of joint antenna selection and user scheduling in multi-band multi-user scenario on the basis of the proposed power allocation scheme in last section.

From Eq. (3.7), to find the highest R_p for system sum rate maximization, we have to maximize $|\mathbf{h}|$ and $|\mathbf{g}|$ in each subband, which means maximization of the channel gain associated with selected antennas and users. If we merely focus on antenna / user selection, we can select the antenna / user with the highest contribution to total channel gain. Antenna / user side contribution is defined as the ratio of channel gain occupied by certain antenna / user to total channel gain. Since we consider joint antenna selection and user scheduling problem, we should jointly take account of antenna and user side contribution. To be specific, the antenna / user side contribution should be weighted from the other side.

The process of joint AU contribution algorithm is presented in Algorithm 4 and a contribu-

tion update example is illustrated in Fig. 3.3. For this algorithm, define $\mathcal{U}_S^k$ as the selected user subset in k-th subband. The channel matrix for k-th subband ($k = 1, 2, 3, ..., K$) is denoted as $\mathbf{H}_k$. $h_{m,n,k}$ is the m-th row, n-th column element of $\mathbf{H}_k$. Define $\mathbf{h}_{m,k} = [|h_{1,m,k}|^2, |h_{2,m,k}|^2, ..., |h_{M_R,m,k}|^2]$, $\mathbf{h}_m = [\mathbf{h}_{m,1}, ..., \mathbf{h}_{m,k}, ..., \mathbf{h}_{m,K}]$, $m = 1, 2, ..., M_T$, and $\mathbf{h}_{n,k} = [|h_{n,1,k}|^2, |h_{n,2,k}|^2, ..., |h_{n,M_T,k}|^2]$ as channel vectors. For the contributions, we define $\mathbf{cw}^A = [cw_1^A, ..., cw_m^A, ...cw_{M_T}^A]$ as contribution vectors at antenna side and $\mathbf{cw}_k^U = [cw_{1,k}^U, ..., cw_{n,k}^U, ..., cw_{M_R,k}^U]$ as the ones at user side, where the elements are cw_m^A and $cw_{n,k}^U$ for m-th antenna and n-th user in k-th subband, respectively. Then the complete user side contribution vector is defined as $\mathbf{cw}^U = [\mathbf{cw}_1^U, ..., \mathbf{cw}_k^U, ..., \mathbf{cw}_K^U]$. We also define $\varphi(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{H}\|_2^2}$, where $\mathbf{H} \in \mathbb{C}^{KM_R \times M_T}$ is constructed by $\mathbf{H} = [\mathbf{H}_1^T\ \mathbf{H}_2^T ...\ \mathbf{H}_{K-1}^T\ \mathbf{H}_K^T]^T$.

Algorithm 4 Joint AU Contribution Algorithm

1: initialize the channel matrix $\mathbf{H}_k, k = 1, 2..., K$;
2: initialize antenna side contribution $cw_m^A = \frac{\sum_{k=1}^{K}\sum_{n=1}^{M_R} h_{n,m,k}}{\|\mathbf{H}\|_2^2}, m = 1, 2, ..., M_T$;
3: initialize user side contribution for each subband $cw_{n,k}^U = \frac{\sum_{m=1}^{M_T} h_{n,m,k}}{\|\mathbf{H}\|_2^2}, n = 1, 2, ..., M_R, k = 1, 2, ..., K$;
4: based on antenna side contribution, update user side contribution $cw_{n,k}^U = \phi(\mathbf{h}_{n,k}, \mathbf{cw}^A), n = 1, 2, ..., M_R, k = 1, 2, ..., K$;
5: $\mathcal{U}_S$ = NBUS($\mathbf{cw}^U$);
6: null rows corresponding to unselected users in each $\mathbf{H}_k$;
7: based on updated $\mathbf{H}_k$, update user side contribution $cw_{n,k}^U = \varphi(\mathbf{h}_{n,k}, \mathbf{cw}^A), n = 1, 2, ..., M_R, k = 1, 2, ..., K$;
8: based on updated user side contribution vector $\mathbf{cw}^U$, update antenna contribution $cw_m^A = \varphi(\mathbf{h}_m, \mathbf{cw}^U), m = 1, 2, ..., M_T$;
9: select L_T antennas with L_T highest cw_m^A and form $\mathcal{A}_S$;

Particularly, the NOMA-based User Selection (NBUS) Algorithm is embedded in Joint AU contribution Algorithm for user scheduling where we select desired users in a ordered manner. Firstly we sort all the elements in user side contribution. Then we pick up the element with the highest value, which is with respect to the n-th user in the k-th subband. Next we delete n-th user in other subbands. Then we set **flag** to guarantee that each subband is assigned with

Algorithm 5 NOMA-based User Scheduling Algorithm (NBUS)

input $\mathbf{cw}^U$;
initiate $i \leftarrow 0$, **flag** = $[flag_1, ...flag_k, ..., flag_K] \leftarrow \mathbf{0}$;
while $i < 2K$ **do**
    find out the element with the highest value in $\mathbf{cw}^U$, denoted as $cw^U_{n^*,k^*}$;
    while $flag_{k^*} = 2$ **do**
        find out the element with the next highest value in $\mathbf{cw}^U$, denoted as $cw^U_{n^*,k^*}$;
    end while
    schedule the n^*-th User for communication in the k^*-th Subband;
    $flag_{k^*} \leftarrow flag_{k^*} + 1$;
    $cw^U_{n^*,k} \leftarrow 0, \forall k$;
    $i \leftarrow i + 1$;
end while
output selected user set $\mathcal{U}_S$.

exactly two NOMA users, which are associated with channel gains as high as possible.

It is critical to select users before antennas. If we select users first, we assign each subband with the best suitable user pair then do antenna selection according to this assignment. However, if we select antennas first, the selection is under the assumption that at the user side, there is no restriction on user pairs assigned to subbands. In this way, the selected antennas and users are not suitable enough. And we are able decide the desired subset when the contribution vector on both sides are updated only once after initialization. This is because if we execute mutual weighting followed by single antenna / user elimination for several times, the results are quite similar. Mutual weighting will enlarge the gap between values of all elements in the contribution vector, but it makes little difference on the sorting results.

Here again, the complexity of Joint AU Contribution Algorithm is provided for further comparison. For Step 2 and Step 3, to generate the initial contribution for both the antenna side and user side, because of M_T candidate antennas, M_R candidate users and K subbands, we need $(2KM_TM_R)$ operations. Then for Step 4 and Step 5, the update of user side contribution also requires $(2KM_TM_R)$, followed by (M_RK) complexity for sorting. The operation times for

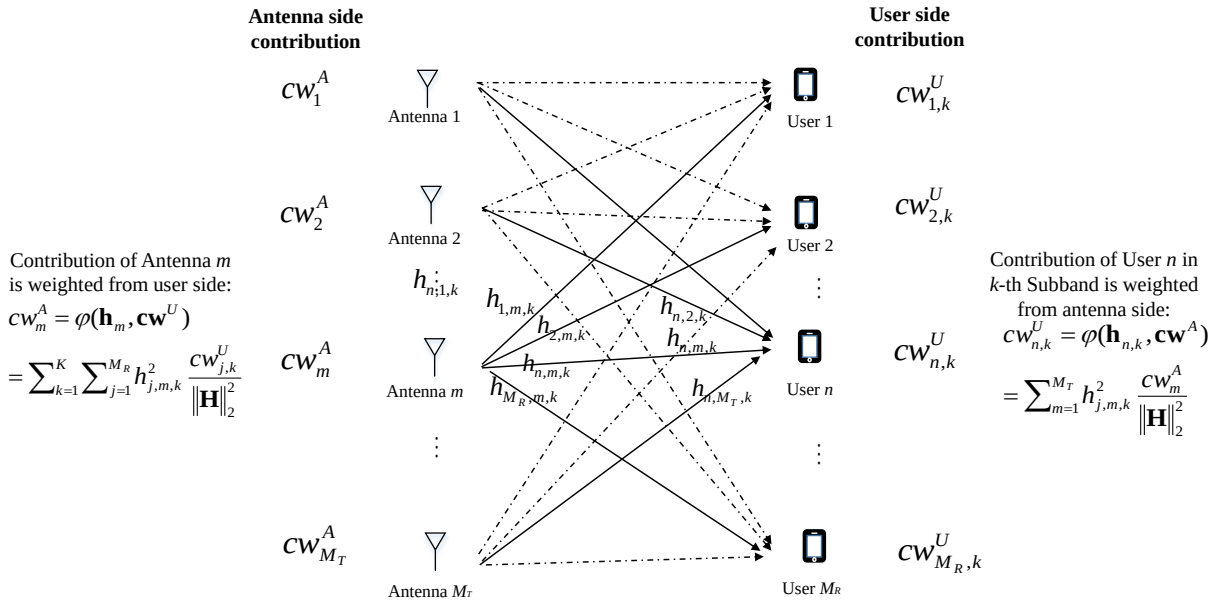

Figure 3.3: A contribution update example of joint AU contribution algorithm.

next updates, including user side and the antenna side (Step 8 and Step 9), are $(2KM_TL_R)$ due to eliminated $(M_R - L_R)$ users (Step 7). Lastly in Step 10, the sorting requires M_T operations. Similarly, by ignoring the factors with low magnitudes, we obtain the complexity of Joint AU Contribution Algorithm as (KM_TM_R).

3.5 Numerical Results

In this section, we compare the performance of user sum rate and outage probability obtained by proposed algorithms and existing methods. Note that our resource allocation scheme in Section 3.3.1 is executed in all methods for fair comparison; the outage probability is the ratio of the number of users not achieving the minimum required PSNR to the total number of users. All results are obtained by Monte Carlo simulations.

Firstly, we compare efficient search algorithm, joint AU contribution algorithm, optimal

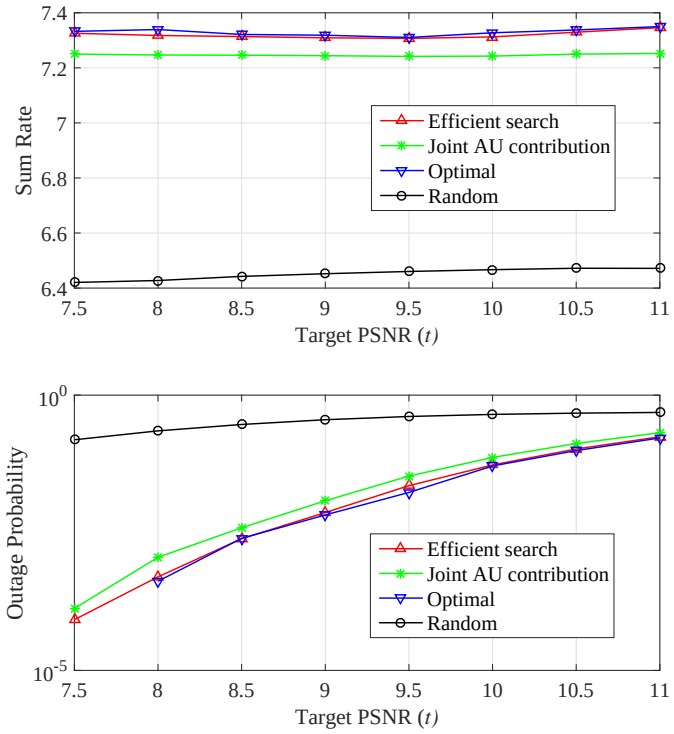

Figure 3.4: Sum rate and outage probability as functions of minimum required PSNR in single-band scenario where $M_T = 18$, $L_T = 6$.

(exhaustive search) and random antenna selection algorithm in single-band two-user scenario where $M_R = 2K = 2$, $L_T = 6$ and $\rho = 22$dB. Channel matrix from all antennas to both users is $\mathbf{H}^{U_1U_2} \in \mathbb{C}^{2\times M_T}$ where the elements are distributed as $\mathcal{CN}(0,1)$. Note that when applying joint AU contribution algorithm to this scenario, we only consider antenna side contribution. In other words, the steps for initialization and update of user side contribution will not be executed.

