MANCHESTER
1824
The University of Manchester

Signal Processing Algorithms for MIMO-NOMA Based 6G Networks

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Science and Engineering

2023

Yunus Dursun

School of Engineering

Department of Electrical and Electronic Engineering

Contents

Word count: 37012

List of figures

List of tables

List of publications

Journal Publications

1. **Y. Dursun**, “Comparative Analysis of NOMA and OMA Schemes: GSVD-based NOMA Systems and the Role of Mobile Edge Computing,” In press, Journal of Telecommunications and Information Technology, 2023.

2. **Y. Dursun**, F. Fang and Z. Ding, “Hybrid NOMA based MIMO offloading for mobile edge computing in 6G networks,” in China Communications, vol. 19, no. 10, pp. 12-20, Oct. 2022, doi: 10.23919/JCC.2022.00.024.

3. **Y. Dursun**, K. Wang and Z. Ding, “Secrecy sum rate maximization for a MIMO-NOMA uplink transmission in 6G networks,” in Physical Communication, vol. 53, pp. 101675, Mar. 2022, doi:10.1016/j.phycom.2022.101675.

4. **Y. Dursun**, M. B. Goktas and Z. Ding, “Green NOMA based MU-MIMO transmission for MEC in 6G Networks,” in Computer Networks, vol. 228, pp. 109749, Apr. 2023, doi:10.1016/j.comnet.2023.109749.

5. **Y. Dursun** and Z. Ding, “Wireless Powered NOMA-based Cognitive Radio for 6G Networks,” submitted to Computer Networks.

Terms and abbreviations

3GPP	3rd Generation Partnership Project
AWGN	Additive White Gaussian Noise
AN	Artificial Noise
AO	Alternating Optimization
AP	Access Point
BPCU	Bits Per Channel Use
BS	Base Station
CPU	Central Processing Unit
CR	Cognitive Radio
CSI	Channel State Information
DC	Difference of Convex
DL	Downlink
EH	Energy Harvesting
FDMA	Frequency Division Multiple Access
GSVD	Generalized Singular Value Decomposition
H-NOMA	Hybrid Non-orthogonal Multiple Access
KKT	Karush-Kuhn-Tucker
MEC	Mobile Edge Computing
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
NOMA	Non-Orthogonal Multiple Access
NSBF	Null Space Beamforming
OMA	Orthogonal Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access

QoS	Quality-of-Service
QCQP	Quadratically Constrained Quadratic Program
PD-NOMA	Power Domain Non-Orthogonal Multiple Access
PLS	Physical Layer Security
RAN	Radio Access Network
RB	Resource Block
RF	Radio Frequency
SCA	Successive Convex Approximation
SC-FDMA	Single-carrier Frequency-division Multiple Access
SDP	Semi-definite Programming
SDR	Semi-definite Relaxation
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single-Input Single-Output
SSR	Secrecy Sum Rate
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access
UE	User Equipment
UL	Uplink
ZF	Zero-Forcing

Abstract

This thesis aims to enhance the anticipated performance of sixth-generation (6G) communication networks by integrating non-orthogonal multiple access (NOMA) and multiple-input multiple-output (MIMO) techniques using the generalized singular value decomposition (GSVD)-based linear beamforming method. To do so, the thesis explores different scenarios with specific objectives, including minimizing mobile edge computing (MEC) offloading delay, maximizing the sum secrecy rate in a MIMO-NOMA system, minimizing the total energy consumption of a MIMO-NOMA-MEC system, and maximizing the sum data rate of the secondary network in a cognitive radio (CR)-based NOMA-MIMO system.

To minimize the offloading delay in a MIMO-MEC system, the Dinkelbach transform and the GSVD method are employed. Analytical and simulation-based evaluations are conducted to assess the performance of the proposed Hybrid-NOMA-MIMO-MEC system. The simulation results show that this system achieves superior delay performance and lower energy consumption than conventional orthogonal multiple access (OMA) approaches.

To improve the sum rate of confidential transmission in an uplink MIMO-NOMA system, the thesis focuses on maximizing the secrecy sum rate (SSR). By leveraging the GSVD method and first-order Taylor approximation, a suboptimal concave problem formulation is derived to tackle the non-convex nature of the SSR problem. The SSR is compared with other algorithms, including conventional orthogonal multiple access, and the simulation results demonstrate the effectiveness of the proposed method.

To minimize the total energy consumption of local computing, task offloading, and MEC computing in a NOMA-MIMO-based system, the base station optimizes power allocation vectors and task assignment coefficients under time and power constraints. The non-convex problem is addressed through successive convex optimization (SCA) and alter-

nating optimization (AO) techniques. The impact of various factors, such as delay tolerance, task size, and user distance, on energy consumption is investigated. Simulation results indicate that the proposed method outperforms orthogonal multiple access (OMA) schemes, particularly for large data sizes and stringent delay requirements.

Finally, the thesis presents a novel approach for wireless-powered NOMA-MIMO systems. This approach is designed for cognitive underlay radio (CR) scenarios where the primary network requires predefined QoS. Given that requirement, the main objective is to maximize the sum rate of the secondary network. A joint beamforming vector for primary and secondary networks and a time-switching coefficient for energy harvesting and information transfer are optimized to achieve this objective. The problem formulation is non-convex. Therefore, we use of semi-definite programming, successive convex approximation, and alternating optimization techniques to solve this problem. The simulation results show that the NOMA-based solution outperforms the TDMA-based benchmark scheme, particularly at low transmit power levels.

In conclusion, this thesis investigates integrating NOMA and MIMO technologies in 6G networks. It addresses delay minimization, maximization of secrecy sum rate, energy consumption optimization, and sum rate in CR scenarios. The proposed solutions demonstrate significant improvements in spectral efficiency, energy efficiency, data rate, and overall system performance, making them valuable contributions to the field of 6G communication networks.

Declaration of originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institutes of learning.

Copyright statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made *only* in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations/`) and in The University's policy on Presentation of Theses.

Acknowledgements

I sincerely thank my supervisor, Prof. Zhiguo Ding, who has been my role model. His kindness, modesty, and generosity in sharing his profound knowledge have inspired me throughout my doctoral journey. I have felt his support at every step, and I am immensely grateful for his guidance.

I want to wholeheartedly dedicate my gratitude to my wife Büşra, my mother Fatma, my father İsrafil, and my brother Hüseyin. Their endless love and unwavering support have been immeasurable throughout this incredible journey.

I would like to thank my dear friends and teachers for their kind help and support.

I am deeply grateful to the Türkiye Ministry of National Education for financial support through the Graduate Education Scholarship (YLSY) program.

Chapter 1

Overview

Data plays a critical role in shaping our world and is a driving factor for developing wireless communications. Due to the rapid increase in the number of connected devices and data traffic in wireless mediums, which is expected to continue growing, academia and industry are giving attention to developing next-generation wireless networks.

Improving multiple access techniques has been considered a crucial way to keep up with cutting-edge network services. Historically, these advancements have been revealed as follows [1]. First-generation (1G) cellular networks utilized frequency division multiple access (FDMA) for transmitting analog voice calls. FDMA involves dividing the frequency band into sub-bands allocated to individual users. However, FDMA has some drawbacks, such as low spectral efficiency, high inter-channel interference, and limited capacity. In second-generation (2G) networks, time division multiple access (TDMA) and code division multiple access (CDMA) technologies were introduced. TDMA allocates different time slots to users, allowing them to share the same frequency band. However, TDMA requires precise synchronization and complex network planning, which limits its flexibility. On the other hand, CDMA assigns a unique code to each user to modulate their signals, enabling simultaneous transmission without interference. However, CDMA is vulnerable to near-far problems, energy inefficiency, and jamming or eavesdropping. Despite these problems, third-generation (3G) networks widely use CDMA due to their high spectral efficiency and improved capacity compared to the aforementioned techniques. Fourth-generation (4G) and fifth-generation (5G) networks host orthogonal frequency division multiple access (OFDMA) and spatial division multiple access (SDMA) technologies, providing a significant improvement in data rate, network architecture, and support services. SDMA separates signals from different users by uti-

lizing multiple antennas based on their spatial locations, while OFDMA allocates multiple sub-carriers from a wide frequency band to different users. OFDMA systems face a problem with peak-to-average power ratios, while SDMA can be expensive because it needs more antennae than the number of users who want to transmit simultaneously [2], [3]. These problems have prompted scientists to develop a new multiple-access method that uses signal quality distinctions and serves multiple users with a single resource block [4].

Non-orthogonal multiple access (NOMA) is gaining momentum as a well-suited technology to meet the requirements of modern communication systems due to its ability to provide high spectral efficiency and support massive connectivity [5]. Additionally, NOMA is compatible with OFDMA, which makes NOMA a natural progression toward sixth-generation (6G) networks. Multiple-input multiple-output (MIMO) is another wireless communication technology that offers various advantages, such as improved data rates, increased spatial diversity, and enhanced spectral efficiency, making it essential for mobile devices [6]. This thesis aims to integrate NOMA and MIMO technologies for different scenarios. To this end, the thesis mainly focuses on the generalized singular value decomposition technique to combine NOMA and MIMO. In NOMA, multiple users share the same frequency and time resources, leading to user interference. The motivation behind using the generalized singular value decomposition (GSVD) technique in NOMA systems is to enhance user signals separation and reduce system interference. We provide numerical results showing that applying GSVD to NOMA can help improve the performance of the NOMA-based systems over orthogonal multiple access (OMA)-based systems.

1.1 The Aims and Objectives of the Thesis

To propose a GSVD-based MIMO-NOMA transmission system built on OFDMA in 5G, enhancing system performance compared to existing literature.

- Minimize the transmission delay in the MIMO-MEC to improve spectral efficiency, energy efficiency, and data rate of MEC offloading.

- Maximize the secrecy sum rate (SSR) for a MIMO-NOMA uplink network under maximum total transmit power and QoS constraints.
- Minimize total energy consumption during local computing, task offloading, and MEC computing in a MIMO-MEC system.
- Maximize the sum rate of the secondary network in a cognitive underlay radio (CR) scenario, where the primary network requires a certain level of QoS and secondary network users can download their data using the same spectrum.

1.2 The main contributions of the Thesis

In this thesis, we go beyond the previous studies on NOMA-based systems, which primarily focused on SISO transmission. Instead, we incorporate both NOMA and MIMO technologies through the GSVD technique, leading to significant advancements in the contribution chapters. The main contributions of this thesis can be listed as follows:

- A hybrid NOMA-based MIMO-MEC system is introduced to minimize the offloading delay. The system model is formulated and transformed into an easily manageable form using mathematical manipulations, the Dinkelbach, the GSVD, and the KKT methods. Performance evaluation of the Hybrid-NOMA-MIMO-MEC system is conducted through numerical results. Simulation results further demonstrate the system's superiority, showcasing improved delay performance and reduced energy consumption compared to OMA.
- We investigate a novel uplink MIMO-NOMA network using GSVD in PLS scenarios, considering multiple external eavesdroppers and a friendly jammer. The system configuration and adopted schemes set it apart from previous studies. A non-concave problem of maximizing SSR is formulated to enhance the system's performance. Through an equivalent transformation of norm functions to trace functions in logarithmic expressions, the problem is reformulated as a difference of convex (DC) programs. The first-order Taylor approximation method is then applied to convert the DC problem into a suboptimal concave problem, and an SCA (successive convex approximation) based algorithm is proposed. The algorithm's

properties, including complexity and convergence, are analyzed. Simulation results demonstrate that the GSVD-based MIMO-NOMA system outperforms conventional MIMO-NOMA regarding SSR. The performance gap between NOMA and OMA schemes is revealed using the same optimization method.

- Another contribution is in a scenario where two multiple antenna-equipped NOMA users collaboratively offload their data to the MEC server, aiming to minimize total energy consumption. The energy minimization problem is formulated by considering various factors such as offloading time limitations, power allocation, task assignment, and energy losses in RF chains. To integrate NOMA, MIMO, and MEC technologies, the GSVD linear beamforming method is also employed. Given the non-convex nature of the problem, an alternating optimization approach is utilized to jointly optimize task assignment coefficients and power allocation vectors, with the task assignment problem solved in the outer layer. The inner layer employs the successive convex approximation method to convert the non-convex power allocation problem into a first-order linear form, enabling the determination of power allocation vectors. Numerical results are provided, analyzing the impact of factors such as offloading time, power budget, users' locations, and data rate on energy consumption. Comparative analysis with OMA as the benchmark scheme demonstrates the significant energy-saving benefits of NOMA-based MIMO-NOMA networks, particularly at high SINR rates with a higher number of antennas. Additionally, the findings emphasize the importance of optimal user pairing in reducing energy consumption within the NOMA framework.

- We introduce an underlay network incorporating cognitive radio and MISO-NOMA technologies. Both the primary and secondary transmitters harvest wireless energy and employ a time-switching protocol to transmit information signals to their respective users. QoS constraints are satisfied, even with potential interference experienced by the far user from the primary user. The secondary transmitter optimizes the sum data rate of the secondary network while considering these constraints. The wireless power transfer in the first time slot is performed through GSVD beamforming. Semi-definite programming and the first-order Taylor series expansion are utilized in the second time slot to optimize the split time variable and the beam-

forming vectors for both the primary and secondary networks. Simulations analyze the impact of QoS and power constraints on the sum rate of the secondary users. The proposed model outperforms the TDMA-based benchmark scheme due to its higher spectral efficiency, allowing the primary users to operate at lower power levels, thus reducing interference for the secondary users. However, as the QoS requirements of the primary users become more stringent, the available transmission time for the secondary users decreases, leading to a widening performance gap between the proposed method and the benchmark.

1.3 Organization of the Thesis

This thesis is motivated by the aforementioned aims and objectives to explore the application of GSVD-based MIMO-NOMA networks. To compare the performance of the proposed schemes with the conventional OMA schemes, we formulated different scenarios with specific objectives: delay minimization, secrecy sum rate maximization, energy minimization, and sum-rate maximization. The formulated problems are solved using convex optimization techniques. The organization of the thesis is outlined as follows:

Chapter 2: This chapter provides an in-depth explanation and comprehensive comparison of the basic concept of NOMA with OMA schemes. The chapter investigates the GSVD method and explores its combination with NOMA while offering a relevant review of GSVD-based NOMA systems. Furthermore, the chapter introduces the concept of MEC and thoroughly discusses its key parameters, accompanied by a comprehensive review of NOMA MEC. Mathematical tools such as the Karush–Kuhn–Tucker (KKT) method, fractional programming, the first-order Taylor approximation, semi-definite programming, and semi-definite relaxation are explained.

Chapter 3: In the context of sixth-generation communication networks, this chapter focuses on three key technologies: NOMA, MIMO, and MEC. These technologies are prominent in satisfying modern communication systems' high data rate demands. The primary objective of this chapter is to minimize transmission delay within the MIMO-MEC system, thereby improving spectral efficiency, energy efficiency, and data rate of MEC offloading.

Chapter 4: NOMA, as a well-qualified candidate for 6G mobile networks, has been attracting remarkable research interests due to high spectral efficiency and massive connectivity. This chapter aims to maximize the secrecy sum rate (SSR) for a MIMO-NOMA uplink network under the maximum total transmit power and quality of service (QoS) constraints. Thanks to the generalized singular value decomposition method, the SSR of NOMA is compared with conventional orthogonal multiple access and other baseline algorithms in different MIMO scenarios.

Chapter 5: MEC is a distributed computing paradigm that brings computing and data storage closer to the network's edge. This chapter considers a MIMO uplink scenario where NOMA users partially offload their data to a MEC server. This chapter aims to minimize the total energy consumption during local computing, task offloading, and MEC computing. To this end, the base station optimizes power allocation vectors and task assignment coefficients under time and power constraints.

Chapter 6: The need for physically charging mobile devices is anticipated to become a thing of the past in the not-too-distant future. In this chapter, a novel approach for wireless-powered mobile devices is presented, which is based on NOMA and multi-user multiple-input multiple-output (MU-MIMO) antenna systems. The proposed method is specifically designed for use in a cognitive underlay radio (CR) scenario, where the primary network requires a certain level of QoS. Meanwhile, secondary network users can download their data using the same spectrum. The main objective of this chapter is to maximize the sum rate of the secondary network.

Chapter 7: This chapter summarizes the conclusions drawn from this thesis and explores potential research topics concerning NOMA-based MIMO systems for 6G networks.