Fig. 3.4 shows the comparison of user sum rate and outage probability as functions of minimum required PSNR t when $M_T = 18$. For all algorithms except random selection algorithm, as t increases, sum rate decreases and then increases; outage probability increases monotonically with t. The reason is as t increases till $t = 9.5$, more power is supposed to be allocated to

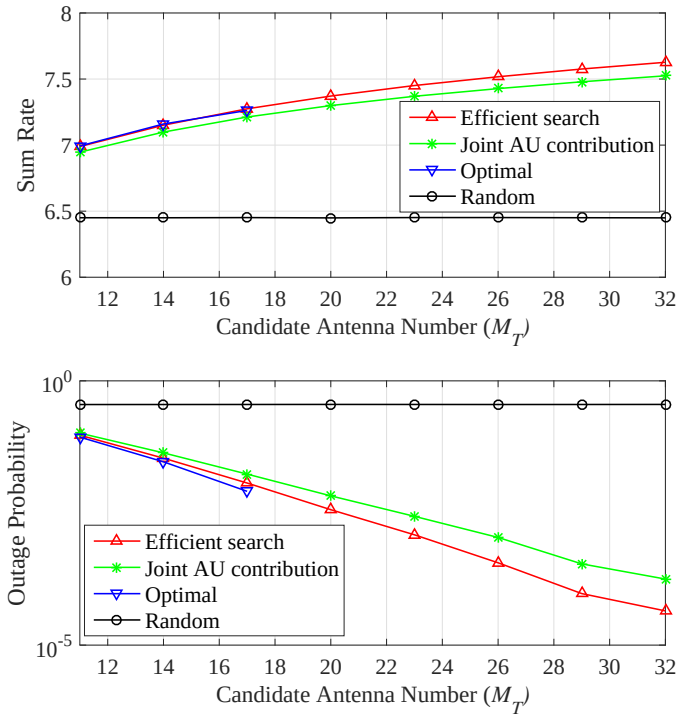

Figure 3.5: Sum rate and outage probability as functions of candidate antenna number in single-band scenario where $t = 9$, $L_T = 6$.

far user to make its PSNR meet higher t, resulting in the reduction of α. Thus, lower α causes sum rate reduction. Furthermore, when $t > 9.5$, it is more difficult to find α for both user PSNR to meet t. As a result, all power will be allocated to near user to increase single-band rate with the occurrence of far user outage. For performance comparison, sum rate and outage probability of efficient search algorithm are better than those of joint AU contribution algorithm and near-optimal.

In Fig. 3.5, we plot sum rate and outage probability as functions of candidate antenna number M_T when $t = 9$. Note that we only apply optimal selection algorithm to $M_T = 11, 14, 17$ due to huge simulation time for greater M_T. For each algorithm except random selection, larger sum rate is achieved and less outage occurs as M_T increases. The reason is as there are

more candidate antennas, the probability of selecting antennas with high channel gain becomes larger. And when $M_T = 17$, the sum rate of efficient search algorithm is higher than the optimal one but the optimal method still holds the a little bit lower outage probability. This is because it is more possible for optimal method to find the antenna subset which serves two NOMA users with data rate guarantee simultaneously. Here again, efficient search algorithm is near-optimal and outperforms joint AU contribution algorithm.

Next, joint AU contribution algorithm, NBJTRAS, JASUS and random selection algorithm are compared in a multi-user multi-band scenario with $M_T = 16, 32$, $L_T = 6$, $K = 3$ and $\rho = 22$dB. Elements in k-th subband channel matrix $\mathbf{H}_k$ are distributed as $\mathcal{CN}(0, 1)$. For NBJTRAS, JASUS algorithms, they are modified for the NOMA protocol. Specifically, there are $K \times M_R$ elements are the user side to denote total K subbands. To lower the outage probability, we should guarantee one user in each subband is with relatively high channel gain. Thus, far users are selected before the selection and identification of all near users are finished. The mechanism can cause that there are no users or only one user selected in some subbands. In this case, we replenish these positions by remaining users with the best channel gains in corresponding subbands.

For Fig. 3.6, comparison of user sum rate and outage probability as functions of t when $M_R = 10$ is presented. Here, NBJTRAS is only applied to $M_T = 16$ because of high computational complexity. One can observe that for each algorithm, sum rate keeps almost constant while outage probability increases monotonically with t. This is caused by the following reason. There are two types of subbands in this scenario. For first type of subband with larger channel gain, both user PSNR can reach t concurrently by power allocation. For the same reason as described in Fig. 3.4 analysis, two-user sum rate decreases as t increases. For second

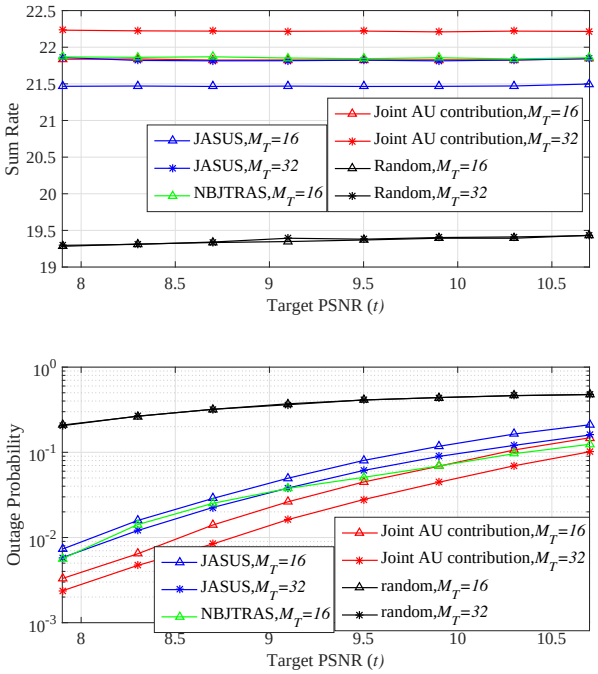

Figure 3.6: Performance of sum rate and outage probability as functions of minimum required PSNR in multi-band scenario where $M_R = 10$, $L_T = 6$, $K = 3$.

type of subband with smaller channel gain, both user PSNR cannot meet t concurrently. So all power will be allocated to near user, causing single-band rate increment and far user outage. In this way, rate increment in the second type compensates for rate reduction in first type of subband with more outages occurring. For performance comparison, joint AU contribution algorithm has similar performance to NBJTRAS and outperform JASUS.

Fig. 3.7 presents sum rate and outage probability as functions of candidate antenna number M_R when $t = 10$. Because large M_T and M_R lead to high computational complexity, we limit the two parameters to low values for NBJTRAS. For all algorithms except random selection, sum rate grows while outage probability decreases. The reason is since there are more candidate users to choose from, the probability of selecting users with high channel gain becomes higher. Here again, joint AU contribution algorithm has performance similar to NBJTRAS and better

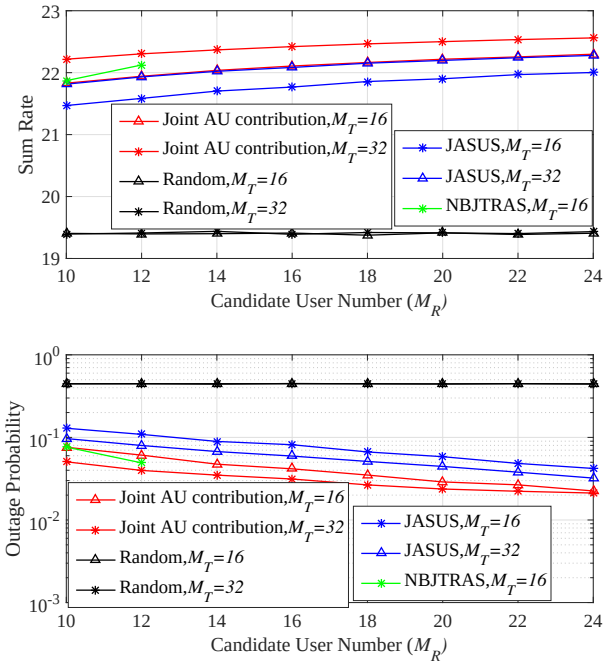

Figure 3.7: Performance of sum rate and outage probability as functions of candidate user number in multi-band scenario where $t = 10$, $L_T = 6$, $K = 3$.

than JASUS.

For single-band scenario, the complexity is $O(L_T C_{M_T}^{L_T})$ for optimal selection but $O(M_T + L_T \min(L_T, M_T - L_T))$ for efficient search algorithm and $O(M_T)$ for joint AU contribution algorithm. See the Appendix B for the complexity analysis of optimal algorithm and joint AU contribution algorithm in this scenario. Hence, with complexity far less than optimal selection and slightly higher than joint AU contribution algorithm, efficient search algorithm attains the trade-off between performance and complexity.

In multi-band scenario, the complexity is $O(KM_T M_R)$ for joint AU contribution algorithm while $O(KM_T C_{KM_R}^{2K})$ for NBJTRAS (by letting $L_R = 2K$, $N_T = M_T$ and $N_R = M_R$ in Chapter 2) and $O(K^3(M_T - L_T)M_T M_R)$ for JASUS (by letting $M = M_T$, $X = KM_R$ in Chapter 2). So for practical multi-subband multi-user massive MIMO-NOMA, joint AU contribution algorithm

presents performance similar to NBJTRAS but requires far lower complexity. This also proves that joint AU contribution algorithm is suitable for massive MIMO where antenna and user numbers are large.

3.6 Discussion: Practical Execution of Proposed Algorithm

The above simulation results can be achieved under the assumption that the perfect channel information is known. To approach the theoretical performance as much as possible, the proposed algorithm can be practically executed as follows.

As the initialization for data transmission, by channel estimation, the channel coefficients between each candidate antennas and candidate users are obtained. Then the proposed algorithm is executed as the BS side, which holds good computational ability, to select the antennas and scheduled users. According to the selected subset, the associated data is transmitted from the selected antennas to the selected users.

Note that the selection process and service provision are in a dynamic manner due to channel condition variation. Therefore, the channel estimation should be executed periodically. Under critical overall channel coefficient change, the proposed algorithm should be executed again to update the selected antenna and user subset accordingly. Sometimes it occurs that the complete channel coefficient information cannot be obtained due to serious noise or possible hardware failure. In this way, the historical channel information will be applied to predicting the channel condition in the current moment.

3.7 Chapter Summary

This chapter studies the antenna selection and user scheduling problems in massive MIMO-NOMA system. We propose efficient antenna selection and user scheduling algorithms for two scenarios. For the single-band two-user scenario, our efficient search algorithm selects desired antennas from limited candidate antennas beneficial to the relevant users. For the multi-band multi-user scenario, joint AU contribution algorithm takes accounts of the contribution of each antenna's and user's channel gain to total channel gain jointly. Numerical results show that efficient search algorithm achieves near-optimal performance, and joint AU contribution algorithm has performance similar to existing methods but requires lower complexity.

Chapter 4

Power Allocation and Performance of Collaborative NOMA Assisted Relaying System

4.1 Introduction

It is well known that serving cell-edge users from the BS is difficult because the signal power decays exponentially with the distance, which is greatly serious in common non-line-of-sight communication environments [59]. Moreover, the rapid proliferation of smart devices in the future 5G causes quite a few cell-edge users in need of guaranteed quality of service [2][60], the standard of which is raised due to various of real-time mobile applications, such as video streaming and other online entertainments. So it is necessary to provide service for multiple cell-edge users simultaneously with guaranteed data rate. However, other than transmitting data to multiple users, the 5G networks have to process massive signallings and coordinate

BSs as well [61]. The power resource is constrained compared to a great number of tasks. Furthermore, communicating with multiple users requires more spectrum resources; If we limit the connections in the existing spectrum, the inter-user interference will be too strong for good quality of service. Conditioned on constrained power and spectrum resource, it is challenging to serve multiple cell-edge users concurrently with data rate guarantee.

Therefore, candidate power and spectral efficient strategies are urgently required to provide concurrent data transmissions to multiple cell-edge users. One strategy is to use the relay, which increases the user receive SNR by short-range communication even with low transmit power. Compared to the amplify-and-forward (AF) relay, the decode-and-forward (DF) relay is preferred since only the desired signals for cell-edge users are sent from the transmitter of the relay. Furthermore, the D2D transmission functions of smart devices make it possible to apply some device as the relay, which save the effort of establishing extra relays. The smart device can receive the message intended for its own and forward the messages for cell-edge users. Another strategy, NOMA, the promising spectral efficient technology for 5G, can improve the performance of relaying systems. NOMA allows messages for several users multiplexed at the same time and frequency band by allocating different power [13][62] to these messages. At the user side, SIC is applied to the message separation. Specifically, users with better channel conditions decode other users' messages and then remove them in order to decode messages for their own [63]. Furthermore, because of superposition coding, the power ratios among messages in a NOMA signal can be maintained at the receivers in relaying systems. Particularly, ratios are maintained by good channel conditions in the S-R link; because of short-range communication, the channel condition in the R-D links may be better than S-R link, resulting in the ratio maintenance, as well. This maintenance helps us better analyze

and control the receive SNR at most of the NOMA users. Such effect cannot be achieved by separated message links in conventional OMA. In this way, using NOMA to provide multiple message inputs and outputs for the relay becomes a promising solution to serving several cell-edge users simultaneously.