Chapter 2

Background Information

This chapter serves as an introductory overview of the key concepts explored in this thesis. Initially, the chapter examines uplink and downlink NOMA systems, specifically in the context of a SISO antenna configuration. Following this, the chapter explores the GSVD technique and its relevance to beamforming frameworks. To clarify these concepts, an illustrative example is provided. Furthermore, the chapter encompasses a concise review of GSVD-based applications in MIMO systems, thereby establishing the context for the subsequent discussions.

The chapter then explains the MEC concept and presents the basic formulations related to MEC. Moreover, a summary of existing works on NOMA MEC is provided, highlighting its significance within the research domain.

Finally, the chapter addresses the mathematical tools used in this thesis. Various techniques, including the Karush-Kuhn-Tucker (KKT) method, fractional programming, successive convex approximation, semi-definite programming, and semi-definite relaxation, are introduced, and their relevance within the research framework is established. These mathematical tools serve as foundational elements for the subsequent analyses presented in the thesis.

2.1 Non-orthogonal Multiple Access

Non-orthogonal multiple access (NOMA) is a multiple access technique that allows multiple users to simultaneously share the same frequency, time, or code resources to communicate with a base station or access point. There are several NOMA schemes in the

literature [3]. Some of the well-known ones are summarized as follows. The sparse code multiple access (SCMA) technique enables multiple users to communicate with a base station simultaneously by utilizing separate sparse codes assigned to each user, based on a multi-dimensional codebook [7]. In pattern division multiple access (PDMA), the available bandwidth is divided into multiple non-overlapping frequency patterns or slots. Thus, the users have a unique pattern or slot to modulate their signals [8]. Resource spread multiple access employs a distinct spreading sequence for users to disperse their data over the frequency band. The receiver then reverses the spreading process by applying the identical spreading sequence to recover the user's information [9]. Multi-user shared access (MUSA) is based on code-domain multiplexing, where symbols are multiplexed using the same spreading code. These symbols are transmitted over an orthogonal channel, such as a sub-carrier, as in OFDMA. At the receiver end, SIC decodes the received symbols [10]. Interleave-grid multiple access (IGMA) is another technique in which the user's data is segmented and interleaved based on a specific pattern, creating a grid-like structure that helps minimize user interference and improves the overall spectral efficiency of the system [11]. Rate-splitting multiple access (RSMA) is also getting huge research interest. In rate-splitting multiple access, users partition their data into shared and exclusive components [12]. The shared components of each user are aggregated and modulated jointly, while each user's unique components are modulated separately. This results in a transmitted signal with shared and unique components for all users. Both users initially decode the shared component at the receiver, treating any interference from the unique signals as noise. Both users use SIC to decode their exclusive signals in the subsequent stage.

In 3GPP Release 13, the standardization of power domain NOMA (PD-NOMA), known as MUST (Multi-User Superposition Transmission), has been introduced for a broadcast channel. In PD-NOMA, multiple users use different power levels to share the same time and frequency resources. At the transmitter, superposition coding is employed to multiplex the users, while the receiver uses successive interference cancellation to decode the superposed signals [13]. PD-NOMA is considered a promising technique for 5G and beyond wireless communication systems, as it can significantly increase spectral efficiency and support multiple users with diverse communication requirements. PD-

NOMA can also improve user fairness and energy efficiency, enabling users with weaker channel conditions to share the same resources with stronger users without sacrificing their quality of service [14].

2.1.1 An overview of Power-Domain NOMA

This section introduces the basic concepts of PD-NOMA for downlink and uplink networks. Additionally, we analyze and compare the sum rate and signal-to-interference-plus-noise ratio (SINR) of NOMA and OMA.

Downlink NOMA Network

Figure 2.1 illustrates a downlink NOMA scheme consisting of a BS/AP and K receivers, where the BS broadcasts a superposed signal to all the receivers. The BS combines complex-valued symbols with superposition coding (SC), and the receivers employ the successive interference cancellation (SIC) technique to decode their respective signals. Each receiver, except for the weakest or those without SIC capabilities, performs the SIC process at the receiver side. Even though there is a tendency for the SIC process as follows: the users first extract the strongest signal from the combined signal and then subtract it to eliminate the interference from the remaining signals, and the SIC process is repeated until the receiver's signal is decoded. This strategy may not be optimal. Ding et al. showed that dynamic decoding orders according to users' QoS and CSI-based SIC orders might improve the system's performance [15].

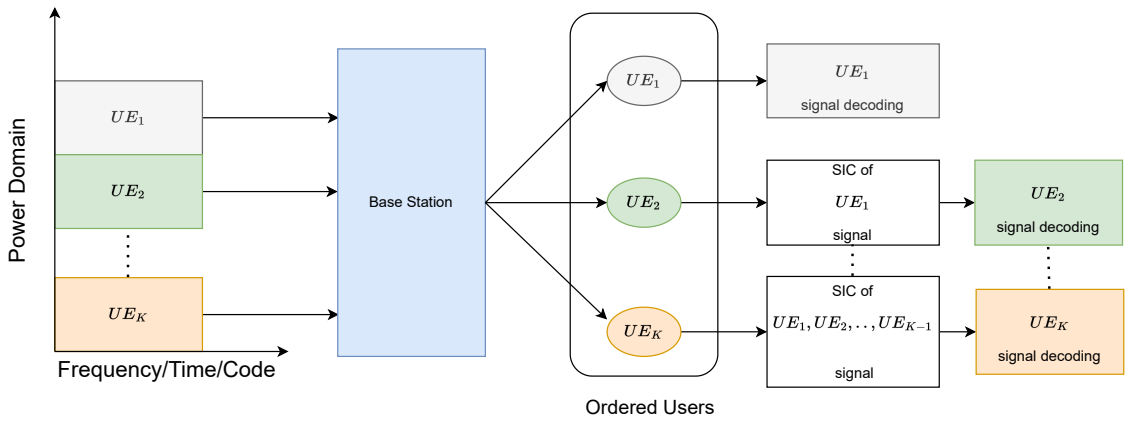

Figure 2.1. A basic concept of downlink NOMA [14]

To simplify the analysis, we consider a downlink NOMA system with a base station and

two users to derive the SINRs and sum rates. Additionally, we assume that the base station and users are equipped with a single antenna and the system bandwidth (B) is one. The information-bearing signals, x_N for the near user (UE_1) and x_F for the far user (UE_2), are superimposed at the transmitter as follows:

$$x = \sqrt{P_N}x_N + \sqrt{P_F}x_F, \tag{2.1}$$

where P_N and P_F denote the transmission power allocation coefficients for the near and far users, respectively. P_{tot} represents the total transmit power which equals the sum of P_N and P_F. The received signal at the receivers are

$$y_i = h_i x + n_i,\ i \in \{N, F\}, \tag{2.2}$$

h_i denotes the channel coefficient between the BS and user UE_i, and n_i represents the additive white Gaussian noise (AWGN) with zero mean and σ_i^2 variance for UE_i. Let's assume the users are ordered using the CSI-based method at the receiver and the near user has a strong signal than of the far user, i.e., $\frac{|h_N|^2}{\sigma_N^2} \geq \frac{|h_F|^2}{\sigma_F^2}$. Therefore, the SINR expression of the near user and far user are given by

$$SNR_N = \frac{P_N|h_N|^2}{\sigma_N^2}, \tag{2.3}$$

$$SINR_F = \frac{P_F|h_F|^2}{P_N|h_N|^2 + \sigma_F^2}. \tag{2.4}$$

Accordingly, the data rate for the near user and far user can be written as follows:

$$R_N = \log_2\left(1 + \frac{P_N|h_N|^2}{\sigma_N^2}\right), \tag{2.5}$$

$$R_F = \log_2\left(1 + \frac{P_F|h_F|^2}{P_N|h_N|^2 + \sigma_F^2}\right). \tag{2.6}$$

Uplink NOMA Network

As illustrated in Figure 2.2, an uplink NOMA system allows K users to simultaneously transmit their data to the BS using the same spectrum. The base station employs SIC to decode the signals from different users. We again assume that the near user has better

channel gain than the far user, i.e., $\frac{|h_N|^2}{\sigma_N^2} \geq \frac{|h_F|^2}{\sigma_F^2}$. The received signal at the receiver is

$$y = h_N x_N + h_F x_F + n_B, \tag{2.7}$$

where n_B is a AWGN with zero mean and σ_N^2 variance at the receiver. If the BS decodes the received signal in descending order, the data rate for the near and far users are:

$$R_N = \log_2 \left(1 + \frac{P_N |h_N|^2}{P_F |h_F|^2 + \sigma_B^2}\right), \tag{2.8}$$

$$R_F = \log_2 \left(1 + \frac{P_F |h_F|^2}{\sigma_B^2}\right). \tag{2.9}$$

On the other hand, if the BS decodes the received signal in ascending order, the data rate for the near and far users becomes:

$$R_N = \log_2 \left(1 + \frac{P_N |h_N|^2}{\sigma_B^2}\right), \tag{2.10}$$

$$R_F = \log_2 \left(1 + \frac{P_F |h_F|^2}{P_N |h_N|^2 + \sigma_B^2}\right). \tag{2.11}$$

It is worth mentioning that in each case, the sum rate for the users is the same as given in Equation 2.12. In other words, the sum rate in the uplink NOMA does not depend on the order of SIC, assuming no error propagation occurs in the SIC process.

$$R_N + R_F = \log_2 \left(\frac{P_N |h_N|^2 + P_F |h_F|^2 + \sigma_B^2}{\sigma_B^2}\right) \tag{2.12}$$

However, according to Benjebbour [13], performing SIC in the descending order of channel quality is more practical.

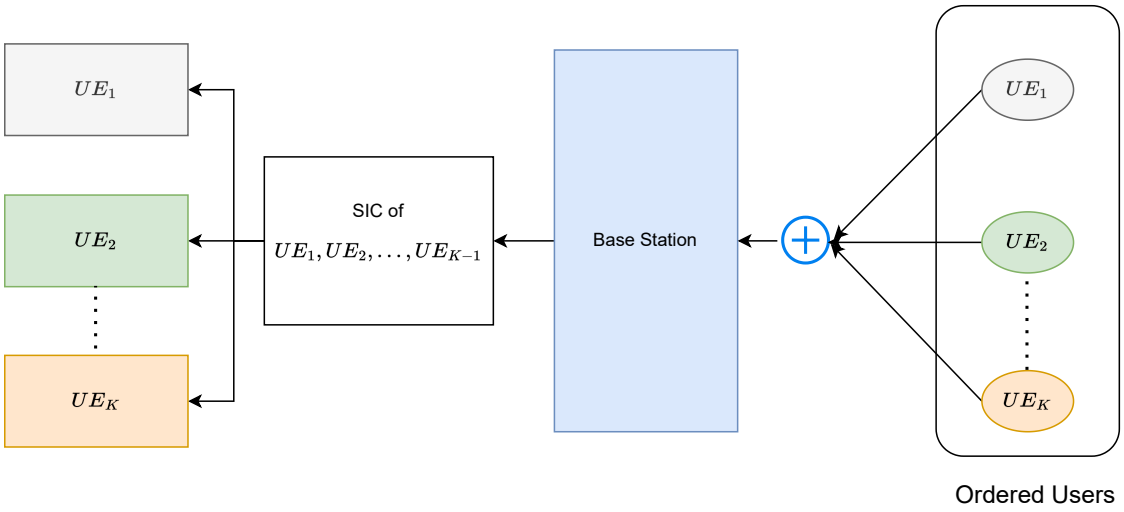

Figure 2.2. A basic concept of uplink NOMA [14]

An example for sum rate comparison between NOMA, TDMA, and FDMA Networks

This section compares NOMA with TDMA and FDMA schemes for an uplink scenario as illustrated in Fig. 2.3. The data rate of the near and far users in the NOMA, FDMA, and TDMA systems are given as follows [16]:

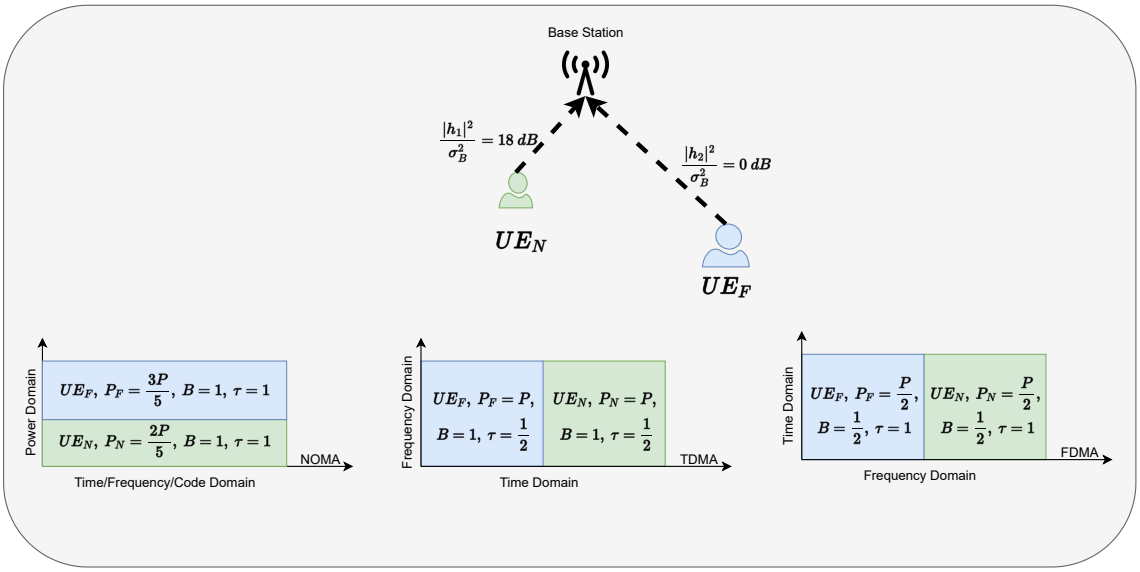

Figure 2.3. Sum rate comparison between NOMA, TDMA, and FDMA networks

$$NOMA = \begin{cases} R_N^{NOMA} = B\log_2\left(1+\frac{P_N|h_N|^2}{B\sigma_B^2}\right) \\ R_F^{NOMA} = B\log_2\left(1+\frac{P_F|h_F|^2}{P_N|h_N|^2+B\sigma_B^2}\right) \\ 0 \leq P_N,\ P_F \leq P, \end{cases} \tag{2.13}$$

$$TDMA = \begin{cases} R_N^{TDMA} = B(1-\tau)\log_2\left(1+\frac{P_N|h_N|^2}{B\sigma_B^2}\right) \\ R_F^{TDMA} = B\tau\log_2\left(1+\frac{P_F|h_F|^2}{B\sigma_B^2}\right) \\ 0 \leq \tau \leq 1, \end{cases} \tag{2.14}$$

$$FDMA = \begin{cases} R_N^{FDMA} = B(1-\omega)\log_2\left(1+\frac{P_N|h_N|^2}{B(1-\omega)\sigma_B^2}\right) \\ R_F^{FDMA} = B\omega\log_2\left(1+\frac{P_F|h_F|^2}{B\omega\sigma_B^2}\right) \\ 0 \leq \omega \leq 1, \end{cases} \tag{2.15}$$

Assuming two uplink users UE_N and UE_F with channel gains of $\frac{h_N}{\sigma_B^2} = 18\ dB$ and $\frac{h_F}{\sigma_B^2} = 0\ dB$, respectively, the total power is the same in all schemes such that $P_N + P_F = P$, where P is the maximum transmit power [13]. In the TDMA scheme, the users are allocated equal time slots, i.e., $\tau = 0.5$. The data rates for the near and far users are $R_N^{TDMA} = 3.0011\ bps$ and $R_F^{TDMA} = 0.5\ bps$, respectively. In the FDMA scheme, the bandwidth is split equally between the users, i.e., $\omega = 0.5$, and the resulting data rates are $R_N^{FDMA} = 3.0011\ bps$ and $R_F^{FDMA} = 0.5\ bps$ for the near and far users, respectively. In the NOMA case, the power is split between the users by δ, with two out of five for the near user and three out of five for the far user. Thus, the data rates for the near and far users are $R_N^{NOMA} = 4.0682\ bps$ and $R_F^{NOMA} = 0.6781\ bps$. The total sum rates achieved by the TDMA, FDMA, and NOMA schemes are $3.5011\ bps$, $3.5011\ bps$, and $4.7463\ bps$, respectively. Based on this example, it can be concluded that NOMA offers a significant advantage over OMA schemes in terms of spectral efficiency, resulting in 35.57% higher sum data rates.