In this chapter, we propose the Collaborative NOMA Assisted Relaying (CNAR) system in 5G featured by the collaboration of the S-R NOMA link as macro-cell communication and R-D NOMA link as small-cell communication. The BS superimposes messages intended for the relay and cell-edge users as a NOMA signal and transmits. The relay decodes its own message from the NOMA signal by SIC scheme; then the relay superposes messages for cell-edge users with adjusted power in NOMA protocol and transmits them to the destinations. The SIC scheme at the relay fits DF relaying mechanism very well because by SIC, some messages in the source NOMA signal are decoded at the relay and the relay only transmits the messages preferred by the cell-edge users. Since multiple users need to be served simultaneously, the system throughput should be improved. Thus, the S-R and R-D phases are connected by a double-antenna relay which operates in full-duplex mode to provide high throughput. Moreover, the S-R and R-D phases are executed in licensed and unlicensed band, respectively, to avoid the interference. Particularly, signals in the S-R phase are carried by cellular communication protocol, like the prevalent LTE, while the device applied as the relay can transmit signals to other users using D2D communication protocols, e.g., WiFi and Bluetooth. However, the performance of the proposed system is yet to be characterized and power allocation among NOMA signals should be determined to guarantee the data rate for multiple users.

Several works have been done on the performance evaluation for the NOMA assisted relaying systems. In [64], the outage performance of NOMA-assisted AF-relaying networks is

analyzed but the drawbacks of AF relaying system undermine the advantage of NOMA. In particular, at the relay, the power allocated to each message cannot be adjusted for R-D links, and the interference from the S-R link is amplified as well. In [65], the capacity of a NOMA-based DF relaying system is analyzed where the destination receives two messages from two time slots, respectively. And the author proposes a suboptimal power allocation scheme to maximize the sum rate associated with these messages. S. Rini et al. in [66] also considers rate maximization at multiple cell-edge users assisted by several relays. For [67], a relay selection mechanism is proposed to obtain the lowest outage probability among all other relay selection schemes. However, the above researchers fails to consider applying certain user device as the relay. Furthermore, the authors in [65][66] fail to consider the fairness where the rate related to each message should reach a target level. For [49][68], the outage probability for user fairness and ergodic sum rate expressions for NOMA downlink cooperative networks are investigated. In [69], a cooperative SWIPT NOMA protocol is raised where some devices with DF relaying functions are selected to assist cell-edge devices. In [50], the author derives the outage probability of a extremely complex system where sequential NOMA links from multiple relaying devices are used to support cell-edge users. User fairness is considered in [50][65]. However, these works only focus on low-throughput half-duplex relaying mode where the cell-edge user receive SNR depends on links from the BS and the relay.

As we propose our CNAR system based on full-duplex relaying mode, the system performance is analyzed in the following way. Firstly, we define that the outage occurs when the achievable rate with regard to any message doesn't reach the target rate. And the exact system outage probability expression is derived by analyzing the outage behavior of S-R and R-D phases, separately. Following this, to guarantee the data rate, we raise the optimal power allo-

cation scheme by minimizing the system outage probability. Then, the ergodic sum capacity of CNAR system in high SNR regime is approximated based on the optimal power allocation scheme. Specifically, the capacity approximation for cell-edge users are obtained by discussing the interference at them on the basis of NOMA protocol.

The remainder of this chapter is organized as follows. Section 4.2 describes the CNAR system model. In Section 4.3, with exact expression of system outage probability derived, an optimal power allocation scheme is proposed by minimizing the outage probability; the ergodic sum capacity analysis and approximation are also provided. Then numerical results are presented and analyzed in Section 4.4. Finally Section 4.5 concludes the paper.

4.2 System Model

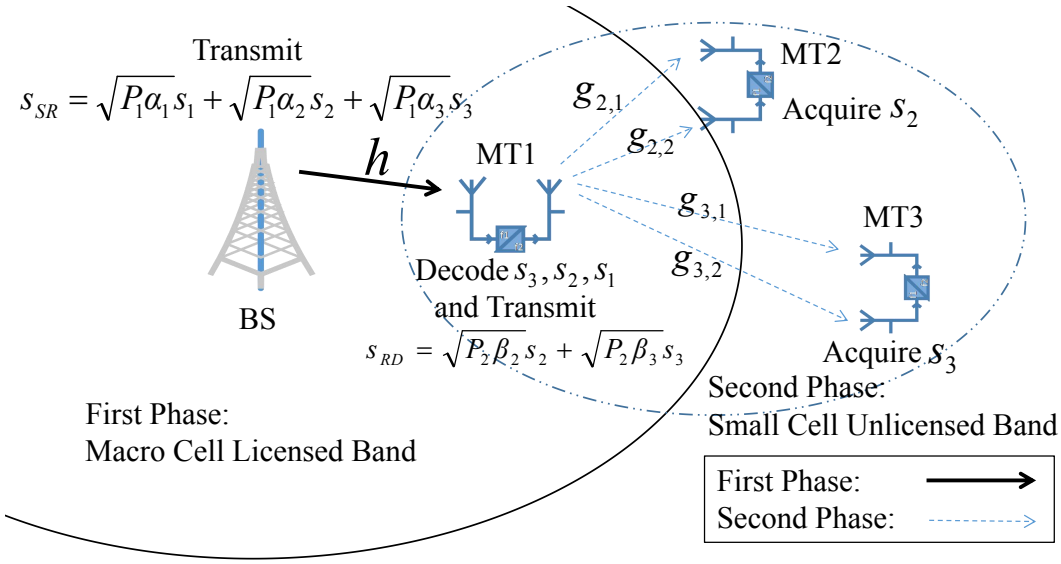

Figure 4.1: CNAR system model.

The CNAR communication system is proposed in Fig. 4.1. We assume that a group of three mobile terminals (MTs) are supported by the BS, where MT1 is the nearest to BS. Due to

long distances from the BS to MT2 and MT3, regarded as cell-edge users, it is hard for BS to provide good quality of service via direct links. Hence, we apply MT1 as a relay to establish connections from BS to MT2 and MT3. In this way, there are two phases for serving the group of three MTs. In the first phase, i.e., the S-R phase, BS superposes three messages as a NOMA signal to be transmitted to MT1; MT1 receives the signal and decodes all three messages. In the second phase, i.e., the R-D phase, MT1 superimposes messages for MT2, MT3 with adjusted power as another NOMA signal and transmits; MT2, MT3 receive the signal and acquire their own messages. Assume that each MT is equipped with double antennas. To achieve high throughput, MT1 operates in full-duplex relaying mode. It receives the NOMA signal from BS with one antenna (receive antenna) and, at the same time, transmits the decoded messages in last time slot with the other antenna (transmit antenna) to MT2, MT3. Moreover, to avoid the interference, the first phase is executed in licensed band as macro-cell communication, and the second phase is in unlicensed band as small-cell communication. We also assume that the both the BS and the relay transmit messages at a rate matching the Shannon channel capacity for each messages.

Specifically, in the first phase, BS superposes three messages as the NOMA signal $s_{SR} = \sqrt{P_1\alpha_1}s_1 + \sqrt{P_1\alpha_2}s_2 + \sqrt{P_1\alpha_3}s_3$, which is transmitted to MT1. s_1, s_2 and s_3 are messages intended for MT1, MT2 and MT3, respectively. P_1 is the transmit power from one BS antenna assigned to serve this group of three MTs. α_1, α_2 and α_3 are the ratios of power allocated to s_1,

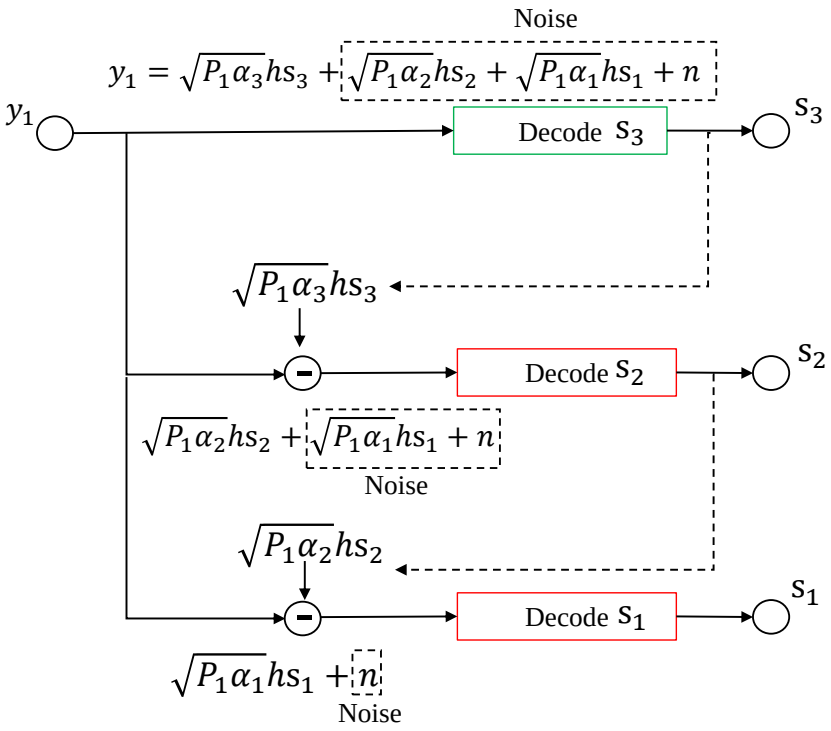

Figure 4.2: Decoding process at MT1 for the first phase based on SIC.

s_2 and s_3 where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Thus, the received signal at MT1 is given by

$$\begin{aligned} y_1 &= hs_{SR} + n_1 \\ &= \sqrt{P_1\alpha_1}hs_1 + \sqrt{P_1\alpha_2}hs_2 + \sqrt{P_1\alpha_3}hs_3 + n_1 \end{aligned} \tag{4.1}$$

where $h \sim \mathcal{CN}(0, \sigma_1^2)$ denotes the BS-MT1 Rayleigh channel coefficient and $n_1 \sim \mathcal{CN}(0, N)$ denotes the AWGN at MT1 receive antenna.

After receiving s_{SR}, MT1 decodes the three messages using the principle of SIC. Particularly, it treats s_2, s_1 and n_1 as the noise to decode s_3. Then by removing s_3, which is part of the interference when decoding s_2, it treats s_1 and n_1 as the noise to decode s_2. Next, with s_2 removed, n_1 is treated as the noise to decode s_1. The SIC-based decoding process is illustrated in Fig. 4.2. Considering the user fairness, we set the target rate as R_0 related to each message as

the standard for quality of service. Hence, the corresponding target SNR becomes $\gamma_0 = 2^{R_0} - 1$. To reach the target SNR, the following conditions should be met consecutively:

$$\begin{aligned} SNR_{3,1} &= \frac{\alpha_3 \rho_1 |h|^2}{(1-\alpha_3)\rho_1 |h|^2 + 1} \geqslant \gamma_0, \\ SNR_{2,1} &= \frac{\alpha_2 \rho_1 |h|^2}{\alpha_1 \rho_1 |h|^2 + 1} \geqslant \gamma_0, \\ SNR_{1,1} &= \alpha_1 \rho_1 |h|^2 \geqslant \gamma_0, \end{aligned} \tag{4.2}$$

where $SNR_{k,1}$ is the PSNR [70] at MT1 for the k-th message and $\rho_1 \triangleq \frac{P_1}{N}$ is BS transmit SNR. Here, when $|h|^2$ is high, it holds that $SNR_{3,1} \approx \frac{\alpha_3}{1-\alpha_3}$ and $SNR_{2,1} \approx \frac{\alpha_2}{\alpha_1}$, which demonstrates that the power ratios for s_3 and s_2 can be maintained. Moreover, the achievable rate associated with s_1 is given by

$$C_1 = \log_2(1 + \alpha_1 \rho_1 |h|^2). \tag{4.3}$$

For the second phase, by superposing s_2 and s_3 as the NOMA signal s_{RD}, MT1 transmits $s_{RD} = \sqrt{P_2 \beta_2} s_2 + \sqrt{P_2 \beta_3} s_3$ to MT2 and MT3. P_2 is the transmission power from MT1. β_2 and β_3 are the power allocation ratios for s_2 and s_3 where $\beta_2 + \beta_3 = 1$. Because MT2, MT3 have double antennas for receiving, their received signals are

$$\begin{aligned} \mathbf{y}_2 &= \mathbf{g}_2 s_{RD} + \mathbf{n}_2 \\ \mathbf{y}_3 &= \mathbf{g}_3 s_{RD} + \mathbf{n}_3 \end{aligned} \tag{4.4}$$

where $\mathbf{g}_2 = [g_{2,1}, g_{2,2}]^T$, $\mathbf{n}_2 = [n_{2,1}, n_{2,2}]^T$ and $\mathbf{g}_3 = [g_{3,1}, g_{3,2}]^T$, $\mathbf{n}_3 = [n_{3,1}, n_{3,2}]^T$. $g_{2,1}, g_{2,2} \sim \mathcal{CN}(0, \sigma_2^2)$ denote Rayleigh channel coefficients from MT1 single transmit antenna to MT2 double antennas, corresponding to AWGN $n_{2,1}, n_{2,2} \sim \mathcal{CN}(0, N)$. And $g_{3,1}, g_{3,2} \sim \mathcal{CN}(0, \sigma_3^2)$

denote Rayleigh channel coefficients from MT1 single transmit antenna to MT3 double antennas, with AWGN $n_{3,1}, n_{3,2} \sim \mathcal{CN}(0, N)$.