2.2 Enhancing MIMO-NOMA Systems through GSVD: Leveraging Joint Precoding, Interference Mitigation, and Power Optimization

Generalized singular value decomposition (GSVD) is a powerful matrix factorization technique that extends the standard singular value decomposition (SVD) to accommodate rectangular matrices of potentially different dimensions. This technique encompasses two primary types: real-valued GSVD and complex-valued GSVD. Real GSVD is utilized for real-valued matrices, while complex GSVD is tailored for complex-valued matrices. Various algorithms, such as Van Loan's, which was first introduced in 1976, and Paige and Sounders' algorithms [17], facilitate the computation of GSVD. By decomposing matrices into their singular components, GSVD finds widespread application across diverse domains. It is employed in signal processing tasks like adaptive filtering, blind source separation, and channel estimation and in analyzing biological data in bioinformatics [18], [19]. In wireless communication, the GSVD decomposes MIMO channels into orthogonal SISO channels. In other words, the main use case of the GSVD is beamforming design. Table 2.1 lists various studies that employ the GSVD to solve some problems in wireless communication.

2.2.1 A definition of GSVD

Let us consider two matrices, $\mathbf{H_1} \in \mathbb{C}^{m\times n}$ and $\mathbf{H_2} \in \mathbb{C}^{m\times n}$. By applying the GSVD method, we can decompose these matrices into three components: a unitary matrix, a non-singular matrix, and a non-negative singular matrix. The decomposition can be expressed as follows [28]:

$$\mathbf{\Sigma_1} = \mathbf{U}\mathbf{H_1}\mathbf{Q} \quad \text{and} \quad \mathbf{\Sigma_2} = \mathbf{V}\mathbf{H_2}\mathbf{Q}, \tag{2.16}$$

where the matrices $\mathbf{U} \in \mathbb{C}^{m\times m}$ and $\mathbf{V} \in \mathbb{C}^{m\times m}$ are unitary, while $\mathbf{Q} \in \mathbb{C}^{n\times n}$ is non-singular, and $\mathbf{\Sigma_1} \in \mathbb{C}^{m\times n}$ and $\mathbf{\Sigma_2} \in \mathbb{C}^{m\times n}$ are diagonal matrices with non-negative elements. The dimension of the matrices form $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ as follows:

Table 2.1. GSVD-based MIMO applications

Ref.	Objective	Technology	System analysis/Optimization	Optimization Variable	Constraints	UL/DL	Result
[20]	To minimize outage probability while improving physical layer security	NOMA-MIMO	Performance analysis and optimization	Power allocation coefficients	N/A	DL	Compared with GSVD-OMA based transmission, NOMA has superior outage performance
[21]	To design low complexity and highly efficient GSVD-based beamforming to maximize secrecy capacity	OMA-MIMO	Optimization	Power allocation coefficients	Average power consumption	DL	GSVD-MIMO achieves nearly identical performance with secure dirty paper coding (S-DPC)
[22]	Maximize secrecy rate	OMA-MIMO	Optimization	Sub-channel and power allocation	Quality of service	DL	GSVD-based precoding outperforms a TDMA-based system
[23]	To obtain the expressions of the average data rate and outage in a MIMO-NOMA relaying	NOMA-MIMO	Performance analysis	N/A	Finite number of users	DL	GSVD-NOMA achieves a higher sum rate than GSVD-OMA
[24]	Maximize minimum data rate	NOMA-MIMO	Optimization	Power allocation coefficients	Imperfect channel estimation	DL	The SINR balancing problem was solved using error bounds. The proposed solution has better performance than non-robust or OMA-based solutions
[25]	Minimize offloading delay	H-NOMA-MIMO	Optimization	Power allocation coefficients	Total power	UL	Hybrid NOMA-MIMO based solution has better delay performance compared with OMA-based solution
[26]	Minimize energy consumption	NOMA-MIMO	Optimization	Task assignment and power allocation coefficients	Total power, offloading time, and RF chains energy consumption	UL	NOMA-MIMO performs better than OMA, especially when the data is high and time is stringent
[27]	Secrecy sum rate maximization	NOMA-MIMO	Optimization	Power allocation coefficients	Total power and QoS	UL	NOMA has better SSR performance than OMA

- If $m \geq n$, then $\mathbf{\Sigma_1} = \begin{pmatrix} \mathbf{0}_{(m-n)\times n} \\ \mathbf{S}_1 \end{pmatrix}$ and $\mathbf{\Sigma_2} = \begin{pmatrix} \mathbf{S}_2 \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix}$.
- If $m \leq n \leq 2m$, $r = n - m$ and $q = 2m - n$, then
  $\mathbf{\Sigma_1} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0}_{r\times q} & \mathbf{0}_{r\times r} \\ \mathbf{0}_{q\times r} & \mathbf{S}_1 & \mathbf{0}_{q\times r} \end{pmatrix}$ and $\mathbf{\Sigma_2} = \begin{pmatrix} \mathbf{0}_{q\times r} & \mathbf{S}_2 & \mathbf{0}_{q\times r} \\ \mathbf{0}_{r\times r} & \mathbf{0}_{r\times q} & \mathbf{I}_r \end{pmatrix}$.
- If $2m \geq n$, then $\mathbf{\Sigma_1} = \begin{pmatrix} \mathbf{I}_m & \mathbf{0}_{m\times(n-m)} \end{pmatrix}$ and $\mathbf{\Sigma_2} = \begin{pmatrix} \mathbf{0}_{m\times(n-m)} & \mathbf{I}_m \end{pmatrix}$,

where $\mathbf{0}$ and $\mathbf{I}$ represent the zero and identity matrices, respectively. Moreover, $\mathbf{S}_1$ and $\mathbf{S}_2$ are non-negative diagonal matrices, with elements between zero and one. Notably, the elements of $\mathbf{S}_1$ are sorted in descending order, while those of $\mathbf{S}_2$ are sorted in ascending order. Table 2.2 illustrates the key distinctions between the SVD, the GSVD, and the multi-linear GSVD methods.

Multilinear GSVD (ML-GSVD) is an advanced variant of GSVD, specifically designed to handle multiple tensors rather than just two matrices. Multilinear GSVD identifies a shared underlying structure among them, which is then represented through common factor matrices. Multilinear GSVD finds applications in various domains such as multilinear subspace learning, tensor factorization, and MU-MIMO systems.

2.2.2 Application of GSVD in MIMO-NOMA

We consider a base station (BS) with n antennas communicating with two downlink users, each equipped with m antennas. The channels between the BS and the users can be represented by $\mathbf{G}_i = \frac{\mathbf{H}_i}{\sqrt{d_i^\tau}}$, where $\mathbf{H}_i$ denotes the small-scale fading coefficients. The near user denoted as UE_N, and the far user, denoted as UE_F, are sorted based on their large-scale fading element $\sqrt{d_i^\tau}$. Here, d represents the distance of the i^{th} user, and τ represents their path loss component. The received signals at the receivers are:

$$\mathbf{y}_N = \frac{\mathbf{H}_N\mathbf{x}}{\sqrt{d_N^\tau}} + \mathbf{n}_N \quad \text{and} \quad \mathbf{y}_F = \frac{\mathbf{H}_F\mathbf{x}}{\sqrt{d_F^\tau}} + \mathbf{n}_F, \tag{2.17}$$

The noise at the i^{th} receiver, $\mathbf{n}_i,\ i \in N, F$, modeled by the additive white Gaussian noise, is given by mutually independent and identically distributed elements with zero

Table 2.2. Key difference between SVD, GSVD and ML-GSVD

	SVD	GSVD	ML-GSVD
Application Domain	Single matrix factorization for various applications (i.e., data compression, diemntion reduction, etc.)	Deals with two sets of matrices (i.e., Canonical Correlation Analysis, multi-user MIMO, etc.)	A set of matrices with one common dimension (more than two matrices)
Matrix Types	Single square or rectangular matrices	Square or rectangular two matrices with the same row size	A tensor consisting of complex matrices with the same number of rows and various number of columns
Purpose	Decomposes a single matrix into three separate matrices representing its singular vectors and singular values	To find common structures or correlations between two sets of data	To analyze a collection of more than two matrices
Solution	Produces unique solutions for square matrices (the best low-rank approximation)	GSVD may have multiple solutions (non-optimal)	Designed to handle multiple tensors and can reveal shared patterns among them

mean and variance σ_i. The transmitted signal, denoted as $\mathbf{x} \in \mathbb{C}^{n\times 1}$, is subject to interference mitigation techniques using precoding and decoding matrices $\mathbf{P}_b \in \mathbb{C}^{n\times n}$ and $\mathbf{D}_i \in \mathbb{C}^{m\times m}$, respectively. The decomposition of channels, as described in Eq. 2.16, leads to the selection of detection matrices $\mathbf{D}_i$ as $\mathbf{U}$ and $\mathbf{V}$ for the near and far users. Additionally, the precoding matrix $\mathbf{P}_b$ is modified to $\mathbf{Q}\sqrt{P}/t$, where P represents the maximum transmission power and t is a power normalization coefficient [28]. Consequently, the MIMO receivers obtain the transmitted signal as follows:

$$\mathbf{U}\mathbf{y}_N = \mathbf{U}\mathbf{H}_N\mathbf{P}_b\mathbf{x} + \mathbf{U}\mathbf{n}_\mathbf{N} = \frac{P}{t\sqrt{d_N^\tau}}\mathbf{\Sigma}_N + \mathbf{U}\mathbf{n}_N, \tag{2.18}$$

$$\mathbf{V}\mathbf{y}_F = \mathbf{V}\mathbf{H}_F\mathbf{P}_b\mathbf{x} + \mathbf{V}\mathbf{n}_\mathbf{F} = \frac{P}{t\sqrt{d_F^\tau}}\mathbf{\Sigma}_F + \mathbf{V}\mathbf{n}_F. \tag{2.19}$$

Please note that $\mathbf{U}$ and $\mathbf{V}$ are the unitary matrices; therefore, the unitary matrices $\mathbf{U}$ and $\mathbf{V}$ preserve the variance of noise after multiplication with them.

Example: In this example, we analyze a basic setup comprising a base station with four transmitter antennas. The near user is equipped with three receiver antennas, while the distant users have two receiver antennas. The near, which small-scale channel coefficient denoted as $\mathbf{H}_N \in \mathbb{C}^{3\times 4}$, is located $d_N = 40$ meters away from the base station. On the other hand, the far user, represented by $\mathbf{H}_F \in \mathbb{C}^{2\times 4}$, is located $d_F = 75$ meters away from the base station. The value of the path loss component, denoted as α, is 3.2.

$$\mathbf{H}_N = \begin{bmatrix} 0.629+0.735i & 0.066+0.931i & 0.193+0.616i & 0.924+0.556i \\ 0.210+0.772i & 0.260+0.013i & 0.639+0.949i & 0.263+0.915i \\ 0.752+0.907i & 0.804+0.234i & 0.524+0.950i & 0.065+0.641i \end{bmatrix},$$

$$\mathbf{H}_F = \begin{bmatrix} 0.390+0.173i & 0.604+0.135i & 0.926+0.021i & 0.394+0.827i \\ 0.485+0.126i & 0.549+0.505i & 0.918+0.947i & 0.963+0.015i \end{bmatrix}.$$

Decoding matrices $\mathbf{U} \in \mathbb{C}^{3\times 3}$ and $\mathbf{V} \in \mathbb{C}^{2\times 2}$ can be found as follows:

$$\mathbf{U} = \begin{bmatrix} 0.5886+0.0000i & 0.0000+0.0000i & 0.8084+0.0000i \\ 0.2571-0.6199i & 0.4855+0.2742i & -0.1872+0.4513i \\ -0.2400+0.3816i & 0.6442+0.5235i & 0.1747-0.2778i \end{bmatrix},$$

$$\mathbf{V} = \begin{bmatrix} -0.5294 + 0.7844i & 0.3201 + 0.0453i \\ -0.2029 - 0.2517i & 0.4024 - 0.8565i \end{bmatrix}.$$

And the precoding matrix $\mathbf{Q}$ becomes

$$\mathbf{Q} = \begin{bmatrix} -0.2004 + 0.3013i & 0.2468 - 0.5664i & 1.2740 - 0.5080i & 0.6976 - 0.7226i \\ -0.4527 + 0.5098i & -0.0139 - 0.7738i & 0.7713 + 0.3349i & 0.0869 - 0.8978i \\ -0.8984 + 0.6990i & -0.1623 - 1.2709i & 1.4061 - 0.6229i & 0.2928 - 0.3442i \\ 0.2405 + 0.5082i & 0.6040 - 1.2088i & 0.7572 - 0.7511i & 0.9449 - 0.2902i \end{bmatrix}.$$

Let's rearrange the super-positioned transmitted signal $\mathbf{x} \in \mathbb{C}^{4\times 1}$ with power allocation such that $\mathbf{x} = \mathbf{Ps}$, where $\mathbf{P} \in \mathbb{C}^{4\times 4}$ is the diagonal nonnegative power allocation matrix and $\mathbf{s} \in \mathbb{C}^{4\times 1}$ contains the coded signals for both users as follows:

$$\mathbf{x} = \underbrace{\begin{bmatrix} p_{1,1} & 0 & 0 & 0 \\ 0 & p_{2,2} & 0 & 0 \\ 0 & 0 & p_{3,3} & 0 \\ 0 & 0 & 0 & p_{4,4} \end{bmatrix}}_{\mathbf{P}} \times \underbrace{\begin{bmatrix} l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2} \\ l_2 s_{2,1} + \sqrt{(1-l_2)^2} s_{2,2} \\ l_3 s_{3,1} + \sqrt{(1-l_3)^2} s_{3,2} \\ l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2} \end{bmatrix}}_{\mathbf{s}} = \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2)^2} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3)^2} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix},$$

where $s_{i,1}$ and $s_{i,2}$ represent the corresponding message, and $l_{i,1}$ and $l_{i,2}$ are the power allocation coefficients for the near and far users. Also we assume that $\mathbf{s}$ encoded with unit power i.e, $\|\mathbf{s}_i\|^2 = 1, i \in \{1, 2, 3, 4\}$. After the GSVD is applied to the downlink channels, the channels become

$$\mathbf{\Sigma}_N = \begin{bmatrix} 0 & 0.4526 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}, \quad \mathbf{\Sigma}_F = \begin{bmatrix} 1.0000 & 0 & 0 & 0 \\ 0 & 0.8917 & 0 & 0 \end{bmatrix}.$$

The observations at the near user is equal to $\mathbf{y}_N = \frac{\mathbf{\Sigma}_N \mathbf{x}}{\sqrt{d_N^\alpha}} + \mathbf{n}_N$ that can be written as

follows:

$$\mathbf{y}_N = \begin{bmatrix} 0 & 0.4526 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix} \frac{1}{\sqrt{d_N^{\alpha}}} + \begin{bmatrix} n_{N,1} \\ n_{N,2} \\ n_{N,3} \end{bmatrix},$$

$$= \begin{bmatrix} 0.4526 \times (l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \sqrt{d_N^{-\alpha}} + n_{N,1} \\ 1 \times (l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \sqrt{d_N^{-\alpha}} + n_{N,2} \\ 1 \times (l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \sqrt{d_N^{-\alpha}} + n_{N,3} \end{bmatrix}.$$

Similarly, the observations at the far user $\mathbf{y}_F$ is equal to $\mathbf{y}_F = \frac{\mathbf{\Sigma}_F \mathbf{x}}{\sqrt{d_F^{\alpha}}} + \mathbf{n}_F$ and it can be given as follows:

$$\mathbf{y}_F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8917 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} p_{1,1}(l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \\ p_{2,2}(l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \\ p_{3,3}(l_3 s_{3,1} + \sqrt{(1-l_3^2)} s_{3,2}) \\ p_{4,4}(l_4 s_{4,1} + \sqrt{(1-l_4^2)} s_{4,2}) \end{bmatrix} \frac{1}{\sqrt{d_F^{\alpha}}} + \begin{bmatrix} n_{F,1} \\ n_{F,2} \end{bmatrix},$$

$$= \begin{bmatrix} 1 \times (l_1 s_{1,1} + \sqrt{(1-l_1^2)} s_{1,2}) \sqrt{d_F^{-\alpha}} + n_{F,1} \\ 0.8917 \times (l_2 s_{2,1} + \sqrt{(1-l_2^2)} s_{2,2}) \sqrt{d_F^{-\alpha}} + n_{F,2} \end{bmatrix}.$$

We can consider each sub-channel an individual SISO channel; therefore, they may require different SIC ordering regarding their effective channel gains. Furthermore, by examining the expressions for $\mathbf{y}_N$ and $\mathbf{y}_F$, we can deduce that $\mathbf{s}_1$ corresponds to a private stream intended for the far user. On the other hand, $\mathbf{s}_3$ and $\mathbf{s}_4$ represent private streams dedicated to the near user, while $\mathbf{s}_2$ serves as the common stream shared by both users. The OMA transmission strategy can be employed for private streams. For example, the first steam can equal $\mathbf{s}_1 = s_{1,2}$. Chapter 4 will provide an in-depth examination of secrecy aspects for further analysis. Now, let us have a look at the decoding of the common stream. It can be calculated that the near user has a stronger channel gain than the far user as follows, i.e., $\frac{0.4525}{\sqrt{40^{3.2}}} \geq \frac{0.8917}{\sqrt{75^{3.2}}}$. Therefore, the near and far users decode their signals as follows:

$$R_{N,2} = \log_2 \left(1 + \frac{p_{2,2} \times l_2^2 \frac{0.4525}{40^{3.2}}}{\text{Var}(n_{N,2})} \right), \tag{2.20}$$

$$R_{F,2} = \log_2 \left(1 + \frac{p_{2,2} \times (1 - l_2^2) \times \frac{0.8917}{75^{3.2}}}{p_{2,2} \times l_2^2 \frac{0.4525}{40^{3.2}} + \text{Var}(n_{F,2})} \right). \tag{2.21}$$

Likewise, data rate expressions can be derived for private channels by eliminating inter-user interference. Additionally, the transmitted power from each antenna, denoted as **P**, holds a specific physical interpretation, such as maximum transmit power from each antenna element. Hence, the opportunity for further exploration emerges from optimizing the elements within **P**, encompassing enhancing secrecy, improving data rate, ensuring fairness, and selecting optimal antennas. In Chapter 5, we have conducted additional research on power consumption in RF chains.