In every time slot, we should identify NOMA far user and near user to apply SIC at the receiver of near user [51]. Assuming MRC is applied at MT2, MT3, the combining channel gains for them are $\|\mathbf{g}_2\|_2^2$ and $\|\mathbf{g}_3\|_2^2$, respectively. Hence, when $\|\mathbf{g}_2\|_2^2 > \|\mathbf{g}_3\|_2^2$, MT2 is identified as near user with power ratio for near user $\beta_N = \beta_2$ and MT3 as far user with power ratio for far user $\beta_F = \beta_3$; otherwise, their statuses are swapped with $\beta_N = \beta_3, \beta_F = \beta_2$. The decoding processes are explained by Fig. 4.3. In this figure, s_F, $\mathbf{g}_F$, n_F are the message, channel gain and AWGN corresponding to far user while s_N, $\mathbf{g}_N$, n_N are for near user. For far user, it considers s_N and n_N as the noise to decode s_F. Since $\|\mathbf{g}_N\|_2^2 > \|\mathbf{g}_F\|_2^2$, s_F can be decoded and then removed by near user. So for near user, it executes the same step as far user first; then it removes s_F from the NOMA signal and decodes s_N with only n_N as the noise.

Again, with the target rate R_0 for each message, the conditions for far user and near user to meet the target SNR are:

$$\begin{aligned} SNR_{F,2} &= \frac{\beta_F \min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}\rho_2}{\beta_N \min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}\rho_2 + 1} \geqslant \gamma_0, \\ SNR_{N,2} &= \beta_N \max\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}\rho_2 \geqslant \gamma_0. \end{aligned} \tag{4.5}$$

where $SNR_{F,2}$ denotes the PSNR at far user, $SNR_{N,2}$ is the PSNR at near user and $\rho_2 \triangleq \frac{P_2}{N}$ is MT1 transmit SNR. When $\min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}$ is large due to short-range communication, $SNR_{F,2} \approx \frac{\beta_F}{\beta_N}$ means that the power ratio between far user and near user message is maintained.

In fact, MT1 is a DF relay to connect transmissions in the first and second phase. Since the end-to-end data rate of DF relaying is dominated by the weakest link [71], the achievable rate

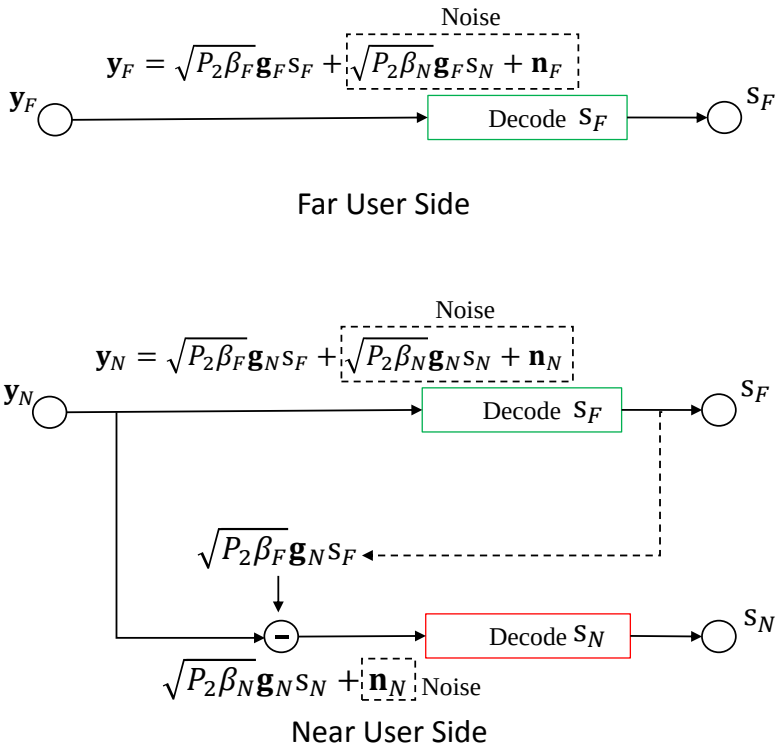

Figure 4.3: SIC-based decoding process at NOMA far user and near user for the second phase.

associated with s_2 and s_3 are

$$C_2 = \log_2(1 + \min\{SNR_{2,1}, SNR_{2,2}\}) \tag{4.6}$$

$$C_3 = \log_2(1 + \min\{SNR_{3,1}, SNR_{3,2}\}) \tag{4.7}$$

where $SNR_{2,2}$ and $SNR_{3,2}$ are PSNR at MT2 and MT3, respectively.

4.3 Power Allocation and Performance

In this section, we derive the system outage probability and propose the optimal power allocation scheme by minimizing the outage probability, which aims to achieve single-user data rate

guarantee and inter-user fairness. Then we provide the analysis of the ergodic sum capacity based on the proposed power allocation scheme.

4.3.1 Power Allocation Scheme and Outage Performance

Define that during the two-phase communication in the proposed CNAR system, if the achievable rate associated with any message is lower than the target rate, the outage occurs in our CNAR system. Let Event A = {All constraints in (4.2) are met} for the first-phase transmission, Event B = {All constraints in (4.5) are met} for the second-phase transmission. So the system outage probability is expressed as

$$P_{out} = 1 - \Pr[A]\Pr[B]. \tag{4.8}$$

According to (4.2), we derive $\Pr[A]$ as follows

$$\Pr[A] = \Pr[|h|^2 > \max\{\gamma_1, \gamma_2, \gamma_3\}] \tag{4.9}$$

where $\gamma_1 \triangleq \frac{\gamma_0}{\rho_1\alpha_1}$, $\gamma_2 \triangleq \frac{\gamma_0}{\rho_1(\alpha_2-\gamma_0\alpha_1)} > 0$ and $\gamma_3 \triangleq \frac{\gamma_0}{\rho_1[\alpha_3-\gamma_0(1-\alpha_3)]} > 0$.

To derive $\Pr[B]$, (4.5) can be rewritten as

$$\begin{aligned} \min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} &> \frac{\gamma_0}{\rho_2[\beta_F(1+\gamma_0)-\gamma_0]} \triangleq w_F > 0 \\ \max\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} &> \frac{\gamma_0}{\rho_2\beta_N} \triangleq w_N \end{aligned} \tag{4.10}$$

where $\min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}$ is the channel gain for far users and $\max\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\}$ is the channel

gain for near user. Since far user has lower channel gain, it is reasonable to set lower channel gain threshold for far user than near user to achieve higher probability of meeting the constrains in (4.10). Hence, we assume $w_N \geqslant w_F$, indicating $\frac{1+\gamma_0}{2+\gamma_0} \leqslant \beta_F < 1$. Then $\Pr[B]$ is given by

$$\begin{aligned}\Pr[B] =& \Pr[\min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} > w_F, \max\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} > w_N] \\ =& \phi(w_N, \sigma_2)\phi(w_F, \sigma_3) + \phi(w_F, \sigma_2)\phi(w_N, \sigma_3) - \phi(w_N, \sigma_2)\phi(w_N, \sigma_3)\end{aligned} \tag{4.11}$$

where $\phi(w, \sigma) \triangleq \frac{1}{\sigma^2}\exp(-\frac{w}{\sigma^2})w + \exp(-\frac{w}{\sigma^2})$. See the appendix for the proof of $\Pr[B]$.

To maximize $\Pr[A]$, we have to minimize $\max\{\gamma_1, \gamma_2, \gamma_3\}$. So finding the optimal power allocation ratios in the first phase is equivalent to the following problem

$$\begin{aligned}&\min_{\alpha_1,\alpha_2,\alpha_3} \max\{\gamma_1, \gamma_2, \gamma_3\} \\ =& \max_{\alpha_1,\alpha_2,\alpha_3} \min\{\alpha_1, \alpha_2 - \gamma_0\alpha_1, \alpha_3 - \gamma_0(1-\alpha_3)\} \\ &s.t. \quad \begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \alpha_2 - \gamma_0\alpha_1 > 0 \\ \alpha_3 - \gamma_0(1-\alpha_3) > 0. \end{cases}\end{aligned} \tag{4.12}$$

Inspired by [72], we introduce α_4 to represent the objective. So problem (4.12) is equivalent

to to following form

$$\begin{aligned} &\min_{\alpha_1,\alpha_2,\alpha_3} (-\alpha_4) \\ &s.t. \quad \begin{cases} 0 < \alpha_4 \leqslant 1 - (1+\gamma_0)(\alpha_1+\alpha_2) \\ 0 < \alpha_4 \leqslant \alpha_2 - \gamma_0\alpha_1 \\ 0 < \alpha_4 \leqslant \alpha_1 \end{cases} \end{aligned} \tag{4.13}$$

Here, we can apply Lagrange multiplier method to find a possible solution [73]. The Lagrange function is written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}) = &- \alpha_4 + \mu_1[(1+\gamma_0)(\alpha_1+\alpha_2) + \alpha_4 - 1] \\ &+ \mu_2(\gamma_0\alpha_1 - \alpha_2 + \alpha_4) + \mu_3(-\alpha_1 + \alpha_4). \end{aligned} \tag{4.14}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_4]$ and $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]$.

Based on Karush-Kuhn-Tucker (KKT) conditions [73], by letting the partial derivatives with respect to $\boldsymbol{\alpha}$ of $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu})$ equal to zero, i. e., $\frac{\partial \mathcal{L}}{\partial \alpha_1} = \frac{\partial \mathcal{L}}{\partial \alpha_2} = \frac{\partial \mathcal{L}}{\partial \alpha_4} = 0$, we obtain

$$\begin{cases} \mu_1 = \frac{1}{\gamma_0^2+3\gamma_0+3}, \\ \mu_2 = \frac{1+\gamma_0}{\gamma_0^2+3\gamma_0+3}, \\ \mu_3 = 1 - \frac{2+\gamma_0}{\gamma_0^2+3\gamma_0+3}. \end{cases} \tag{4.15}$$

KKT conditions also require $\mu_1[(1+\gamma_0)(\alpha_1+\alpha_2)+\alpha_4-1]+\mu_2(\gamma_0\alpha_1-\alpha_2+\alpha_4)+\mu_3(-\alpha_1+\alpha_4) =$

0. Since in (4.15) μ_1, μ_2 and μ_3 are positive values, the following inequations should be satisfied

$$\begin{cases} (1+\gamma_0)(\alpha_1+\alpha_2)+\alpha_4-1=0, \\ \gamma_0\alpha_1-\alpha_2+\alpha_4=0, \\ -\alpha_1+\alpha_4=0. \end{cases} \tag{4.16}$$

So the candidate solution obtained from (4.16) is given by

$$\begin{cases} \alpha_1 = \frac{1}{\gamma_0^2+3\gamma_0+3}, \\ \alpha_2 = \frac{1+\gamma_0}{\gamma_0^2+3\gamma_0+3}, \\ \alpha_3 = 1-\frac{2+\gamma_0}{\gamma_0^2+3\gamma_0+3}, \\ \alpha_4 = \frac{1}{\gamma_0^2+3\gamma_0+3}. \end{cases} \tag{4.17}$$

To check the feasibility of the candidate solution (4.17) with second order sufficient conditions (SONC) [73], we obtain $\boldsymbol{y}$ for SONC by calculating the Jacobian evaluated at the solution, which is given by

$$\boldsymbol{y} = \boldsymbol{J}(\boldsymbol{\alpha}^{(1)})^T \cdot \mathbf{0} = \begin{bmatrix} 1+\gamma_0 & 1+\gamma_0 & 1 \\ \gamma_0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}^T \cdot \mathbf{0} = \mathbf{0} \tag{4.18}$$

And the Hessian of Lagrangian (4.14) is calculated as

$$\nabla^2_{\boldsymbol{\alpha}}\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\mu}) = \begin{bmatrix} \frac{\partial\mathcal{L}^2}{\partial\alpha_1^2} & \frac{\partial\mathcal{L}^2}{\partial\alpha_1\partial\alpha_2} & \frac{\partial\mathcal{L}^2}{\partial\alpha_1\partial\alpha_4} \\ \frac{\partial\mathcal{L}^2}{\partial\alpha_2\partial\alpha_1} & \frac{\partial\mathcal{L}^2}{\partial\alpha_2^2} & \frac{\partial\mathcal{L}^2}{\partial\alpha_2\partial\alpha_4} \\ \frac{\partial\mathcal{L}^2}{\partial\alpha_4\partial\alpha_1} & \frac{\partial\mathcal{L}^2}{\partial\alpha_4\partial\alpha_2} & \frac{\partial\mathcal{L}^2}{\partial\alpha_4^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.19}$$

So the SONC equation gets the following relationship:

$$\boldsymbol{y}^T \cdot \nabla^2_{\boldsymbol{\alpha}}\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\mu}) \cdot \boldsymbol{y} \geqslant \mathbf{0}, \tag{4.20}$$

proving solution (4.17) is feasible [74].

In this way, the maximized $\Pr[A]$ is given by

$$\begin{aligned} \Pr[A] &= \Pr[|h|^2 \geqslant \frac{\gamma_0(\gamma_0^2+3\gamma_0+3)}{\rho_1}] \\ &= \exp(-\frac{\gamma_0(\gamma_0^2+3\gamma_0+3)}{\rho_1\sigma_1^2}). \end{aligned} \tag{4.21}$$

For power allocation ratios in the second phase, generally we should solve the equation $\frac{\partial \Pr[B]}{\partial\beta_F} = 0$ to figure out possible optimal β_F. Due to many exponential components in $\frac{\partial \Pr[B]}{\partial\beta_F}$, it is hard to obtain the explicit expression of optimal β_F. However, we find it very easy to exhaustively search possible β_F values within $\frac{1+\gamma_0}{2+\gamma_0} \leqslant \beta_F < 1$ to find out the optimal β_F. Then the β_N is obtained by $\beta_N = 1 - \beta_F$.

4.3.2 Ergodic Capacity Performance

In the field of wireless communications, the ergodic capacity is a critical metric for evaluating the system performance. It is defined in [75] that "The ergodic capacity is the maximum mutual information between the input and output if the code spans an infinite number of independent realizations of the channel matrix". In other words, the ergodic capacity is the average theoretical channel capacity over time if the capacity is estimated over a fading channel.

It is hard to obtain the exact expressions of ergodic capacity in CNAR system. So ergodic sum capacity approximations in high ρ_1 and ρ_2 regime are given by analyzing the achievable rate for each message as follows.

In the first-phase transmission for s_1, by using $\int_0^\infty \log_2(1+\mu x) f_X(x)dx = \frac{\mu}{\ln 2}\int_0^\infty \frac{1-F_X(x)}{1+\mu x}dx$, we obtain the achievable rate of BS-MT1 channel as

$$\begin{aligned} E[C_1] &= \int_0^\infty \log_2(1+\rho_1\alpha_1 x) f_{|h|^2}(x)dx \\ &= \frac{\rho_1\alpha_1}{\ln 2}\int_0^\infty \frac{\exp(-\frac{x}{\sigma_1^2})}{1+\rho_1\alpha_1 x}dx \\ &\overset{(a)}{=} -\frac{1}{\ln 2} e^{\frac{1}{\rho_1\alpha_1\sigma_1^2}} \mathrm{Ei}(-\frac{1}{\rho_1\alpha_1\sigma_1^2}) \\ &\overset{(b)}{\approx} -\frac{1}{\ln 2}(1+\frac{1}{\rho_1\alpha_1\sigma_1^2})[C+\ln(\frac{1}{\rho_1\alpha_1\sigma_1^2}) - \frac{1}{\rho_1\alpha_1\sigma_1^2}] \end{aligned} \tag{4.22}$$

where $\mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{e^t}{t}dt, x<0$ is the exponential integral function. (a) is obtained according to Eq. (3.352.4) in [76]. Approximation (b) is made by applying $e^{-x} \underset{x\to 0}{\sim} 1-x$ and $\mathrm{Ei}(x) \underset{x\to 0}{\sim} [C+\ln(-x)+x]$ where C denotes the Euler constant in high ρ_1 regime.

According to (4.6), we calculate $E[C_2]$ by discussing $\min\{SNR_{2,1}, SNR_{2,2}\}$ as follows. For $SNR_{2,1}$, when ρ_1 is large, $SNR_{2,1} \approx \frac{\alpha_2}{\alpha_1} = 1+\gamma_0$. For $SNR_{2,2}$ in high ρ_2 regime, if MT2 is far

user, it holds that $SNR_{2,2} = SNR_{F,2} \approx \frac{\beta_F}{\beta_N} > SNR_{2,1}$, leading to $\min\{SNR_{2,1}, SNR_{2,2}\} = SNR_{2,1}$. And if MT2 is near user with the condition $SNR_{N,2} = \beta_N \rho_2 \|\mathbf{g}_2\|_2^2 > SNR_{2,1}$, $\min\{SNR_{2,1}, SNR_{2,2}\}$ also equals to $SNR_{2,1}$. However, if MT2 is near user conditioned on $\beta_N \rho_2 \|\mathbf{g}_2\|_2^2 \leqslant SNR_{2,1}$, $\min\{SNR_{2,1}, SNR_{2,2}\}$ becomes $\beta_N \rho_2 \|\mathbf{g}_2\|_2^2$. The above cases for the discussion are listed in Table 4.1.

Table 4.1: Table for value of $\min\{SNR_{2,1}, SNR_{2,2}\}$ under different conditions

MT2 Status	Condition	$\min\{SNR_{2,1}, SNR_{2,2}\}$
Far user	None	$SNR_{2,1}$
Near user	$\beta_N \rho_2 \|\mathbf{g}_2\|_2^2 > SNR_{2,1}$	$SNR_{2,1}$
Near user	$\beta_N \rho_2 \|\mathbf{g}_2\|_2^2 \leqslant SNR_{2,1}$	$\beta_N \rho_2 \|\mathbf{g}_2\|_2^2$

Define Event D = {MT2 is near user with the condition $\beta_N \rho_2 \|\mathbf{g}_2\|_2^2 \leqslant SNR_{2,1}$}. We prove in the appendix that when ρ_2 is large, $\Pr[D] \approx 0$. So Event D seldom happens, which indicates

$$E[C_2] \approx \log_2(1 + SNR_{2,1}) = \log_2(2 + \gamma_0). \tag{4.23}$$

Table 4.2: A list for value of $\min\{SNR_{3,1}, SNR_{3,2}\}$ in different conditions

MT3 Status	Condition	$\min\{SNR_{3,1}, SNR_{3,2}\}$
Far user	None	$SNR_{3,1}$
Near user	$\beta_N \rho_2 \|\mathbf{g}_3\|_2^2 > SNR_{3,1}$	$SNR_{3,1}$
Near user	$\beta_N \rho_2 \|\mathbf{g}_3\|_2^2 \leqslant SNR_{3,1}$	$\beta_N \rho_2 \|\mathbf{g}_3\|_2^2$

Next, to calculate $E[C_3]$ based on (4.7), the discussion on $\min\{SNR_{3,1}, SNR_{3,2}\}$ is also required. Since $SNR_{F,2} > SNR_{3,1} \approx \frac{(1+\gamma_0)^2}{2+\gamma_0}$ in high ρ_1 and ρ_2 regime, the approximation method used for $E[C_2]$ is also applicable with similar calculating process. A similar discussion is provided in Table 4.2. So we obtain

$$E[C_3] \approx \log_2(1 + SNR_{3,1}) = \log_2(\tfrac{\gamma_0^2+3\gamma_0+3}{2+\gamma_0}). \tag{4.24}$$

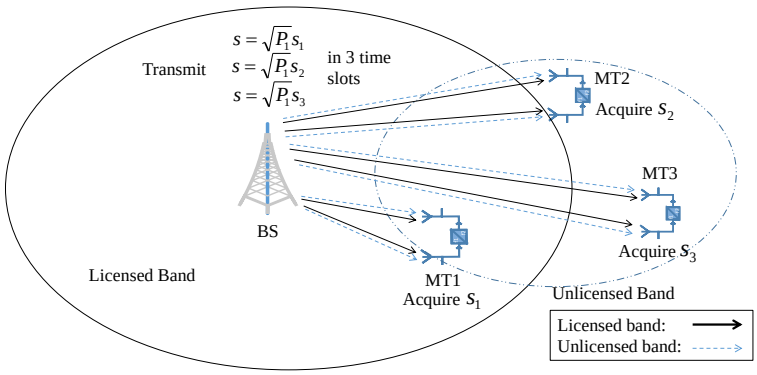

Figure 4.4: Illustration of conventional OMA system.

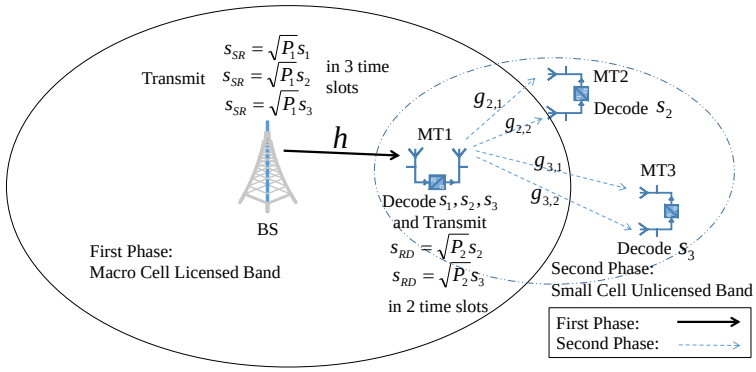

Figure 4.5: Depiction of OMA-based relaying system.

Therefore, in high ρ_1 and ρ_2 regime, the ergodic sum capacity for CNAR system is approximated as

$$\begin{aligned} &E[C_1] + E[C_2] + E[C_3] \\ &\approx -\frac{1}{\ln 2}(1 + \frac{1}{\rho_1\alpha_1\sigma_1^2})[C + \ln(\frac{1}{\rho_1\alpha_1\sigma_1^2}) - \frac{1}{\rho_1\alpha_1\sigma_1^2}] + \log_2(\gamma_0^2 + 3\gamma_0 + 3). \end{aligned} \tag{4.25}$$

4.4 Numerical Results

We compare the outage probability and ergodic capacity obtained by CNAR system, conventional OMA system and the relaying system based on OMA. In all systems, for the purpose of illustration, we suppose the BS and all MTs are on a straight line. Assume BS-MT1 dis-

tance d_1 = 55m, MT1-MT2 distance d_2 = 55m and MT1-MT3 distance d_3 = 82.5m. Suppose $\sigma_i^2 = (\frac{d_i}{\mathcal{D}_0})^{-\theta}$ where the reference distance $\mathcal{D}_0$ is 100m and the path loss exponent θ is 4. So $\sigma_1 = \sigma_2 = 3.31$ and $\sigma_3 = 1.47$.

In the conventional OMA system, for fairness comparison, we set the BS transmit SNR as $(\rho_1 + \rho_2)$. The BS transmits s_1, s_2 and s_3 to each MT using $\frac{1}{3}$ time resource, respectively. Each MT receives signal from lincensed and unlicensed band at the same time by double antennas and uses MRC for receive signal processing, which is depicted in Fig. 4.4. In the OMA-based relaying system, we assume that s_1, s_2 and s_3 are transmitted using $\frac{1}{3}$ time resource in the S-R link, respectively; s_2 and s_3 are sent using half time resource in the R-D link, respectively. Other assumptions are the same as CNAR system. The OMA-based relaying system is illustrated in Fig. 4.5. Specially, the outage probability of conventional OMA system is calculated by taking the average of the three MTs' outage probabilities in Monte-Carol simulations. This reason is the three MTs communicate with the BS separately in OMA, making the outage behavior considered in an individual manner. But for relaying systems, the three MTs influence the communication quality of each other. So it's better to consider the outage behavior from the perspective of a group.