2.3 Mobile Edge Computing

Mobile edge computing, also known as multi-access edge computing, brings the processing of traffic and services from centralized cloud servers to the edge of the network, closer to the end-users, as illustrated in Fig. 2.4. Instead of transmitting all the data to the cloud for analysis, the MEC devices carry out the processing, storage, and analysis of the data [29]. This approach minimizes latency, improving high-bandwidth applications' performance in real-time [30]. The combination of NOMA and MEC holds immense potential, as it not only enhances the spectral efficiency of MEC users but also empowers IoT devices at the edge to handle computationally intensive tasks. Combining NOMA and MEC requires optimal resource and power allocation and time management.

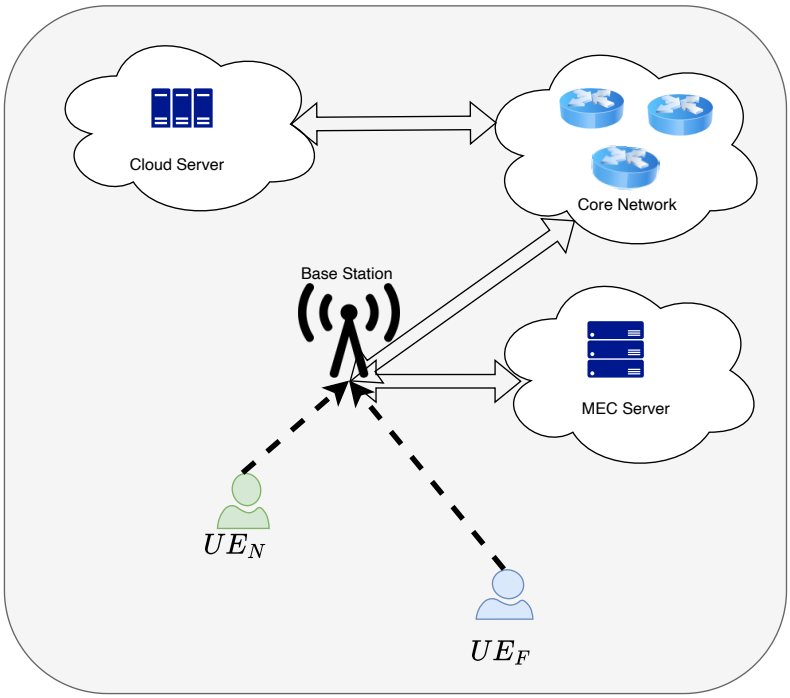

Figure 2.4. NOMA assisted MEC model

In order to minimize offloading time, UE needs to determine the optimal task partition coefficient (β) and power allocation (p_{off}). The offloading time (T_{off}) can be defined as follows:

$$T_{off} = \frac{\beta N}{R},\ s \tag{2.22}$$

Here, N represents the data size of the task, and R is the data rate of the UE. The energy consumed during the offloading time (T_{off}) can be calculated as:

$$E_{off} = T_{off} \times p_{off},\ joule \tag{2.23}$$

In the above equation, p_{off} denotes the transmit power of the UE. Once the data is offloaded to the MEC server, the duration for the mobile execution time T_{mec} can be determined using the following equation:

$$T_{mec} = \frac{\beta N C_m}{f_m},\ s \tag{2.24}$$

In this equation, C_m represents the required CPU cycles to execute a bit, and f_m is the CPU frequency of the MEC server. The energy consumption during T_{mec} can be calculated as:

$$E_{mec} = \xi \beta N C_m f_m^2,\ joule \tag{2.25}$$

where ξ denotes the energy consumption coefficient for the MEC. Table 2.3 summarizes existing works on combining NOMA and MEC. It provides valuable insights into different research papers that have explored this area regarding the optimization perspective.

2.4 Mathematical Tools Used in the Thesis

2.4.1 Karush-Kuhn-Tucker Method

The KKT (Karush-Kuhn-Tucker) method is a mathematical technique utilized to solve optimization problems containing equality and inequality constraints, regardless of whether the problem is linear or nonlinear. The KKT conditions, namely stationary, primal feasibility, dual feasibility, and complementary slackness, are necessary but insufficient to determine the global optimal solution. In other words, meeting the KKT conditions at a point does not guarantee that it is the global optimum but a local optimum. Second-order sufficient conditions must be examined to ascertain a point's global optimality. In this part, we will discuss the necessary and sufficient conditions for the following problem:

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \tag{2.26a}$$

$$\text{s.t.} \quad g_i(\mathbf{x}) \leq 0 \; i \in \{1, 2, ..m\}, \tag{2.26b}$$

$$h_j(\mathbf{x}) = 0 \; j \in \{1, 2, ..n\}, \tag{2.26c}$$

Necessary conditions

Stationary: The stationary condition states that if a solution point satisfies the stationary condition, the point is either a local minimum or local maximum, which means that the first-order derivative of the objective function on this point is equal to zero, i.e.,$\nabla f(\mathbf{x}^*) = 0$.

Primal feasibility: Primal feasibility refers to the process of verifying whether a candidate solution ($\mathbf{x}^*$) fulfills the equality and inequality constraints of an optimization prob-

Table 2.3. Summary of some existing works on NOMA-MEC

Ref.	Objective	Method	Technology	Constraints	Optimization parameters	Offloading policy
[31]	Minimize task offloading and computing delay	Linear problem using auxiliary variable	massive MIMO	Transmit power, MEC computing capacity	Power allocation, computation frequency allocation	Partial
[29]	Minimize delay	Bisection search	SISO-MEC	Energy and offloading power, computation time at MEC	Task partition coefficient, power allocation	Partial
[26]	Minimize total energy minimization during local computing task offloading and MEC computing	AO-SCA	MIMO-MEC	Total power, time, energy consumption on RF chains	Task partition coefficient, power allocation	Partial
[27]	Minimize delay	Dinkelbach transform-SCA	MIMO-MEC	Total power	Power allocations	Full
[32]	Minimize system energy consumption	SCA	SISO-MEC	Time, transmission power, decoding power	Power allocation	Full
[33]	Minimize delay	Alternating optimization (AO)	UAV assisted SISO-MEC	Energy and QoS	Trajectory of UAV, power allocation, user scheduling	Full
[34]	Maximize computation capacity	AO, DC programming	Backscatter-assisted SISO_NOMA	Energy and QoS	Energy harvesting time coefficient, BackCom time coefficient, transmission time, computing resource allocation	Partial
[35]	Maximize computation capacity	AO, Concave-convex procedure and SDR	IRS and UAV assisted SISO-NOMA	Total transmit power	Phase shift of IRS, transmit power, computational resource allocation, the trajectory of UAV	Binary
[36]	Minimize system energy	AO, Matching algorithm, SCA	SISO-MEC	Latency	Power allocation, time, sub-channel allocation	Binary
[37]	Maximize computation capacity	Deep reinforcement learning	SISO-multi-MEC	Delay, limited sub-channel	Task scheduling, power allocation	Binary
[30]	Minimize latency	AO	WPT, IRS	Transmit power	Power allocation, phase shift of IRS	Partial
[38]	Minimize total energy consumption	TD3	SISO-MEC	Transmit power, latency	Task partition, power allocation	Binary
[39]	Maximize computation probability	Meta-heuristic-based algorithms, PSO, GA	SISO-MEC	Transmit power, computational resource	Task partition, power allocation	Binary/partial/full
[40]	Minimize power consumption	AO, Riemann gradient descent	IRS-MEC	Delay, computational resource	Bandwidth allocation, computational resource allocation, power allocation, the phase shift of IRS	Binary

lem, i.e., $g_i(\mathbf{x}^*) \leq 0$ and $h_j(\mathbf{x}^*) = 0$.

Dual feasibility: Dual feasibility requires that the Lagrange multipliers associated with the inequality constraints are nonnegative, i.e., $\mu_j \geq 0$ for $i \in \{1, 2, ..m\}$ where m is the number of inequality constraint.

Complementary slackness: Complementary slackness establishes a connection between two key aspects of constrained optimization: primal feasibility and dual feasibility. The complementary slackness indicates whether the inequality constraints are effective or not on the objective function. The complementary slackness equations for the inequality constraints in (2.26b) can be given as follows:

$$\mu_i g_i(\mathbf{x}^*) = 0 \; i \in \{1, 2, ..m\}. \tag{2.27}$$

Eq. 2.27 indicates that if $\mu_i > 0$, i^{th} inequality is binding; otherwise, μ_i is not effective on the objective function. Eq. 2.27 indicates that if the Lagrange multiplier associated with the i^{th} inequality constraint, denoted as μ_i, is greater than zero, it indicates that the i^{th} inequality constraint is binding, meaning it affects the optimization problem's solution. Conversely, if μ_i is zero, it implies that the constraint does not impact the objective function.

Sufficient conditions

The sufficient condition involves evaluating the second-order derivative of the objective function with respect to a potential solution point. If the variable $\mathbf{x}$ is a vector, the second-order derivative is represented by the Hessian matrix. A solution point is considered the global optimum if the Hessian matrix is either positive-definite or negative-definite. Let's define the Lagrangian function as follows:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\lambda}_j) = f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) \tag{2.28}$$

where

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(x) \\ : \\ g_m(x) \end{bmatrix}, \ \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(x) \\ : \\ h_l(x) \end{bmatrix}, \ \boldsymbol{\mu}(\mathbf{x}) = \begin{bmatrix} \mu_1(x) \\ : \\ \mu_m(x) \end{bmatrix} \text{ and } \boldsymbol{\lambda}(\mathbf{x}) = \begin{bmatrix} \lambda_1(x) \\ : \\ \lambda_l(x) \end{bmatrix} \tag{2.29}$$

The sufficient condition states that if the inequality

$$\mathbf{s}^T \nabla^2_{xx} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*_i, \boldsymbol{\lambda}^*_j)\mathbf{s} \geq 0 \tag{2.30}$$

holds for all nonzero vectors **s**, then the vector $\mathbf{x}^*$ corresponds to the minimum solution for the given optimization problem (2.26).

An example for the optimal solution using KKT

Let's find the optimal solution for the following problem.

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = 4x_1^2 + 3x_2^2 \tag{2.31a}$$

$$\text{s.t.} \quad g(\mathbf{x}) = 3x_1 + 5x_2 \leq 20 \tag{2.31b}$$

$$h(\mathbf{x}) = 2x_1 + x_2 = 10 \tag{2.31c}$$

The Lagrangian function can be formulated as follows:

$$\mathcal{L}(\mathbf{x}, \mu, \lambda) = f(\mathbf{x}) + \mu g(\mathbf{x}) + \lambda h(\mathbf{x}) \tag{2.32}$$

where μ and λ are the Lagrange multipliers associated with the inequality and equality constraints, respectively. Substituting the objective function $f(\mathbf{x})$ and the constraint functions $g(\mathbf{x})$ and $h(\mathbf{x})$, we have:

$$\mathcal{L}(\mathbf{x}, \mu, \lambda) = 4x_1^2 + 3x_2^2 + \mu(3x_1 + 5x_2 - 20) + \lambda(2x_1 + x_2 - 10) \tag{2.33}$$

Determine the partial derivatives of the Lagrangian with respect to the decision variables:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 8x_1 + 3\mu + 2\lambda \tag{2.34}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 6x_2 + 5\mu + \lambda \tag{2.35}$$

Set the derivatives equal to zero to find the stationary points:

$$8x_1 + 3\mu + 2\lambda = 0 \tag{2.36}$$

$$6x_2 + 5\mu + \lambda = 0 \tag{2.37}$$

Solve the above equations along with the equality constraint and complementary slackness:

$$2x_1 + x_2 = 10 \tag{2.38}$$

$$\mu(3x_1 + 5x_2 - 20) = 0 \tag{2.39}$$

Case 1 (μ=0): We assume that μ equals zero to solve (2.26). Therefore, we got

$$8x_1 + 2\lambda = 0 \tag{2.40}$$

$$6x_2 + \lambda = 0 \tag{2.41}$$

$$2x_1 + x_2 = 10 \tag{2.42}$$

From the equations above, $x_1 = \frac{15}{4}$, $x_2 = \frac{5}{2}$ and $\lambda = -15$. Check if **x** satisfies the KKT conditions. Primal feasibility: $3\frac{15}{4} + 5\frac{5}{2} \leq 20$ is not correct because $23.75 > 20$. Therefore, the given solution is not a feasible solution. Thus, there is no need to check the other conditions, and we continue to the second case, where $\mu > 0$.

Case 2 (μ>0): From Eq. (2.39) then we can say that $3x_1 + 5x_2 - 20 = 0$. Therefore,

$$3x_1 + 5x_2 - 20 = 0 \tag{2.43}$$

$$8x_1 + 3\mu + 2\lambda = 0 \tag{2.44}$$

$$6x_2 + 5\mu + \lambda = 0 \tag{2.45}$$

$$2x_1 + x_2 = 10 \tag{2.46}$$

From the equations above, $x_1 = \frac{30}{7}$, $x_2 = \frac{10}{7}$, $\lambda = -94.2857$ and $\mu = 17.1429$. Check if **x** satisfies the KKT conditions. Primal feasibility: g(x_1, x_2)=$3\times\frac{30}{7} + 5 \times \frac{10}{7} \leq 20$, and h$(x_1, x_2)$=$2 \times \frac{30}{7} + \frac{10}{7} = 10$. Therefore, primary feasibility constraints are satisfied. Dual feasibility: the Lagrange multipliers μ are non-negative ($\mu \geq 0$). Therefore, the solution passes the dual feasibility. Finally, the minimum value for $f(\mathbf{x})$ equals 79.5918. Figure 2.5 provides an illustration of the objective function, as well as the equality and inequality constraints.