Fig. 4.6 shows comparison of outage probability as a function of target rate R_0 when ρ_2 = 15dB. One can observe that the theoretical analysis is verified by the simulation results. For lower value R_0, the outage probability for conventional OMA is the lowest. But it grows significantly as R_0 increases, which is not suitable for 5G as the data rate requirement is higher in the future 5G. Moreover, the outage probability of CNAR system is much lower than the OMA-based system for different R_0 values.

In Fig. 4.7 and Fig. 4.8, we plot outage probability with respect to ρ_1 and ρ_2, respectively

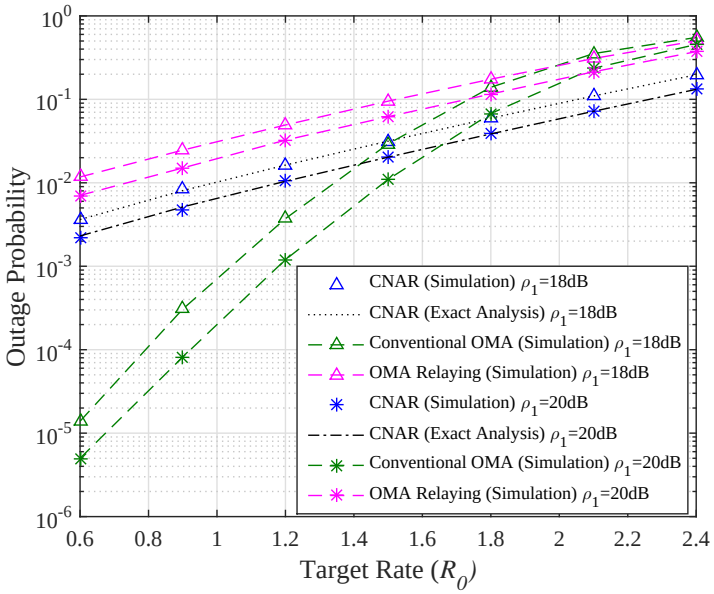

Figure 4.6: Performance of Outage probability as a function of target rate R_0 when $\rho_2 = 15$dB.

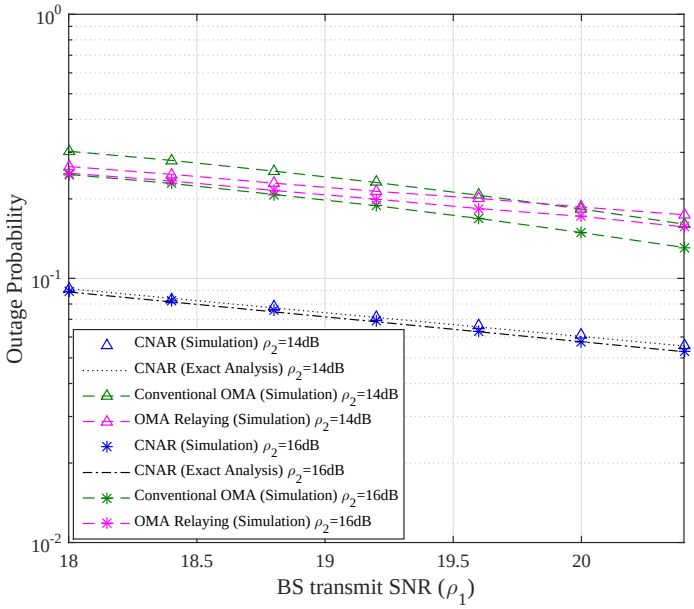

Figure 4.7: Outage probability as a function of BS transmit SNR ρ_1 when $R_0 = 2$.

when $R_0 = 2$. Here again, our theoretical derivations are validated and the CNAR system has much lower outage probability than conventional OMA and OMA-based relaying system. Moreover, in Fig. 4.7 with ρ_1 increasing, the outage probability for CNAR doesn't decrease significantly. The reason is according to Eq. (4.8), when large ρ_1 makes $\Pr[A] \approx 1$, P_{out} is lower-bounded by $1 - \Pr[B]$, which is determined by the MT1 transmit SNR and target rate. Similarly, the outage probability for CNAR doesn't reduce obviously in Fig. 4.8 as ρ_2 increases. The reason is similar to the one in Fig. 4.8. As a result, in high SNR regime, the choice of target rate has much more influence than BS and MT1 transmit power. This phenomenon can be considered in a practical view. In the past several years when the target rate is in low level, (where the services are mainly texts and online music), the traditional OMA is enough for required communication quality. But when the target rate is higher currently and in the future, the relaying systems have much performance gain over traditional OMA when the total power, spectrum and antennas are the same.

Fig. 4.9 presents the capacity results for CNAR and OMA-based system when the target rate $R_0 = 2$ and $\rho_1 = 19$dB. The figure shows that the approximations of the capacity in CNAR system well match the simulation results. One can also observe that the simulation curve for C_2 is not as close to its approximation as the curve for C_3. The reason is as MT2 is nearer to MT1, MT2 is more possible to be near user. When Event D occurs, $\log_2(1 + \beta_N \rho_2 \|\mathbf{g}_2\|_2^2) \leqslant \log_2(1 + SNR_{2,1})$ lower the capacity from the approximation to some extent. And our power allocation scheme makes the ergodic capacity of each CNAR user above the target rate.

From Fig. 4.10, CNAR has sum capacity gain over the OMA-based one. This is because NOMA enables each user to exploit all time resource while OMA limits the time resource that each user can use. Moreover, it's easier to maintain the receive SNR at cell-edge users

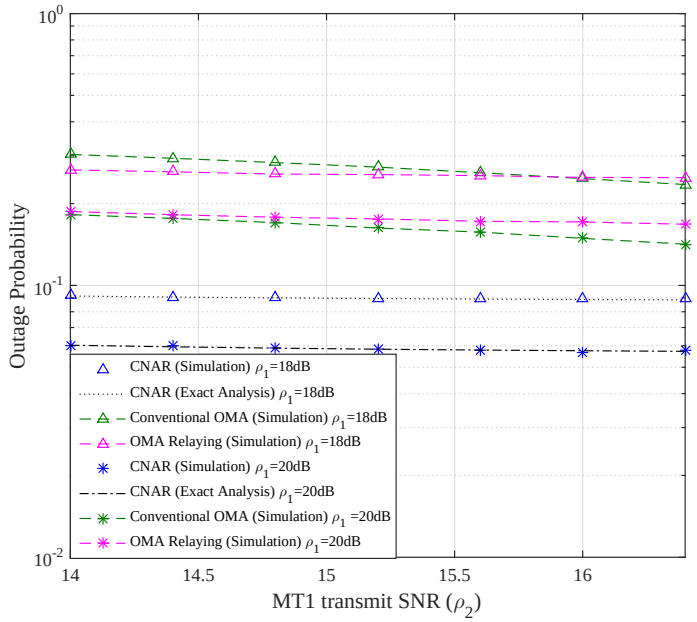

Figure 4.8: Outage probability as a function of MT1 transmit SNR ρ_2 when $R_0 = 2$.

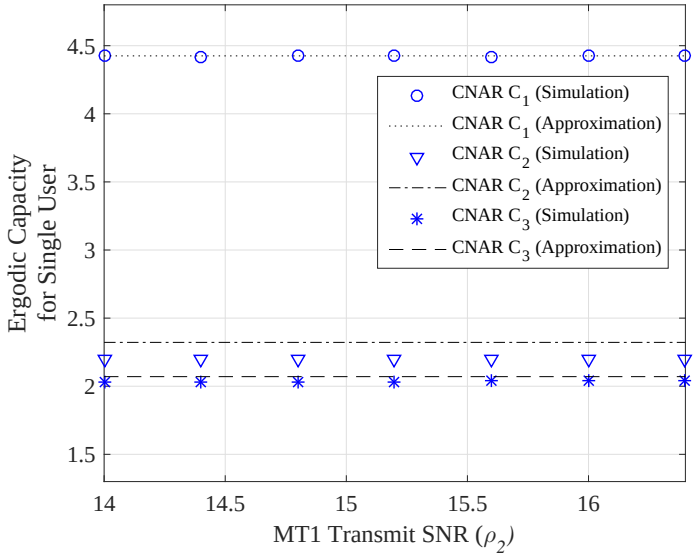

Figure 4.9: Single-user ergodic capacity as a function of MT1 transmit SNR ρ_2 when $R_0 = 2$ and $\rho_1 = 19$dB.

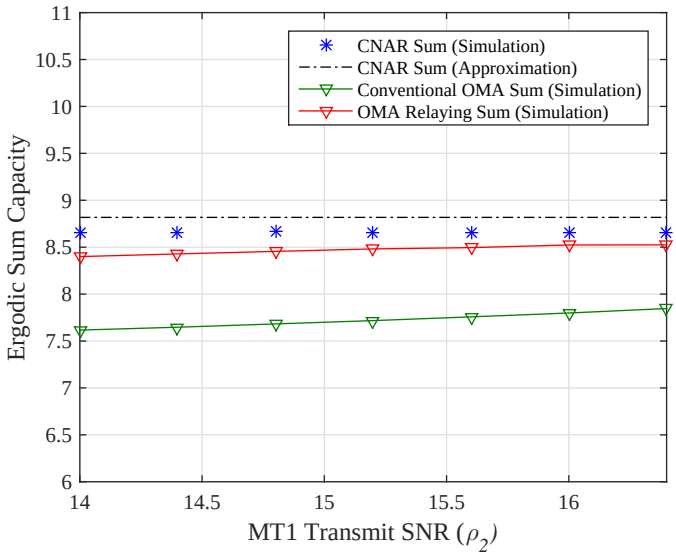

Figure 4.10: Sum ergodic capacity as a function of MT1 transmit SNR ρ_2 when $R_0 = 2$ and $\rho_1 = 19$dB.

as a constant based on power ratios in NOMA, while conventional OMA causes their receive SNR reduced due to the poor channel quality in long-distance transmission. According to the tendency of curves, sum capacity of OMA-based system increases a little bit faster than CNAR system, which indicates under extremely high transmit SNR, the sum capacity of CNAR system will be lower than OMA. However, the practical situation where transmit power is limited is suitable for CNAR to outperform OMA-based relaying system.

4.5 Chapter Summary

This chapter proposes a CNAR system in 5G to serve multiple cell-edge users concurrently with data rate guarantee. To characterize the system performance, we derive the exact system outage probability by analyzing the outage behavior in the S-R and R-D links separately. Then we provide the optimal power allocation scheme to guarantee the data rate by minimizing the

outage probability. Additionally, we give the approximation of ergodic sum capacity in high SNR regime from the discussion on the interference at cell-edge users based on NOMA principles. Numerical results validate the theoretical analysis. And they demonstrate that though conventional OMA has low outage probability when the target rate is very small, our CNAR has better outage and capacity performance than conventional OMA and OMA-based relaying system at high target rate for the future 5G networks.

Chapter 5

Conclusions

5.1 Thesis Summary

In this thesis, the application of resource efficient techniques, i.e., massive MIMO, NOMA and relaying technology, in the future 5G has been investigated mainly from three aspects. They involves a general introduction of these techniques, following problems and literature survey of existing solutions, our proposed algorithms for efficient antenna selection and user scheduling in 5G massive MIMO-NOMA system, and a design along with performance a analysis of CNAR system.

In particular, firstly, the principles of massive MIMO, NOMA to achieve high spectral efficiency and relaying technology to achieve power efficiency are introduced. Following this, relevant problems, including antenna selection, user scheduling, power allocation and relaying scheme design, are described with the necessity to be solve. A literature survey on currently available approaches is also provided.

Next, the antenna selection and user scheduling strategies in massive MIMO-NOMA sys-

tem is investigated. With power allocation scheme among NOMA users figured out as the basis, efficient antenna selection algorithm is proposed for single-band scenario. It searches desired antennas from candidate antennas beneficial to relevant users. The joint AU contribution algorithm is proposed for multi-band multi-user scenario, which considers the contribution of each antenna's and user's channel gain to total channel gain jointly. Simulation results show that efficient search algorithm reaches near-optimal performance, and joint AU contribution algorithm achieves performance similar to existing methods with reduces complexity. These algorithms achieve high-level inter-user orthogonality with maintaining system stability.