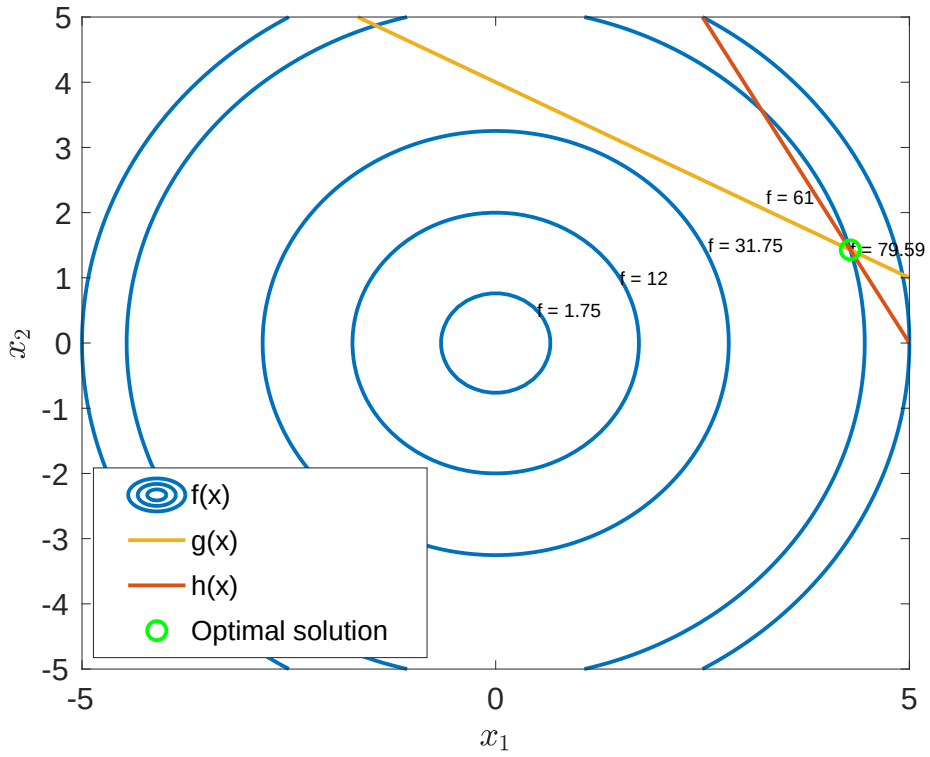

Figure 2.5. Illustration of the objective and constraint functions

2.4.2 Fractional Programming

Fractional programming is an optimization technique that addresses problems involving fractional objectives or constraints. In the realm of communication system design, fractional expressions commonly arise in scenarios like power control, beamforming design, and energy efficiency. These expressions can encompass various types, including linear, quadratic, polynomial, logarithmic, and exponential forms. Fractional programming can

be categorized into two groups: single-ratio problems and multiple-ratio problems.

Dinkelbach method

The Dinkelbach method is an iterative approach to solving single-ratio problems, ensuring convergence by updating an auxiliary variable, denoted as y. To begin, let us define the single-ratio problem as presented in [41]:

$$\max_{\mathbf{x}} \quad \frac{A(\mathbf{x})}{B(\mathbf{x})} \tag{2.47a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} \tag{2.47b}$$

where $A(\mathbf{x})$ is nonnegative function on the numerator and $B(\mathbf{x})$ is a positive-valued function on the denominator, i.e., $A(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}_+$ and $B(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}_{++}$. By introducing the auxiliary variable y, we can transform (2.47) into the following form:

$$\max_{\mathbf{x}} \quad A(\mathbf{x}) - yB(\mathbf{x}) \tag{2.48a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} \tag{2.48b}$$

The variable y is updated as follows:

$$y[t+1] = \frac{A(\mathbf{x}[t])}{B(\mathbf{x}[t])}, \tag{2.49}$$

where t denotes the iteration number. The Dinkelbach method is effective in providing the optimal solution for single-ratio problems. However, its applicability to multi-ratio problems is limited. To address this, Shen et al. introduced the quadratic transform for multi-ratio fractional problems [42].

Quadratic transform

Let us consider a generalized form of a single-ratio problem, also referred to as a multi-ratio problem [42]:

$$\max_{\mathbf{x}} \quad \sum_{m=1}^{M} \frac{A_m(\mathbf{x})}{B_m(\mathbf{x})} \tag{2.50a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} \tag{2.50b}$$

where $A_m(\mathbf{x})$ is nonnegative function on the numerator and $B_m(\mathbf{x})$ is a positive-valued function on the denominator, i.e., $A_m(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}_+$ and $B_m(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}_{++}$ and $m = 1, ..., M$. By using the quadratic transform, (2.50) can be equivalently rewritten as:

$$\max_{\mathbf{x},\mathbf{y}} \quad \sum_{m=1}^{M} (2y_m \sqrt{A(\mathbf{x})} - y_m^2 (B(\mathbf{x}))) \tag{2.51a}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X},\ y_m \in \mathbb{R} \tag{2.51b}$$

where $\mathbf{y} = \{y_1\ y_2\ ...\ y_M\}$ is a variable that can be updated as follows:

$$y_m[t+1] = \frac{\sqrt{A(\mathbf{x}[t])}}{B(\mathbf{x}[t])},\ \forall m = 1, ..., M. \tag{2.52}$$

The Dinkelbach algorithm and the Quadratic transform are two well-known methods used to solve fractional programming problems. While the Dinkelbach method is proven to converge to the optimal solution for single-variable fractional problems, the Quadratic transform can be applied to multi-variable fractional problems. Another notable difference is the complexity: the Quadratic transform converts the fractional expression into a quadratic optimization form, making it more computationally efficient than the iterative Dinkelbach solution [42].

2.4.3 Successive Convex Approximation

Suppose $k \geq 1$ be an integer and the function $f : \mathbb{R} \rightarrow \mathbb{R}$ be k times differentiable at the point $a \in \mathbb{R}$. Then, f can be linearly approximated at the point of a as follows [43]:

$$f(x) \approx f(a) + \frac{\partial f(x)}{\partial x}|_{x=a}(x - a) \tag{2.53}$$

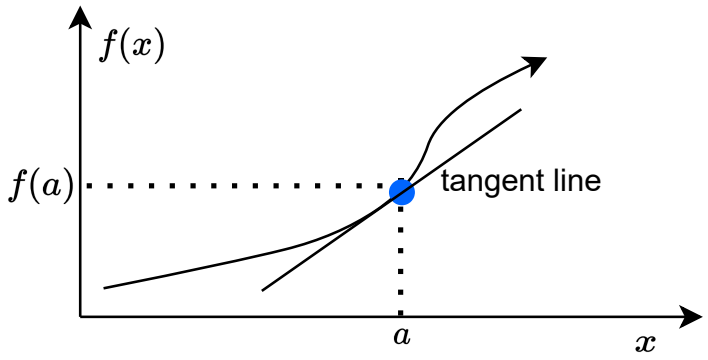

Figure 2.6. An illustration of first-order Taylor approximation

The first-order Taylor approximation can be applied to multivariate functions. Let's assume that $f(\mathbf{x})$ is a multivariate function taking $\mathbf{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ as the variable. The linear approximation of $f(\mathbf{x})$ at the point $\mathbf{a} = [a_1, a_2, a_3]^{\mathrm{T}}$ can be written as follows:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \frac{\partial f(\mathbf{x})}{\partial x_1}|_{\mathbf{x}=\mathbf{a}}(x_1 - a_1) + \frac{\partial f(\mathbf{x})}{\partial x_2}|_{\mathbf{x}=\mathbf{a}}(x_2 - a_2) + \frac{\partial f(\mathbf{x})}{\partial x_3}|_{\mathbf{x}=\mathbf{a}}(x_3 - a_3). \quad (2.54)$$

Further, the first-order Taylor expansion can be extended for matrix-valued functions. Let assume that $f(\mathbf{X})$ is a matrix-valued scalar function, at $\mathbf{A}$, it can be linearized as follows:

$$f(\mathbf{X}) \approx f(\mathbf{A}) + \frac{\partial f(\mathbf{A})}{\partial \mathbf{X}}(\mathbf{X} - \mathbf{A}) \quad (2.55)$$

where $\frac{\partial f(\mathbf{A})}{\partial \mathbf{X}}$ is the Jacobian matrix of a function f evaluated at a point $\mathbf{A}$, and it can be written as $\begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{A}) & \frac{\partial f}{\partial x_2}(\mathbf{A}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{A}) \\ \frac{\partial f}{\partial x_1}(\mathbf{A}) & \frac{\partial f}{\partial x_2}(\mathbf{A}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{A}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_1}(\mathbf{A}) & \frac{\partial f}{\partial x_2}(\mathbf{A}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{A}) \end{bmatrix}$

Example:

Linearize the function $f(\mathbf{x}) = x_1^3 + 3x_2^2 - x_1 + 6x_2 + 7$ at $\mathbf{a} = [3\ 1]^{\mathrm{T}}$.

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \frac{\partial f(\mathbf{x})}{\partial x_1}|_{\mathbf{x}=\mathbf{a}}(x_1 - a_1) + \frac{\partial f(\mathbf{x})}{\partial x_2}|_{\mathbf{x}=\mathbf{a}}(x_2 - a_2)$$

$$f(\mathbf{a}) = f([3\ 1]^{\mathrm{T}}) = 40$$

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 3x_1^2 - 1 = 26$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 6x_2 + 6 = 12$$

$$f(\mathbf{x}) \approx 40 + 26(x-3) + 12(x-1)$$

The result of the example is illustrated in Fig. 2.7.

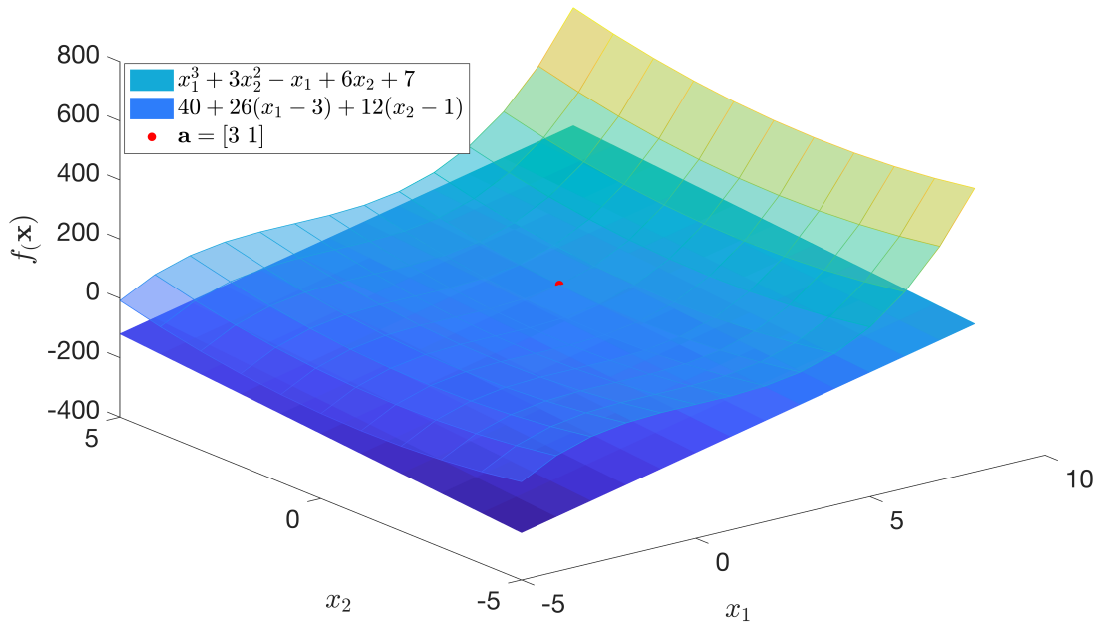

Figure 2.7. The first-order Taylor approximation of a vector-valued function

2.4.4 Semi-definite Programming

Semi-definite programming (SDP) is an advanced optimization technique that encompasses and extends the capabilities of linear, quadratic, and second-order cone programming. While linear programming typically involves vector decision variables, SDP extends the scope to positive semi-definite matrices as decision variables. The objective of SDP is to minimize or maximize an objective function, with positive semi-definite matrices as the decision variables, subject to constraints involving affine combinations of symmetric matrices [44]. SDP in standard form can be given as follows:

$$\max_{\mathbf{X} \in \mathbb{S}^n} \quad \mathrm{Tr}(\mathbf{CX}) \tag{2.56a}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{A_i X}) \succeq b_i,\ i = \{1, .., p\} \tag{2.56b}$$

$$\mathrm{Tr}(\mathbf{A_j X}) = b_j,\ j = \{p+1, .., m\} \tag{2.56c}$$

$$\mathbf{X} \succeq 0 \tag{2.56d}$$

where $\mathbf{X}$ is a square decision variable matrix. The matrices $\mathbf{C}$, $\mathbf{A_i}$, and $\mathbf{A_j}$ represent objective and constraint matrices, all of which have the same size as $\mathbf{X}$. The notation $\mathbf{X} \succeq 0$ indicates that $\mathbf{X}$ is positive semi-definite. b_i and b_j are scalar values. PSD matrix can be defined as follows. If a symmetric square matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ satisfies the condition that $\mathbf{v}^\top \mathbf{X} \mathbf{v} \succeq 0$ where $\mathbf{v}$ is a non-zero real vector, $\mathbf{X}$ is a semi-definite matrix. In certain types of QCQP problems, such as principal component analysis or beamforming design (discussed in Chapter 6), a rank one solution for the matrix variable $\mathbf{X}$ is desired. However, the rank one constraint is non-convex and cannot be directly incorporated into the SDP formulation.

2.4.5 Semi-definite Relaxation

Semi-definite relaxation efficiently approximates non-convex QCQP problems by excluding the rank-one constraint. This approach allows for solving vector-valued QCQP problems via the matrix-valued SDP method. Figure 2.8 illustrates the relationship between QCQP and SDP. A real-valued homogeneous QCQP problem can be expressed as [45]:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \quad \mathbf{x^T C x} \tag{2.57a}$$

$$\text{s.t.} \quad \mathbf{x^T A_i x} \succeq b_i, \; i = \{1, .., p\} \tag{2.57b}$$

$$\mathbf{x^T A_j x} = b_j, \; j = \{p+1, .., m\} \tag{2.57c}$$

where the matrices $\mathbf{A_i}$ and $\mathbf{A_i}$ are symmetric; however, their definiteness is not specified. It is important to note that the main distinction between (2.56) and (2.57) lies in the variable that requires optimization. The first step is to reformulate (2.57) as in (2.56). In this process, the vector variable $\mathbf{x}$ is transformed into a new matrix variable $\mathbf{X} = \mathbf{x}\mathbf{x}^T$.

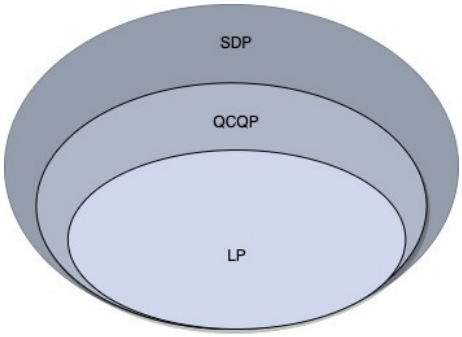

Figure 2.8. Illustrating the relationship between linear programming, quadratically constraint quadratic programming, and semi-definite programming

Importantly, it should be noted that the vector $\mathbf{x}$ is equivalent to a rank one PSD matrix. Therefore, (2.57) can be rewritten as follows:

$$\begin{aligned}
\max_{\mathbf{X}\in\mathbb{S}^n} \quad & \text{Tr}(\mathbf{CX}) && \text{(2.58a)}\\
\text{s.t.} \quad & \text{Tr}(\mathbf{A_i X}) \succeq b_i,\ i = \{1,..,p\} && \text{(2.58b)}\\
& \text{Tr}(\mathbf{A_j X}) = b_j,\ j = \{p+1,..,m\} && \text{(2.58c)}\\
& \mathbf{X} \succeq 0 && \text{(2.58d)}\\
& \text{rank}(\mathbf{X}) = 1. && \text{(2.58e)}
\end{aligned}$$

By relaxing the rank one constraint, (2.58) can be written as follows:

$$\begin{aligned}
\max_{\mathbf{X}\in\mathbb{S}^n} \quad & \text{Tr}(\mathbf{CX}) && \text{(2.59a)}\\
\text{s.t.} \quad & \text{Tr}(\mathbf{A_i X}) \succeq b_i,\ i = \{1,..,p\} && \text{(2.59b)}\\
& \text{Tr}(\mathbf{A_j X}) = b_j,\ j = \{p+1,..,m\} && \text{(2.59c)}\\
& \mathbf{X} \succeq 0. && \text{(2.59d)}
\end{aligned}$$

The problem denoted by (2.59) represents a convex optimization problem, which can be effectively solved using a convex optimization toolbox such as [46]. Upon obtaining the optimal solution $\mathbf{X}^*$, it is necessary to derive the corresponding optimal $\mathbf{x}^*$ from it. However, the extraction process is not always as straightforward as in the case where $\text{rank}(\mathbf{X}^*)$ is equal to one, resulting in $\mathbf{X} = \mathbf{x}\mathbf{x}^T$. In scenarios where $\text{rank}(\mathbf{X}^*)$ is greater than one, alternative methods exist to obtain the optimal solution. One commonly used approach involves considering the largest eigenvalue. By performing eigenvalue decom-

position, the eigenvalues corresponding to $\mathbf{X}^*$ can be arranged in descending order as follows: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$. The associated eigenvectors are denoted as $\mathbf{q}_1, \ldots, \mathbf{q}_r \in \mathbb{R}^n$, where r represents the rank of $\mathbf{X}^*$. A sub-optimal solution for problem (2.59) can be obtained using the largest eigenvalue. Specifically, the approximate solution is given by $\tilde{\mathbf{x}} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T$. This method provides an effective strategy for scenarios where rank$(\mathbf{X}^*)$ exceeds one. The Gaussian randomization method is another widely used technique. For a more comprehensive explanation and utilization of the Gaussian randomization method, please refer to Chapter 6, which is discussed in detail.

Chapter 3

Hybrid NOMA Based MIMO Offloading for Mobile Edge Computing in 6G Networks

3.1 Introduction

Increasing demand for both achieving higher data rate to solve computationally intensive tasks timely and connecting more user equipment (UEs) simultaneously have prompted researchers to develop new technologies in the area of wireless communications. The transmission delay time is a comprehensive metric for satisfying these demands.

Table 3.1. Comparison of OMA, NOMA, and H-NOMA

MA Scheme	Advantages	Disadvantages
OMA	Lowest receiver complexity, Lowest signaling overhead	Lowest spectrum efficiency, Serve the lowest number of UEs, Lowest fairness, Near-Far problem
NOMA	Exploit Near-Far problem, Serve the highest number of UEs, Fulfil different QoS requirements, Lower latency compared with OMA, Higher spectral efficiency than OMA	Similar channel gain leads to SIC error propagation, Higher receiver complexity than OMA
H-NOMA	Combining the advantages of OMA and NOMA, Balanced complexity-performance, Lower latency compared with NOMA under limited energy constraint	Highest signaling overhead, Serve a lower number of users compared with NOMA

Non-orthogonal multiple access (NOMA), multiple-input multiple-output (MIMO), and

mobile edge computing (MEC) are promising technologies for minimizing the uplink/downlink transmission delay [47], [5]. Specifically, NOMA, which hosts more than one user in the same sub-carrier by exploiting the power domain, could play a vital role in the next-generation communication networks due to its higher spectral efficiency, lower latency, user fairness, and greater connectivity features compared with the traditional orthogonal multiple access (OMA) techniques [48]. Motivated by the advantage of high throughput due to array and spatial diversity gains, several studies have shown that MIMO will maintain its importance in 5G and beyond [47]. Driven by the increasing applications with computationally intensive tasks, MEC was proposed to reduce the computation time. The main idea behind MEC technology is to bring mini cloud computers to the edge. Therefore, UEs in the cell can enjoy the cloud computing-like facilities by offloading their computationally complex tasks to the MEC server [49].

Existing studies with MEC mainly utilized OMA protocols [50], [51]. Joint optimization of radio and computation resources have been investigated to reduce energy consumption with latency constraints for the OMA-based MIMO-MEC system [50]. In [52], the weighted sum of energy consumption and round transmission delay for OMA-based multi-user MIMO-MEC offloading was minimized by using the semi-definite relaxation (SDR) method. In [53], an inter-user task dependency problem was investigated while minimizing a weighted sum of energy and offloading delay in the time division multiple access (TDMA) based SISO-MEC systems. In [51], a TDMA-based multiple input single output (MISO)-MEC system was integrated with secure wireless power transfer (WPT).

Recently, researchers have demonstrated the superiority of NOMA over OMA in the single-input single-output (SISO)-MEC for a delay minimization problem [5]. In [54], offloading tasks partition ratio and offloading transmit power of the users were jointly optimized to minimize the offloading delay. In [55], the energy minimization problem was studied for a multi-user multi-BS NOMA-MEC network with imperfect channel state information (CSI). In [56], total energy consumption was minimized by optimizing the user clustering, computing and communication resource allocation, and transmit power for the NOMA-based SISO-MEC. In [6], a NOMA-based secure and energy-efficient massive MIMO system was investigated.

Hybrid NOMA (H-NOMA) is a hybrid multiple-access concept that combines NOMA and OMA. More specifically, if there are two H-NOMA users in a cluster, they start uploading/downloading their data concurrently using the NOMA protocol. Once one of the users completes its transmission, the other user switches to the OMA protocol to upload/download its remaining task. The advantages and disadvantages of the H-NOMA, in comparison with OMA and NOMA, are presented in Table 3.1. H-NOMA achieves better delay performance than pure NOMA and OMA as energy consumption is considered [5]. In [57], power allocation, time slot allocation, task assignment, and user grouping methods were utilized to minimize energy consumption in the H-NOMA-based SISO-MEC system.

In SISO-NOMA, two channels can be compared and their corresponding transmit powers can be allocated to the channels, but it is not as easy for MIMO as it is in SISO. The generalized singular value decomposition (GSVD) method, which simultaneously decomposes two matrices into their singular values, was proposed for MIMO-NOMA uplink and downlink transmissions in [58].

Existing studies on NOMA-based MEC were mostly built on SISO transmission [5], [56], [57]. To exploit MIMO's diversity gain and H-NOMA's superior delay performance [5] with balanced spectral efficiency and system complexity features [59], we integrate H-NOMA, MIMO and MEC technologies by the GSVD technique. To this end, an optimal power allocation problem is formulated. The problem is non-convex. Therefore, some insights are provided to transform the non-convex problem into a suboptimal convex form. The delay minimization problem is divided into two subproblems which are represented by two time-frames: T_1 and T_2. T_1 represents the total offloading delay during NOMA transmission, and T_2 is for OMA transmission. In addition, the MIMO channels between UEs and the MEC-assisted base station are decomposed into SISO channels by using the GSVD and the singular value decomposition (SVD) techniques according to the H-NOMA method. Moreover, we mainly focus on T_2 because T_1 is a basic concave problem [5]. In other words, T_1 could be easily solved by numerical methods. Due to the fractional form of T_2, the Dinkelbach method [41] is applied to transform T_2 into a subtractive form. After the transformation, an iterative closed-form solution for T_2 is derived by using the Karush-Kuhn-Tucker (KKT) conditions. Finally, the delay perfor-

mance of the OMA-MIMO-MEC and H-NOMA-MIMO-MEC systems are compared. The effect of the antenna number on delay in the H-NOMA-MIMO-MEC system is also investigated.

3.2 System Model

We consider a MIMO-NOMA-MEC uplink communication scenario in which one MEC-assisted eNodeB communicates with two UEs, as shown in Fig. 3.1. We assume the base station has M antennas, and each UE has K antennas. We consider the H-NOMA scheme in this system model due to its superior delay minimization performance [5]. The model aims to minimize the total offloading time for UE_1 and UE_2. We assume that UE_1 is the near user and has a higher SINR than UE_2. In order to achieve a further reduction in system complexity, it is assumed that the data amount for each user is identical. As shown in Fig. 3.2, UE_1 offloads its task during T_1. Concurrently, UE_2 offloads its task. However, UE_2 might not complete its task in T_1 due to its lower SINR. Therefore, UE_2 needs to continue offloading during T_2 to complete its task. Accordingly, the total delay time for UE_2 can be found by $T_1 + T_2$.

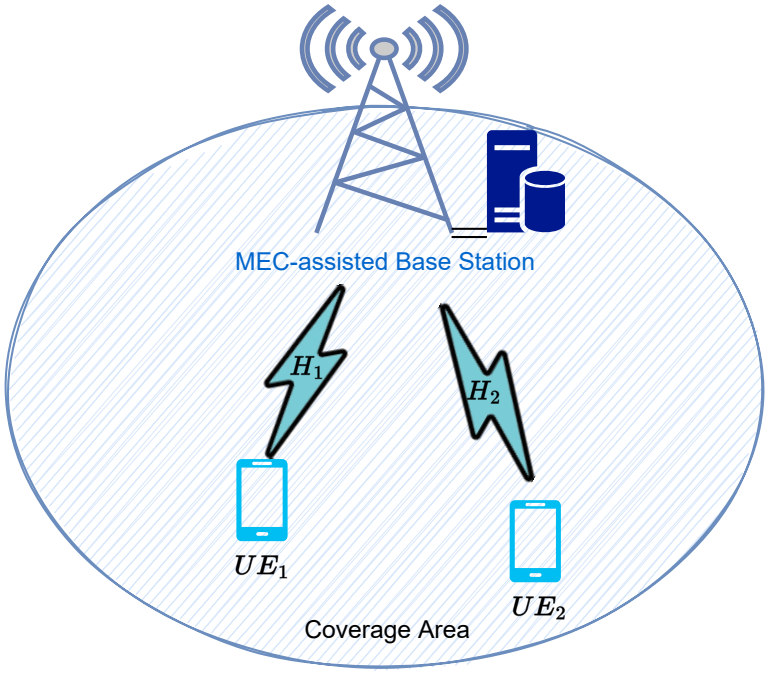

Figure 3.1. H-NOMA based MIMO MEC system model

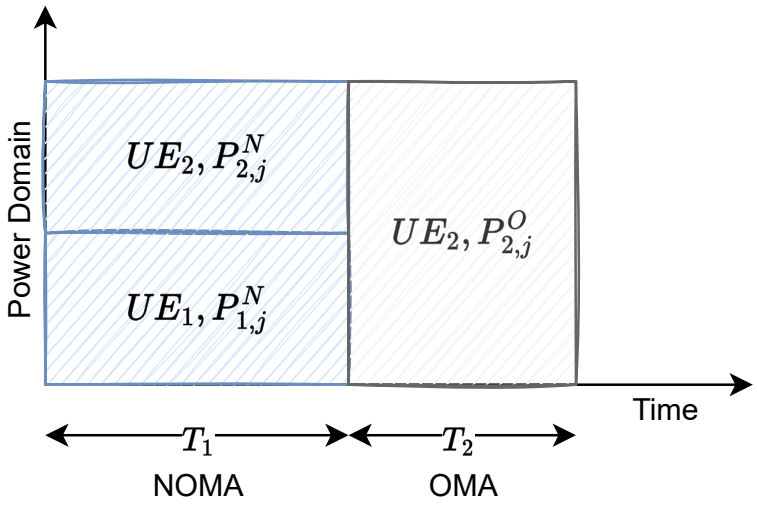

Figure 3.2. A basic concept of hybrid NOMA

Under the time-invariant wireless channel condition, the received signal at the base station can be formulated as follows:

$$\mathbf{y} = \sum_{i=1}^{2} \mathbf{H}_i \mathbf{x}_i + \mathbf{n}, \tag{3.1}$$

where $\mathbf{y}$ is an $M \times 1$ dimensional vector. $\mathbf{x}_i \in \mathbb{C}^{1\times K}$ denotes information vector created by the i-th user, $\mathbf{n} \in \mathbb{C}^{M\times 1}$ denotes a complex additive noise with zero mean and σ_n^2 variance. $\mathbf{H}_i \in \mathbb{C}^{M\times K}$ represents a complex Gaussian channel matrix between UE_i and the base station. $\mathbf{H}_i$ can be decomposed into SISO channels by GSVD as follows:

$$\mathbf{H}_i = \mathbf{U}\mathbf{\Lambda}_i\mathbf{V}_i^H,\ i \in \{1, 2\}, \tag{3.2}$$

where $\mathbf{U}$ is an $M \times M$ matrix, $\mathbf{V}_i$ is an $M \times M$ unitary matrix and $\mathbf{\Lambda}_i$= diag$(\sigma_{i,1}, ..., \sigma_{i,K})$. In addition, we assume that the users have perfect channel state information (CSI). The power of the transmitted signal $x_{i,j}$ is set to be normalized. Therefore, the received signal at the MEC-assisted base station can be expressed as:

$$\mathbf{y} = \sum_{i=1}^{2}\sum_{j=1}^{K} \Lambda_{i,j} P_{i,j} x_{i,j} + \hat{n}_j, \tag{3.3}$$

where $\hat{n}_j$ is still a complex AWGN. $P_{i,j}$ is the power allocation expression for the i-th user's j-th sub-channel. The SIC technique can be used to decode the received signals. According to the applied SIC order, $\mathbf{x_2}$ is decoded first and then it is subtracted from

the superposed signals. Finally, $\mathbf{x_1}$ is decoded from the residue. This strategy is mainly used for the delay-sensitive far user with low target data rate [60]. Therefore, this strategy eliminates the occurrence of non-convex expressions in (3.5).

In the H-NOMA-based system, the achievable maximum data rates are denoted by R_1 and R_2 for UE_1 and UE_2, respectively. Compared to NOMA, these rates in an OMA system, i.e., orthogonal frequency division multiple access (OFDMA), are given by R_3 and R_4.

$$R_1 = B \times \sum_{j=1}^{K} \log_2 \left(1 + \frac{\sigma_{1,j}^2 P_{1,j}^N}{\hat{n}_j^2}\right),\ bits/s/Hz \tag{3.4a}$$

$$R_2 = B \times \sum_{j=1}^{K} \log_2 \left(1 + \frac{\sigma_{2,j}^2 P_{2,j}^N}{\sigma_{1,j}^2 P_{1,j}^N + \hat{n}_j^2}\right),\ bits/s/Hz \tag{3.4b}$$

$$R_3 = \frac{B}{2} \times \sum_{j=1}^{K} \log_2 \left(1 + \frac{\sigma_{3,j}^2 P_{1,j}^O}{\hat{n}_j^2}\right),\ bits/s/Hz \tag{3.4c}$$

$$R_4 = \frac{B}{2} \times \sum_{j=1}^{K} \log_2 \left(1 + \frac{\sigma_{4,j}^2 P_{2,j}^O}{\hat{n}_j^2}\right),\ bits/s/Hz \tag{3.4d}$$

where B is the bandwidth. $\sigma_{1,j}$ and $\sigma_{2,j}$ are the generalized singular values; $\sigma_{3,j}$ and $\sigma_{4,j}$ are singular values of the strong channel and the weak channels, respectively. The power allocation expressions of the j^{th} subchannels for UE_1 and UE_2 are denoted by $P_{1,j}^N$ and $P_{2,j}^N$ in NOMA; $P_{1,j}^O$ and $P_{2,j}^O$ are the power allocation expressions in OMA as illustrated in Fig. 3.2.

3.3 Problem Formulation and Solution

As 6G networks are expected to serve an unprecedented number of UEs with different data rate requirements and power constraints, we formulate the transmission delay problem based on the H-NOMA technique in this section. According to H-NOMA, T_1 and T_2 correspond to the delay time during NOMA and OMA transmission in (3.5). The delay minimization problem can be formulated as

$$\min_{P_{1,j}^N, P_{2,j}^N, P_{2,j}^O} \quad T_1 + T_2 \tag{3.5a}$$

$$\text{s.t.} \quad -P_{max} \leq -\sum_{j=1}^{K} P_{2,j}^{O} \leq 0, \tag{3.5b}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{i,j}^{N} \leq 0,\ i = \{1,2\}, \tag{3.5c}$$

$$T_1 = \frac{N}{B \times \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{1,j}^2 P_{1,j}^N}{\hat{n}_j^2})} \tag{3.5d}$$

$$T_2 = \frac{N - B \times T_1 \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{2,j}^2 P_{2,j}^N}{\sigma_{1,j}^2 P_{1,j}^N + \hat{n}_j^2})}{\frac{B}{2} \times \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{3,j}^2 P_{2,j}^O}{\hat{n}_j^2})}. \tag{3.5e}$$

where (3.5a) is the objective function minimizing total transmission time. Particularly, NOMA and OMA-based transmission time expressions are given in (3.5d) and (3.5e), respectively. The inequality constraints in (3.5b) and (3.5c) denote the transmit power limits for the users. (3.5) can be divided into two sub-problems (3.6) and (3.7).

$$\min_{P_{1,j}^N} \quad T_1 \tag{3.6a}$$

$$\text{s.t.} \quad T_1 = \frac{N}{B \times \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{1,j}^2 P_{1,j}^N}{\hat{n}_j^2})} \tag{3.6b}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{1,j}^{N} \leq 0. \tag{3.6c}$$

$$\min_{P_{1,j}^N, P_{2,j}^N, P_{2,j}^O} \ T_2 \tag{3.7a}$$

$$\text{s.t. } T_2 = \frac{N - T_1 B \times \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{2,j}^2 P_{2,j}^N}{\sigma_{1,j}^2 P_{1,j}^N + \hat{n}_j^2})}{\frac{B}{2} \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{3,j}^2 P_{2,j}^O}{\hat{n}_j^2})} \tag{3.7b}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{2,j}^{O} \leq 0, \tag{3.7c}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{i,j}^{N} \leq 0,\ i = \{1,2\}. \tag{3.7d}$$

where (3.6) only depends on $P_{1,j}^N$ and for this reason (3.6) is a concave optimization problem as in [48]. To solve (3.6), we use CVX, a package for specifying and solving convex programs [46]. Therefore, we assume that $P_{1,j}^N$ and T_1 are fixed for T_2 [1]. However,