Furthermore, a collaborative NOMA assisted relaying system is proposed for serve multiple cell-edge users simultaneously with data rate guarantee. The system is featured by the collaboration of S-R and R-D links to provide multiple input and output for the relay. Then with analysis of the outage behavior in the S-R and R-D links separately, we derive the system outage probability, by minimizing which the optimal power allocation ratios are obtained. To further characterize the system performance, we analyze the sum ergodic capacity by discussing the interference based on NOMA protocol. Simulation results validate the mathematical analysis and demonstrate that the proposed relaying system assisted by NOMA has better outage and capacity performance than OMA.

5.2 Future Works

Some aspects of the application of the abovementioned 5G technologies still remain uncovered in this thesis. They are worthy further consideration and exploration. Some aspects are listed as follows:

- In the massive MIMO-NOMA system, we focus on the power allocation scheme among single NOMA user pair in each subband, with the assumption that the same power is allocated to each subband. The global power allocation strategy requires future investigation to improve the system capacity.

- We merely consider one user group forming a R-D connection in our CNAR system, but massive users need to be served in practical 5G situation. As a result, the strategies to designate certain user devices as relaying devices and assign cell-edge users to them are worth investigating.

- For capacity maximization, some far users in the massive MIMO-NOMA system are not scheduled for communications in certain time slots. However, they can possibly be scheduled if assisted by relaying devices. Hence, potential combination of massive MIMO-NOMA and CNAR systems to improve the communication quality can be considered.

Bibliography

[1] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selen, and J. Skold. "5G wireless access: requirements and realization". *IEEE Commun. Mag.*, 52(12):42–47, Dec. 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J.C. Zhang. "What will 5G be?". *IEEE J. Sel. Areas Commun*, 32(6):1065–1082, Jun. 2014.

[3] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah. "Massive MIMO systems with non-ideal hardware: Energy effciency, estimation, and capacity limits". *IEEE Trans. Inf. Theory*, 60(11):7112–7139, Nov. 2014.

[4] C. Lai, H. Li, X. Liang, R. Lu, K. Zhang, and X. Shen. "CPAL: A conditional privacy-preserving authentication with access linkability for roaming service". *IEEE Internet Things J.*, 1(1):2327–4662, Feb. 2014.

[5] M. Condoluci, M. Dohler, G. Araniti, A. Molinaro, and J. Sachs. "Enhanced radio access and data transmission procedures facilitating industry-compliant machine-type communications over LTE-based 5G networks". *IEEE Wireless Commun.*, 23(1):56–63, Feb. 2016.

[6] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu. "5G network capacity: Key elements and technologies". *IEEE Trans. Veh. Technol.*, 9(1):71–78, Jan. 2014.

[7] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler, and Y. Koucheryavy. "Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap". *IEEE Commun. Mag.*, 53(9):32–40, Sep. 2015.

[8] S. Scott-Hayward and E. Garcia-Palacios. "Multimedia resource allocation in mmwave 5G networks". *IEEE Commun. Mag.*, 53(1):240–247, Jan. 2015.

[9] T. Bai, A. Alkhateeb., and R. W. Heath. "Coverage and capacity of millimeter-wave cellular networks". *IEEE Commun. Mag.*, 52(9):70–77, Sep. 2014.

[10] Z. Pi and F. Khan. "An introduction to millimeter-wave mobile broadband systems". *IEEE Commun. Mag.*, 49(6):101–107, Jun. 2011.

[11] L. Lei, Z. Zhong, C. Lin, and X. Shen. "Operator controlled device-to-device communications in LTE-advanced networks". *IEEE Wireless Commun.*, 19(3):96C104, Jun. 2012.

[12] S. Liu, Z. Luo, Y. Liu., and J. Gao. "Spreading code design for downlink space-time-frequency spreading CDMA". *IEEE Trans. Veh. Technol.*, 57(5):2933–2946, Feb. 2008.

[13] L. Dai, B. Wang, Y. Yuan, S. Han, C. I, and Z. Wang. "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends". *IEEE Commun. Mag.*, 53(9):74–81, Sep. 2015.

[14] N. Hu, Y. Yao, and Z. Yang. "Analysis of cooperative TDMA in rayleigh fading channels". *IEEE Trans. Veh. Technol.*, 62(3):1158–1168, Nov. 2012.

[15] N. Gupta and A.K. Jagannatham. "Multiuser successive maximum ratio transmission (MS-MRT) for video quality maximization in unicast and broadcast MIMO OFDMA-based 4G wireless networks". *IEEE Trans. Veh. Technol.*, 63(7):3147–3156, Jan. 2014.

[16] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. "Massive MIMO for next generation wireless systems". *IEEE Commun. Mag.*, 52(2):186–195, Feb. 2014.

[17] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, and Z. Turanyi. "Design aspects of network assisted device-to-device communications". *IEEE Commun. Mag.*, 50(3):170–177, Mar. 2012.

[18] J. Deng, A. A. Dowhuszko, R. Freij, and O. Tirkkonen. "Relay selection and resource allocation for D2D-relaying under uplink cellular power control". *Proc. of IEEE GC Wkshps*, pages 1–6, Dec. 2015.

[19] Q. Zhao, Y. Mao, S. Leng, and H. Wang. "Multimedia traffic placement under 5G radio access techniques in indoor environments". *Proc. of IEEE ICC*, pages 3891–3896, Jun. 2015.

[20] N. B. Mehta, S. Kashyap, and A. F. Molisch. "Antenna selection in LTE: from motivation to specification". *IEEE Commun. Mag.*, 50(10):144–150, Oct. 2012.

[21] A. Al-Dulaimi, S. Al-Rubaye, Q. Ni, and E. Sousa. "5G communications race: Pursuit of more capacity triggers LTE in unlicensed band". *IEEE Trans. Veh. Mag.*, 10(1):43–51, Mar. 2015.

[22] Z. Arslan, M. Erel, Y. Ozcevik, and B. Canberk. "SDoff: A software-defined offloading controller for heterogeneous networks". *Proc. of IEEE WCNC*, pages 2827–2832, Apr. 2014.

[23] A. Sniady, M. Sonderskov, and J. Soler. "VoLTE performance in railway scenarios: Investigating VoLTE as a viable replacement for GSM-R". *IEEE Veh. Technol. Mag.*, 10 (3):60–70, Jul. 2015.

[24] W. Nam, D. Bai, J. Lee, and I. Kang. "Advanced interference management for 5G cellular networks". *IEEE Commun. Mag.*, 52(5):52–60, May 2014.

[25] A. Vallejo, A. Zaballos, J. Selga, and J. Dalmau. "Next-generation QoS control architectures for distribution smart grid communication networks". *IEEE Commun. Mag.*, 50(5): 128–134, May 2012.

[26] D. Huang, B. He, and C. Miao. "A survey of resource management in multi-tier web applications". *IEEE Commun. Surveys Tuts.*, 16(3):1574–1590, Jan. 2014.

[27] T. Liu, C. Yang, and L. Yang. "A unified analysis of spectral efficiency for two-hop relay systems with different resource configurations". *IEEE Trans. Veh. Technol.*, 62(3): 3137–3148, Sep. 2013.

[28] R. Dou and G. Nan. "Optimizing sensor network coverage and regional connectivity in industrial IoT systems". *IEEE Syst. J.*, PP(99):1–10, Jun. 2015.

[29] Q. Wang, G. Lim, L. J. Cimini, L. J. Greenstein, D. S. Chan, and A. Hedayat. "Quantifying and comparing energy efficiencies on SU-MIMO and MU-MIMO downlinks". *Proc. of IEEE GLOBECOM*, pages 1–6, Dec. 2015.

[30] D. Tse and P. Viswanat. *Fundamentals of Wireless Communication*. 1st ed. Cambridge, U. K.: Cambridge Univ. Press, 2005.

[31] Y. Jang, K. Min, S. Park, and S. Choi. "Spatial resource utilization to maximize uplink spectral efficiency in full-duplex massive MIMO". *Proc. of IEEE ICC*, pages 1583–1588, Jun. 2015.

[32] T. M. Kim, A. Ghaderipoor, and A. Paulraj. "Antenna selection and power combining for transmit beamforming in MIMO systems". *Proc. of IEEE GLOBECOM*, pages 4600–4605, Dec. 2012.

[33] R. M. Radaydeh and M. Alouini. "On the performance of arbitrary transmit selection for threshold-based receive MRC with and without co-channel interference". *IEEE Trans. Commun.*, 59(11):3177–3191, Oct. 2011.

[34] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. 1st ed. Cambridge, U. K.: Cambridge Univ. Press, 2008.

[35] Q. Nadeem, A. Kammoun, M. Debbah, and M. Alouini. "3D massive MIMO systems: Modeling and performance analysis". *IEEE Trans. Wireless Commun.*, 14(12):6926–6939, Jul. 2015.

[36] A. Chockalingam and B. S. Rajan. *Large MIMO Systems*. Cambridge, U. K.: Cambridge Univ. Press, 2014.

[37] C. He and R. D. Gitlin. "Limiting performance of massive MIMO downlink cellular systems". *Proc. of Information Theory and Application Workshop*, pages 1–6, Feb. 2016.

[38] A. Li, A. Benjebbour, X. Chen, H. Jiang, and H. Kayama. "Investigation on hybrid automatic repeat request (HARQ) design for NOMA with SU-MIMO". *Proc. of IEEE PIMRC*, pages 590–594, Aug. 2015.

[39] X. Chen, A. Benjebbour, A. Li, and A. Harada. "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)". *Proc. of IEEE VTC Spring*, pages 1–5, May 2014.

[40] P. Patcharamaneepakorn, A. Doufexi, and S. Armour. "Reduced complexity joint user and receive antenna selection algorithms for SLNR-based precoding in MU-MIMO systems". *Proc. of IEEE VTC Spring*, pages 1–5, May 2012.

[41] M. Torabi and D. Haccoun. "Performance analysis of joint user scheduling and antenna selection over MIMO fading channels". *IEEE Signal Process. Lett.*, 18(4):235–238, Feb. 2011.

[42] S. Shi, L. Yang, and H. Zhu. "Outage balancing in downlink non-orthogonal multiple access with statistical channel state information". *IEEE Trans. Wireless Commun.*, PP (99):1, Mar. 2016.

[43] Z. Ding, P. Fan, and H. V. Poor. "Impact of user pairing on 5G non-orthogonal multiple access". *IEEE Trans. Veh. Technol.*, PP(99):1, Sep. 2015.

[44] J. Shin and J. Jeong. "Improved outage probability of indoor PLC system for multiple users using resource allocation algorithms". *IEEE Trans. Power Del.*, 28(4):2228–2235, Jul. 2013.

[45] I. Ahmed and A. Mohamed. "Outage optimal resource allocation for two-hop multiuser multirelay cooperative communication in OFDMA upstream". *Proc. of IEEE VTC Spring*, pages 1–6, May 2011.

[46] X. Cai, J. Zheng, Y. Zhang, and H. Murata. "A capacity oriented resource allocation algorithm for device-to-device communication in mobile cellular networks". *Proc. of IEEE ICC*, pages 2233–2238, Jun. 2014.

[47] P. Zhang, S. Chen, and L. Hanzo. "Two-tier channel estimation aided near-capacity MIMO transceivers relying on norm-based joint transmit and receive antenna selection". *IEEE Trans. Wireless Commun.*, 14(1):122–137, Jul. 2014.

[48] M. Benmimoune, E. Driouch, W. Ajib, and D. Massicotte. "Joint transmit antenna selection and user scheduling for massive MIMO systems". *Proc. of IEEE WCNC*, pages 381–386, Mar. 2015.

[49] J. Kim and I. Lee. "Non-orthogonal multiple access in coordinated direct and relay transmission". *IEEE Commun. Lett.*, 19(11):2037–2040, Nov. 2015.

[50] Z. Ding, M. Peng, and H. Vincent Poor. "Cooperative non-orthogonal multiple access in 5G systems". *IEEE Commun. Lett.*, 19(8):1462–1465, Aug. 2015.

[51] J. Choi. "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems". *IEEE Trans. Commun.*, 63(3): 791–800, Jan. 2015.

[52] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson. "The role of small cells, coordinated multipoint, and massive MIMO in 5G". *IEEE Commun. Mag.*, 52(5):44–51, May 2014.