[1]While alternating optimization often converges to a favorable local optimum, which might be in proximity to the global optimum, it lacks a general guarantee of always reaching the global optimum, particularly when dealing with complex non-convex

(3.7) is still a non-convex problem owing to its concave-to-convex fractional expression in (3.7b). Fortunately, (3.7b) can be reformulated as a subtractive optimization problem by applying the Dinkelbach method. As $P_{1,j}^{N}$ is fixed for T_2, (3.7) can be rewritten as follows:

$$\min_{P_{2,j}^{N}, P_{2,j}^{O}} \quad T_2 \tag{3.8a}$$

$$\text{s.t.} \quad N = T_2 \sum_{j=1}^{K} \underbrace{\log_2(1 + \frac{\sigma_{3,j}^2 P_{2,j}^{O}}{\hat{n}_j^2})}_{A_j} + \tag{3.8b}$$

$$\frac{N}{const_1} \sum_{j=1}^{K} \underbrace{\log_2(1 + \frac{\sigma_{2,j}^2 P_{2,j}^{N}}{const_2})}_{B_j} \tag{3.8c}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{2,j}^{O} \leq 0, \tag{3.8d}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{i,j}^{N} \leq 0, \tag{3.8e}$$

$$const_1 = \sum_{j=1}^{K} \underbrace{\log_2(1 + \frac{\sigma_{1,j}^2 P_{1,j}^{N*}}{\hat{n}_j^2})}_{C_j} \tag{3.8f}$$

$$const_2 = \sigma_{1,j}^2 P_{1,j}^{N*} + \hat{n}_j^2, \quad i = \{1, 2\}. \tag{3.8g}$$

Given that $P_{1,j}^{N*}$ is fixed, the values of $const_1$ and $const_2$ remain constant. The Dinkelbach transform is applied to (3.8) as follows:

$$f(q) = \max_{P_{2,j}^{N}, P_{2,j}^{O}} \quad \frac{1}{2} \sum_{j=1}^{K} A_j - q \sum_{j=1}^{K} (C_j - B_j) \tag{3.9a}$$

$$\text{s.t.} \quad -P_{max} \leq -\sum_{j=1}^{K} P_{2,j}^{O} \leq 0, \tag{3.9b}$$

$$-P_{max} \leq -\sum_{j=1}^{K} P_{i,j}^{N} \leq 0, \tag{3.9c}$$

$$i = \{1, 2\}, \; q \in \mathbb{R}, \tag{3.9d}$$

where q is the Dinkelbach parameter. Thus, (3.7) is transformed into a convex optimization problem. To obtain the optimal solutions for $P_{2,j}^{N}$ and $P_{2,j}^{O}$, the Lagrange multipliers

problems.

method is applied to (3.9). The Lagrange function of (3.9) is written as follows:

$$\mathcal{L}(P_{2,j}^{O}, P_{2,j}^{N}, \lambda_{1,j}, \lambda_{2,j}, \lambda_3, \lambda_4) = \\ \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{3,j}^2 P_{2,j}^{O}}{\hat{n}_{2,j}^2}) - q\Bigg(\sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{1,j}^2 P_{1,j}^{N}}{\hat{n}_{1,j}^2}) \\ - \sum_{j=1}^{K} \log_2(1 + \frac{\sigma_{2,j}^2 P_{2,j}^{N}}{const_2}) \Bigg) - \lambda_{1,j}(-P_{2,j}^{O}) - \lambda_{2,j}(-P_{2,j}^{N}) \\ - \lambda_3(\sum_{j=1}^{K} P_{2,j}^{O} - P_{max}) - \lambda_4(\sum_{j=1}^{K} P_{2,j}^{N} - P_{max}) \tag{3.10a}$$

where λ_i are the Lagrange multipliers. The KKT conditions are derived to find the optimal solutions as follows:

$$\frac{\partial \mathcal{L}(P_{2,j}^{O}, P_{2,j}^{N}, \lambda_{1,j}, \lambda_{2,j}, \lambda_3, \lambda_4)}{\partial P_{2,j}^{O}} = 0 \tag{3.11a}$$

$$\frac{\partial \mathcal{L}(P_{2,j}^{O}, P_{2,j}^{N}, \lambda_{1,j}, \lambda_{2,j}, \lambda_3, \lambda_4)}{\partial P_{2,j}^{N}} = 0 \tag{3.11b}$$

$$\lambda_{1,j}(-P_{2,j}^{O}) = 0 \tag{3.11c}$$

$$\lambda_{2,j}(-P_{2,j}^{N}) = 0 \tag{3.11d}$$

$$\lambda_3(\sum_{j=1}^{K} P_{2,j}^{O} - P_{max}) = 0 \tag{3.11e}$$

$$\lambda_4(\sum_{j=1}^{K} P_{2,j}^{N} - P_{max}) = 0 \tag{3.11f}$$

$$\sum_{j=1}^{K} P_{2,j}^{O} - P_{max} \leq 0 \tag{3.11g}$$

$$\sum_{j=1}^{K} P_{2,j}^{N} - P_{max} \leq 0 \tag{3.11h}$$

$$-P_{2,j}^{N} \leq 0 \tag{3.11i}$$

$$-P_{2,j}^{O} \leq 0 \tag{3.11j}$$

$$\lambda_{1,j}, \lambda_{2,j}, \lambda_3, \lambda_4 \geq 0. \tag{3.11k}$$

The optimal solutions for $P_{2,j}^{N}$ and $P_{2,j}^{O}$ are given in Lemma 3.3.1. Algorithm 1 describes Dinkelbach's method-based closed-form solution.

Lemma 3.3.1. *H-NOMA power allocation policy*

$$P_{2,j}^{O*} = \max\left(\frac{P_{max} + \sum_{j=1}^{K} \frac{\hat{n}_j^2}{\sigma_{3,j}^2}}{K} - \frac{\hat{n}_j^2}{\sigma_{3,j}^2}, 0\right), \tag{3.12a}$$

$$P_{2,j}^{N*} = \max\left(\frac{P_{max} + \sum_{j=1}^{K} \frac{const_2}{\sigma_{2,j}^2}}{K} - \frac{const_2}{\sigma_{2,j}^2}, 0\right), \tag{3.12b}$$

where $\max(a, b)$ *denotes the maximum of* a *and* b*.*

Proof. Please see Appendix 3.5. □

Algorithm 1 Optimal power allocation algorithm for T_2

1: Find T_1 by a convex optimization package (e.g., CVX)
2: Plug $P_{1,j}^{N}$ in (3.7b)
3: Set $t = 0, q = 0, \Delta \to 0$
4: **if** $f(q) > \Delta$ **then**
5: $t = t + 1$
6: Update $P_{2,j}^{O}$ and $P_{2,j}^{N}$ by Lemma 3.3.1
7: Update q as $q = \frac{\frac{1}{2}\sum_{j=1}^{K} A_j}{\sum_{j=1}^{K}(C_j - B_j)}$
8: **end if**
9: $P_{2,j}^{N*}$ and $P_{2,j}^{O*}$ are found by using $q = q_t$ in Lemma 3.3.1.

Complexity Analysis

The time complexity of the proposed algorithm is analyzed in this subsection. Algorithm 1 consists of two loops: the outer loop is to apply the Dinkelbach algorithm and the inner loop, which is a water filling-based solution, is to specify the number of power allocated sub-channels (K) and to assign optimal power. The Dinkelbach parameter q is updated in each iteration until $f(q) < \Delta$. The computational complexity of the Dinkelbach algorithm is $\mathcal{O}(T)$, where $\mathcal{O}(.)$ describes the upper bound of the time complexity and T is the number of iterations required for convergence of the Dinkelbach algorithm [61]. The required number of operations, at worst, for the inner loop, is $\mathcal{O}(K)$, where K is the minimum rank of the near and the far users' channel matrices. Therefore, the proposed algorithm has a $\mathcal{O}(TK)$ complexity.

3.4 Numerical Results

In this section, we evaluate the performance of the proposed H-NOMA-MIMO-MEC system. The simulation parameters are given in Table 3.2.

Table 3.2. Simulation parameters

Parameter	Value
Noise spectral efficiency, N_0	-174 dBm/Hz
Bandwidth, B	10 MHz
Path loss component, α	3.2
Error tolerance, ϵ	10e-3
The radius of the inner ring	40 m
The radius of the outer ring	125 m
The number of bits needs to be offloaded, N	1 Gbit

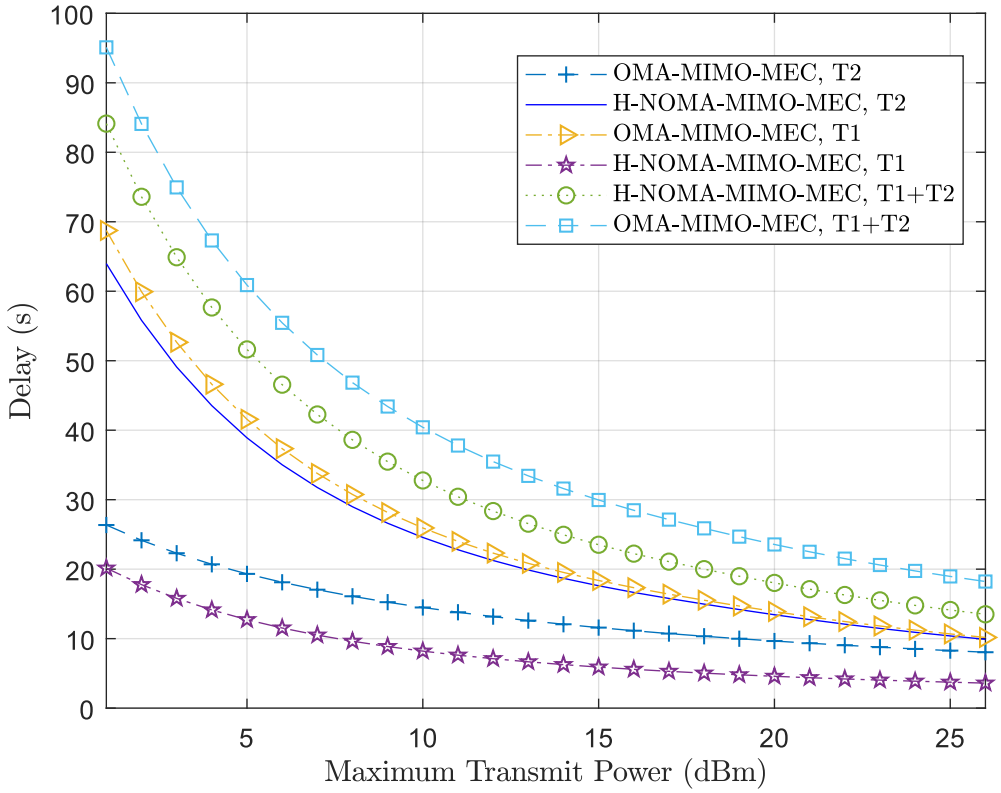

Figure 3.3. Delay performance comparison of H-NOMA-MIMO-MEC with OMA-MIMO-MEC

Fig. 3.3 demonstrates the offloading delay performances of the H-NOMA and OMA-based MIMO-MEC systems. The base station and the UEs are equipped with three an-

tennas. The figure shows that the H-NOMA-MIMO-MEC performs better than OMA-MIMO-MEC, particularly at higher power levels. This is because the weak NOMA user suffers from co-channel interference at low SNR. Also, it can be concluded that increasing transmit power has less impact on delay minimization compared with bandwidth. This is one of the key advantages of using NOMA. Fig. 3.3 shows that the H-NOMA-MIMO-MEC improves delay performance by an average of 11% compared to the OMA-MIMO-MEC.

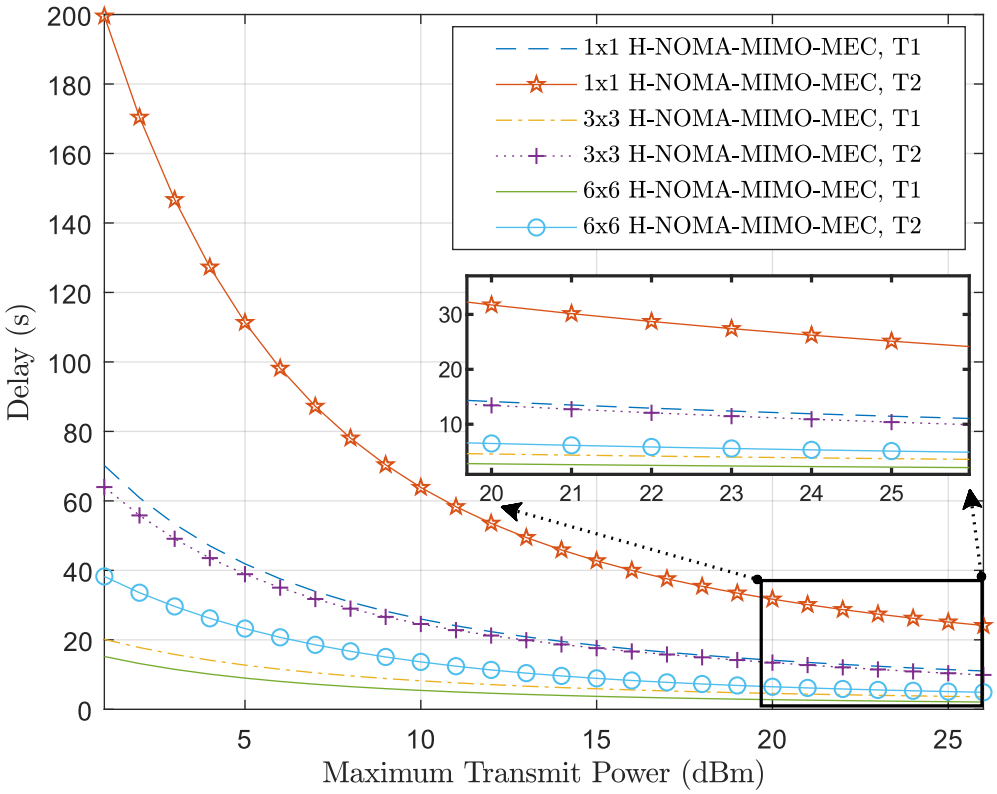

Figure 3.4. Delay performance of antenna numbers in H-NOMA-MIMO-MEC system

In Fig. 3.4, the impact of antenna numbers on transmission delay is demonstrated. Transmission delay is closely related to the antenna number. The figure shows that the proposed H-NOMA-based MIMO-MEC achieves better delay performance than SISO-MEC. The most striking result from the figure is that having more antennas improve delay performance significantly on the low transmit power region.

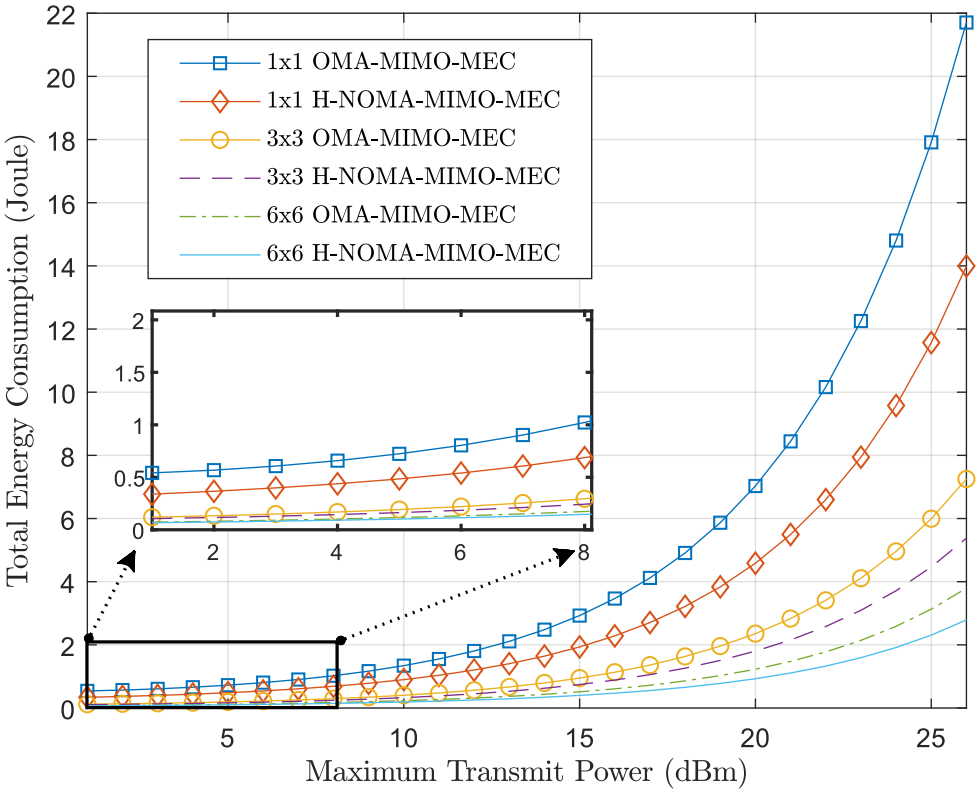

Figure 3.5. Transmit energy consumption versus power budget

In Fig. 3.5, energy consumption of the H-NOMA based MIMO offloading system is compared with OMA. It can be seen that NOMA yields better results for each antenna configuration. Since the power budgets are the same for the UEs in H-NOMA and OMA transmissions and the proposed H-NOMA based system completes offloading earlier than OMA, energy efficiency is improved.