[53] X. Gao, O. Edfors, J. Liu, and F. Tufvesson. "Antenna selection in measured massive MIMO channels using convex optimization". *Proc. of IEEE GC Wkshps*, pages 129–134, Dec. 2013.

[54] M. Hanif, M. Juntti, and L. Tran. "Antenna selection with erroneous covariance matrices under secrecy constraints". *IEEE Trans. Veh. Technol.*, 65(1):414–420, Jan. 2015.

[55] P. V. Amadori and C. Masouros. "Power efficient massive MU-MIMO via antenna selection for constructive interference optimization". *Proc. of IEEE ICC*, pages 1607–1612, Jun. 2015.

[56] A. Benjebbour, Y. Kishiyama A. Li, H. Jiang, and T. Nakamura. "System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements". *Proc. of IEEE GC Wkshps*, pages 706–710, Dec. 2014.

[57] B. Kim, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong. "Non-orthogonal multiple access in a downlink multiuser beamforming system". *Proc. of IEEE MILCOM*, pages 1278 – 1283, Nov. 2013.

[58] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis. "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems". *IEEE Trans. Signal Process.*, 64(1):76–88, Sep. 2015.

[59] J. G. Andrews, A. Ghosh, and R. Muhamed. *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2007.

[60] W. Nam, D. Bai, J. Lee, and I. Kang. "Advanced interference management for 5G cellular networks". *IEEE Commun. Mag.*, 52(5):52–60, May 2014.

[61] X. Ge, H. Cheng, M. Guizani, and T. Han. "5G wireless backhaul networks: challenges and research advances". *IEEE Netw.*, 28(6):6–11, Nov. 2014.

[62] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi. "Non-orthogonal multiple access (NOMA) for cellular future radio access". *Proc. of IEEE VTC Spring*, pages 1–5, Jun. 2013.

[63] X. Chen, A. Benjebboui, Y. Lan, A. Li, and H. Jiang. "Evaluations of downlink non-orthogonal multiple access (NOMA) combined with SU-MIMO". *Proc. of IEEE PIMRC*, pages 1887 – 1891, Sep. 2014.

[64] J. Men and J. Ge. "Non-orthogonal multiple access for multiple-antenna relaying networks". *IEEE Commun. Lett.*, 19(10):1686–1689, Oct. 2015.

[65] M. Lauridsen and I. Lee. "Capacity analysis of cooperative relaying systems using non-orthogonal multiple access". *IEEE Commun. Lett.*, 19(11):1949–1952, Nov. 2015.

[66] S. Rini, L. Ghaghanidze, E. Kurniawan, and A. Goldsmith. "Rate optimization for relay-assisted downlink cellular systems using superposition coding". *Proc. of IEEE ICC*, pages 5371–5375, Jun. 2013.

[67] Z. Ding, H. Dai, and H. V. Poor. "Relay selection for cooperative NOMA". *IEEE Commun. Lett.*, PP(99):1, Jun. 2016.

[68] J. Men and J. Ge. "Performance analysis of non-orthogonal multiple access in downlink cooperative network". *IET Commun.*, 9(18):2267–2273, Dec. 2015.

[69] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor. "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer". *IEEE J. Sel. Areas Commun.*, 34(4):938–953, Mar. 2016.

[70] F. Liu, P. Mahonen, and M. Petrova. "Proportional fairness-based user pairing and power allocation for non-orthogonal multiple access". *Proc. of IEEE PIMRC*, pages pp. 1127–1131, Aug. 2015.

[71] M. R. Bhatnagar. "On the capacity of decode-and-forward relaying over Rician fading channels". *IEEE Commun. Lett.*, 17(6):1100–1103, Jun. 2013.

[72] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, U. K.: Cambridge Univ. Press, 2004.

[73] A. Takayama. *Analytical Methods in Economics*. 839 Greene Street Ann Arbor, MI, USA: University of Michigan Press, 1994.

[74] T. Edgar, D. Himmelblau, and L. Lasdon. *Optimization of Chemical Processes*. 2nd ed. New York City, U. S.: McGraw-Hill, 2001.

[75] H. Jafarkhani. *Space-Time Coding: Theory and Practice*. 1st ed. Cambridge, U. K.: Cambridge Univ. Press, 2005.

[76] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. 7th ed. New York, NY, USA: Academic, 2007.

Appendix A

Proofs of Equations for Performance Analysis in CNAR System

Proof Proof of Pr[B] With the probability density function (PDF) for Rayleigh channel, the cumulative distribution functions (CDF) of $\|\mathbf{g}_2\|_2^2$ can be obtained by

$$\begin{aligned} F_{\|\mathbf{g}_2\|_2^2}(x) = &\int_0^x \int_0^{x-x_2} \frac{1}{\sigma_2^2} \exp(-\frac{x_1}{\sigma_2^2}) \times \frac{1}{\sigma_2^2} \exp(-\frac{x_2}{\sigma_2^2}) dx_1 dx_2 \\ =&1 - \frac{x}{\sigma_2^2} \exp(-\frac{x}{\sigma_2^2}) - \exp(-\frac{x}{\sigma_2^2}), \end{aligned} \tag{A.1}$$

Similarly, the CDF of $\|\mathbf{g}_2\|_2^2$ is calculated as

$$F_{\|\mathbf{g}_3\|_2^2}(x) = 1 - \frac{x}{\sigma_3^2} \exp(-\frac{x}{\sigma_3^2}) - \exp(-\frac{x}{\sigma_3^2}). \tag{A.2}$$

So we get the following PDF of $\|\mathbf{g}_2\|_2^2$ and $\|\mathbf{g}_3\|_2^2$ by:

$$\begin{aligned} f_{\|\mathbf{g}_2\|_2^2}(x) &= F'_{\|\mathbf{g}_2\|_2^2}(x) = \frac{x}{\sigma_2^4}\exp(-\frac{x}{\sigma_2^2}), \\ f_{\|\mathbf{g}_3\|_2^2}(x) &= F'_{\|\mathbf{g}_3\|_2^2}(x) = \frac{x}{\sigma_3^4}\exp(-\frac{x}{\sigma_3^2}). \end{aligned} \tag{A.3}$$

Define Event F=$\{\|\mathbf{g}_2\|_2^2 > \|\mathbf{g}_3\|_2^2, \|\mathbf{g}_2\|_2^2 > w_N, \|\mathbf{g}_3\|_2^2 > w_F\}$ and Event G=$\{\|\mathbf{g}_3\|_2^2 > \|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2 > w_N, \|\mathbf{g}_2\|_2^2 > w_F\}$. Pr[$B$] is calculated by

$$\begin{aligned} \Pr[B] &= \Pr[\max\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} > w_N, \min\{\|\mathbf{g}_2\|_2^2, \|\mathbf{g}_3\|_2^2\} > w_F] \\ &= \Pr[F \cup G] \\ &= \int_{w_F}^{\infty}\int_{w_F}^{\infty} f_{\|\mathbf{g}_2\|_2^2}(x_1) f_{\|\mathbf{g}_3\|_2^2}(x_2) dx_1 dx_2 - \int_{w_F}^{w_N}\int_{w_F}^{w_N} f_{\|\mathbf{g}_2\|_2^2}(x_1) f_{\|\mathbf{g}_3\|_2^2}(x_2) dx_1 dx_2 \\ &= \phi(w_N, \sigma_2)\phi(w_F, \sigma_3) + \phi(w_F, \sigma_2)\phi(w_N, \sigma_3) - \phi(w_N, \sigma_2)\phi(w_N, \sigma_3) \end{aligned}$$

Proof Proof of Logarithmic Equation 1

$$\begin{aligned} &\int_0^{\infty} \log_2(1+\lambda x) f_X(x) dx \\ &= \int_0^{\infty} \log_2(1+\lambda x) dF_X(x) \\ &= \log_2(1+\lambda x) F_X(x)|_{x=\infty} - \int_0^{\infty} F_X(x) d[\log_2(1+\lambda x)] \\ &\overset{(e)}{=} \frac{\lambda}{\ln 2}\int_0^{\infty} \frac{1}{1+\lambda x} dx - \frac{\lambda}{\ln 2}\int_0^{\infty} \frac{F_X(x)}{1+\lambda x} dx \end{aligned} \tag{A.4}$$

where (e) is obtained by $F_X(x)|_{x=\infty} = 1$ and $\log_2(1+\lambda x)|_{x=\infty} = \frac{\lambda}{\ln 2}\int_0^{\infty}\frac{1}{1+\lambda x}dx$.

Proof Proof of Pr[D]

$$\Pr[D] = \int_0^{SNR_{2,1}} \int_0^x f_{\beta_N\rho_2\|\mathbf{g}_2\|_2^2}(x) f_{\beta_N\rho_2\|\mathbf{g}_3\|_2^2}(y) dxdy \tag{A.5}$$

$$\begin{aligned}
=&1 - e^{-\frac{SNR_{2,1}}{\beta_N\rho_2\sigma_2^2}} - \frac{SNR_{2,1}}{\beta_N\rho_2\sigma_2^2} e^{-\frac{SNR_{2,1}}{\beta_N\rho_2\sigma_2^2}} - \frac{1}{\sigma_2^4(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})^2} \\
&+ \frac{1 + \frac{SNR_{2,1}}{\beta_N\rho_2}(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})}{\sigma_2^4(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})^2} e^{-(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})\frac{SNR_{2,1}}{\beta_N\rho_2}} \\
&- \frac{1}{\beta_N^3\rho_2^3\sigma_3^2\sigma_2^4}\{\frac{2\beta_N^3\rho_2^3}{(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})^3} - e^{-(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})\frac{SNR_{2,1}}{\beta_N\rho_2}} \\
&\times \frac{2 + 2\frac{SNR_{2,1}}{\beta_N\rho_2}(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2}) + (\frac{SNR_{2,1}}{\beta_N\rho_2})^2(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})^2}{\frac{1}{\beta_N^3\rho_2^3}(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_3^2})^3}\} \\
\overset{(c)}{\approx}&\frac{1}{\rho_2^2} \times \frac{SNR_{2,1}^2}{\beta_N^2\sigma_2^2}[\frac{2}{\sigma_2^2} - \frac{1}{\sigma_2^2+\sigma_3^2} - \frac{SNR_{2,1}}{\beta_N\rho_2\sigma_2^2\sigma_3^2}]
\end{aligned}$$

where (c) is obtained by $e^{-x} \underset{x\to 0}{\sim} 1 - x$ when ρ_2 is large. Then it makes $\frac{1}{\rho_2^2}$ in (c) close to zero, indicating $\Pr[D] \approx 0$.

Appendix B

Analysis of Complexity for single-band Scenario

The optimal algorithm will exhaustively search all possible $C_{M_T}^{L_T}$ possibilities of antenna subset and figure out the one with the maximal user sum rate. So the operation times are $2L_T C_{M_T}^{L_T}$ for obtaining all possibilities of user sum rate and $C_{M_T}^{L_T}$ for obtaining the optimal one. In this way, the complexity for optimal algorithm in antenna selection is $O(L_T C_{M_T}^{L_T})$.

For Joint AU contribution algorithm, only the steps relevant to antenna side should be executed since users involved in communications have been identified. In particular, we can initialized the user side contribution as $cw^U = 1$ for each user. Then we need to calculate the antenna side contribution only once. Since there are one subband, the computational operations are $2M_T$, followed by M_T operations for sorting. As a result, the final complexity is $O(M_T)$.

Curriculum Vitae

Name:	Xin Liu
Post-Secondary Education and Degrees:	Wuhan University Wuhan, Hubei, China 2010 - 2014 B.E. University of Western Ontario London, ON 2014 - 2016 M.E.Sc
Honours and Awards:	NSERC CREATE Program in Communications Society, Privacy and Cyberethics, 2014-2016
Related Work Experience:	Teaching Assistant The University of Western Ontario 2015 - 2016

Publications:

1. X. Liu and X. Wang, "Efficient Antenna Selecion and User Scheduling in 5G Massive MIMO-NOMA System" *Proc. Vehicular Technology Conference. VTC Spring*, May 2016.

2. X. Liu and X. Wang, "Outage Probability and Capacity Analysis of the Collaborative NOMA Assisted Relaying System in 5G" *Proc. International Conference on Communications in China ICCC*, Jul. 2016.

3. X. Liu, X. Wang, Y. Liu and H. Lin, "Highly Efficient 3D Resource Allocation Techniques in 5G for NOMA Enabled Massive MIMO and Relaying Systems" *IEEE Trans. Wireless Commun.* (to be submitted)