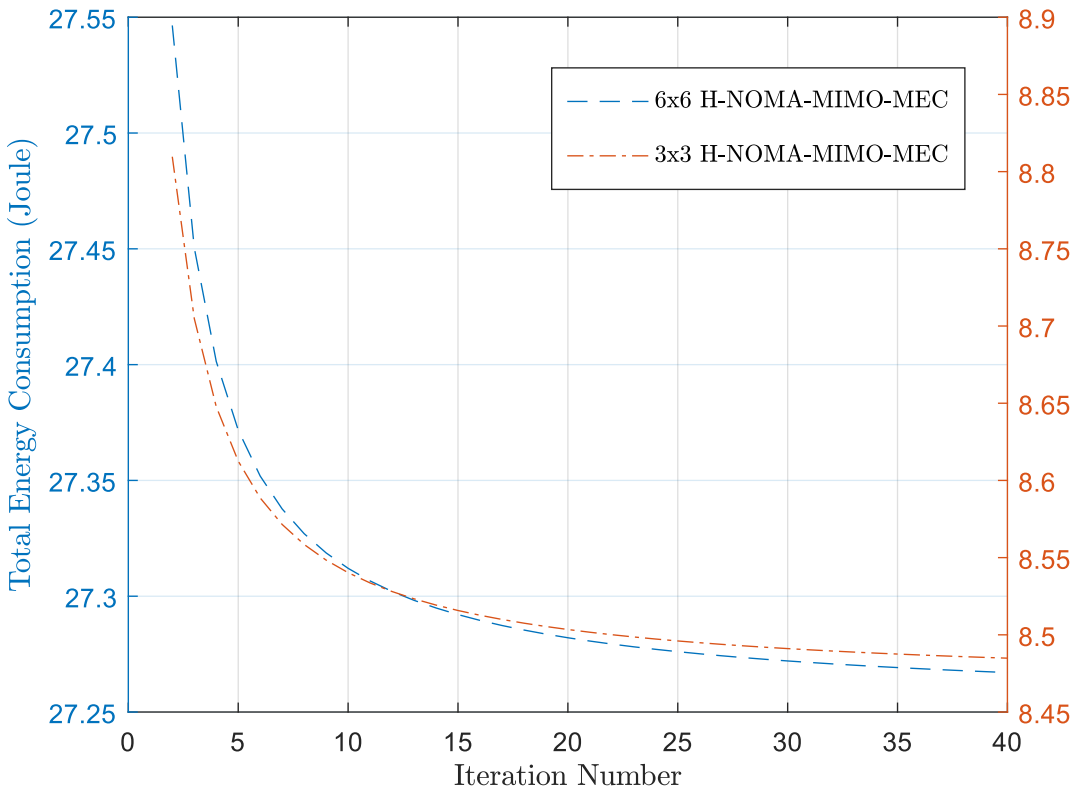

Figure 3.6. Convergence performance of the proposed algorithm in terms of iteration number

Fig. 3.6 presents sub-linearly convergence of the proposed algorithm versus iteration number. It can be observed in Fig. 3.6 that the algorithm significantly converges within 20 iterations for H-NOMA-MIMO-MEC.

3.5 Conclusion of Chapter 3

Recent developments in wireless communication have increased the need for spectrum efficiency, energy efficiency, and data rate. This chapter presented the first study to combine H-NOMA, MIMO and MEC technologies for delay minimization. In this chapter, the H-NOMA-MIMO-MEC offloading delay was investigated. Due to the concave-to-convex fractional nature of the problem, the Dinkelbach method was used to eliminate fractional expression. Finally, an iterative closed-form solution was obtained. According to the simulation results, the proposed method improved the delay performance and reduced the total energy consumption of the MIMO-MEC.

Appendix

Proof of Lemma 1

Proof. To find the optimal value for $P_{2,j}^{O*}$, it is necessary to show that there is no more feasible descent for (3.9) in terms of $P_{2,j}^{O}$. Therefore, (3.11a) is executed as follows:

$$\frac{\frac{\sigma_{3,j}^2}{n_j^2}}{1+\frac{\sigma_{3,j}^2 P_{2,j}^O}{n_j^2}} + \lambda_{1,j} - \lambda_3 = 0. \tag{3.13a}$$

$\boldsymbol{\lambda}_{1,j}$ is the vector consisting of the Lagrange multipliers $(\lambda_{1,1}, ..., \lambda_{1,K})$ corresponding to $(P_{2,1}^O, ..., P_{2,K}^O)$, respectively. From the complimentary slackness condition in $(3.11c)$, either $P_{2,j}^O$ or $\lambda_{1,j}$ must be zero. When a MIMO channel is decomposed into decoupled SISO channels, some may not be feasible for power allocation. Accordingly, we introduce a variable (L) indicating the number of power-allocated SISO channels. Hence, we can eliminate $\lambda_{1,j}$ expression in (3.13a) for the weak SISO channels. Therefore, λ_3 can be written as follows:

$$\lambda_3 = \frac{K}{P_{max} + \sum_{j=1}^{K} \frac{\hat{n}_j^2}{\sigma_{3,j}^2}}. \tag{3.13b}$$

It is clear from (3.13b) that λ_3 is positive. Therefore, (3.13c) can be obtained from (3.11e).

$$\sum_{j=1}^{K} P_{2,j}^O - P_{max} = 0. \tag{3.13c}$$

Furthermore, $P_{2,j}^O$ can be simplified as follows:

$$P_{2,j}^O = \frac{\sigma_{3,j}^2 - \lambda_3 \hat{n}_j^2}{\lambda_3 \sigma_{3,j}^2}. \tag{3.13d}$$

Finally, by combining (3.13b) with (3.13d), the optimal expression for $P_{2,j}^O$ becomes

$$P_{2,j}^{O*} = \frac{P_{max} + \sum_{j=1}^{K} \frac{\hat{n}_j^2}{\sigma_{3,j}^2}}{K} - \frac{\hat{n}_j^2}{\sigma_{3,j}^2}. \tag{3.13e}$$

We follow similar steps to those above to find the optimal expression for $P_{2,j}^{N*}$. (3.10a) is differentiated with respect to $P_{2,j}^N$ as follows:

$$\frac{q\frac{\sigma_{2,j}^2}{const_2}}{1+\frac{\sigma_{2,j}^2 P_{2,j}^N}{const_2}} + \lambda_{2,j} - \lambda_4 = 0. \tag{3.14a}$$

$\lambda_{2,j}$ goes zero for power allocated sub-channels. Thus, λ_4 becomes

$$\lambda_4 = \frac{q\sigma_{2,j}^2}{const_2 + \sigma_{2,j}^2 P_{2,j}^N}. \tag{3.14b}$$

$P_{2,j}^N$ can be manipulated as follows:

$$P_{2,j}^N = \frac{q\sigma_{2,j}^2 - \lambda_4 const_2}{\lambda_4 \sigma_{2,j}^2}. \tag{3.14c}$$

We rewrite (3.11f) by using (3.14c) as

$$\sum_{j=1}^{K} \frac{q\sigma_{2,j}^2 - \lambda_4 const_2}{\lambda_4 \sigma_{2,j}^2} - P_{max} = 0. \tag{3.14d}$$

We combine (3.14b) with (3.14d). Finally, the optimal power allocation expression for $P_{2,j}^N$ is as follows:

$$P_{2,j}^{N*} = \frac{P_{max} + \sum_{j=1}^{K} \frac{const_2}{\sigma_{2,j}^2}}{K} - \frac{const_2}{\sigma_{2,j}^2}. \tag{3.14e}$$

Therefore, the proof for Lemma 3.3.1 is complete. □

Chapter 4

Secrecy Sum Rate Maximization for a MIMO-NOMA Uplink Transmission in 6G Networks

4.1 Introduction

Chapter 3 focused on minimizing offloading delay. The significance of security in critical networks is evident, alongside the objective of minimizing delay. Building upon this foundation, Chapter 4 aims to investigate the problem of maximizing the secrecy sum rate in NOMA-Based MIMO-NOMA networks.

Spectral efficiency, massive connectivity, and low latency are essentials for wireless communication. Besides, data security is an increasingly important area in the sixth generation of radio access networks (6G-RANs). This is because there is a sharp increase in the number of critical Internet of Things (IoT) devices and services, such as autonomous cars and mobile online banking applications; however, wireless channels are vulnerable to malicious eavesdropping attacks. To protect wireless communication, physical layer security (PLS) methods have gained importance before implementing high-level cryptography techniques [22].

Non-orthogonal multiple access (NOMA), superposing multiple users' signals in the power domain, is a promising technique to meet those requirements. Furthermore, being compatible with the present schemes in use, such as time division multiple access

(TDMA) or orthogonal frequency division multiple access (OFDMA), is a remarkable feature of NOMA [62]. There is a growing body of literature that proves the supremacy of NOMA over orthogonal multiple access (OMA) in terms of sum ergodic capacity [63], user fairness [4] and green communication [64].

In recent years, there has been an increasing interest in the physical layer security of NOMA networks. In [65], the optimal closed-form power allocation policy was derived to maximize the secrecy rate in a single-input, single-output (SISO)-NOMA downlink system. This study revealed that NOMA has a better secrecy sum rate (SSR) than OMA in the single antenna configuration [65]. In [66], a maximum ratio transmission beamforming was used to secure the far user from internal eavesdropping in a two-user multiple-input single-output (MISO)-NOMA downlink network. In [67], an artificial noise (AN) based transmit beamforming scheme and an inter-user-interference-based scheme were used to disturb the eavesdropper in a MISO-NOMA downlink network. In [66], the co-operative rate splitting method was introduced to maximize the SSR in a MISO downlink scenario. According to this method, the message for the legitimate users was split into two parts: common and private parts. Herein the common signal was used as AN; the private signals can be decoded by applying successive interference cancellation (SIC) techniques. In [68], a secure beamforming design was studied for a private unicasting user among multicasting users in an unmanned aerial vehicle (UAV)-assisted MISO-NOMA downlink network. In [69], a secure beamforming design was studied in a MIMO-NOMA downlink network with and without having perfect channel state information (CSI). In [70], the SSR was maximized in a MISO-NOMA-based simultaneous wireless information and power transfer (SWIPT) system. The SSR maximization problem for a MISO-NOMA downlink multiple-input multiple-output (MIMO)-NOMA network was converted to a second-order cone programming (SOCP) problem in [69]. To improve PLS for a cognitive radio network (CRN) assisted downlink MIMO-NOMA system, the zero-forcing and eigen beamforming techniques were proposed in [71]. In [72], the artificial noise injection technique was utilized for a massive MIMO-assisted NOMA downlink network. In [73], the SSR optimization problem for a downlink MIMO-NOMA network which hosts multiple MIMO users and a multiple-antenna-equipped eavesdropper was investigated.

However, a few studies in the literature regarding the SSR have only focused on uplink PLS system [74]-[75]. In [74], a secure single input multiple outputs (SIMO)-NOMA system was studied by optimizing the SIC order and the power allocation to the legitimate users. In [76], effective secrecy throughput (EST) was proposed as a new physical layer security metric that helps decide on design parameters in the SIMO-NOMA networks. In [77], a cooperative dynamic jamming system was proposed to increase the EST performance of the SIMO-NOMA network. In [75], an alternating optimization-based algorithm was introduced to solve the secrecy problem in an OMA-based MIMO uplink network.

The generalized singular value decomposition (GSVD) based linear precoding technique has been widely considered in a growing body of literature because the GSVD technique has low complexity and high performance [78]. However, few studies investigate the GSVD-based SSR maximization problem in NOMA-MIMO networks. In [79], the SSR maximization problem against legitimate internal users was investigated in a downlink MIMO-NOMA system. In [80], an energy-efficient precoding design was presented in a cooperative two-way relay transmission with an insecure relay. To the authors' knowledge, the GSVD-based SSR maximization problem with a jammer and multiple eavesdroppers in a NOMA-MIMO uplink network has not been studied. The contributions of this chapter to this growing area of literature are given as follows:

- A novel GSVD-based uplink MIMO-NOMA network is investigated in PLS scenarios, where multiple external eavesdroppers and a friendly jammer are considered. The researched system is different from [74]-[75], [80] due to the adopted schemes or system configuration.
- To improve the considered system, a non-concave SSR maximization problem is formulated. By equivalently transforming the norm functions in the original problem to the trace functions in the logarithmic expressions, a problem with the difference of convex (DC) programs is reformulated. After that, the first-order Taylor approximation method is utilized to convert the DC problem into a suboptimal concave problem, and an SCA-based algorithm is proposed. The property of the proposed algorithm is analyzed, including complexity and convergence.

- Simulation results indicate that GSVD-based MIMO-NOMA can significantly outperform conventional MIMO-NOMA regarding SSR. Moreover, the performance gap between NOMA and OMA schemes is revealed based on the same optimization method.

Notations

Matrices and vectors are presented by uppercase and lowercase boldface letters, respectively. $(\cdot)^H$ and $(\cdot)^{-H}$ denote the Hermitian and the inverse of Hermitian, respectively. Tr$(\cdot)$, $(\cdot)^{-1}$ and $\|\cdot\|$ stand for the trace, the inverse, and the Euclidean norm of a matrix, respectively. $|\cdot|$ denotes for the absolute value of a complex number. The notation $diag(\cdot)$ represents a diagonal matrix where the diagonal elements are from a vector and $[x]^+$ represents $\max(x, 0)$. $BlkDiag(\mathbf{A}, \mathbf{B}, \mathbf{C})$ generates a block diagonal matrix from $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ matrices.

4.2 System Model and Problem Formulation

This chapter investigates a NOMA-based uplink network with one BS and two legitimate users, as depicted in Figure 4.1. The antennas at the BS, at the near user, and at the far user are N_B, N_N, and N_F, respectively. The legitimate users transmit their private data to the BS simultaneously in the presence of a friendly jammer with N_J antennas and P eavesdroppers with a single antenna[1].

[1]This RAN setup may be practical for some powerful base stations [81].

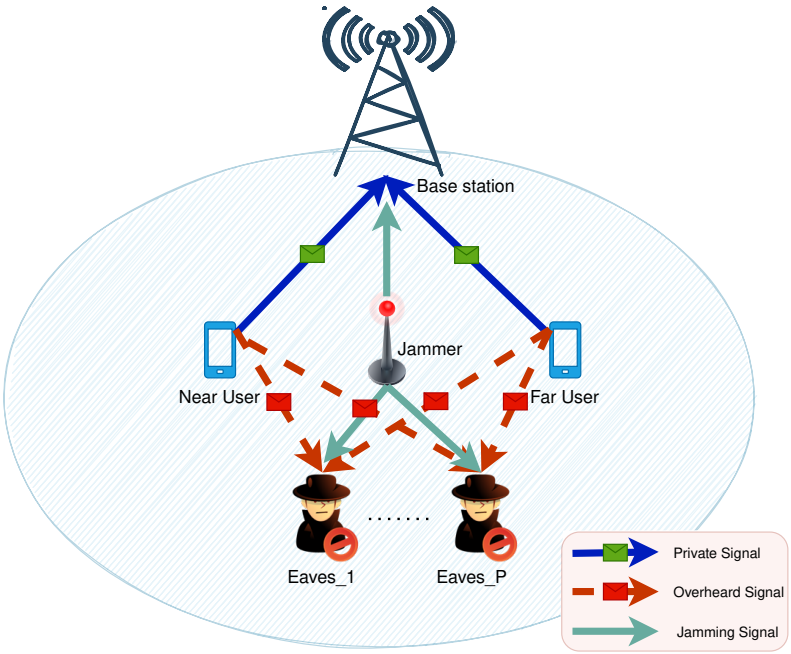

Figure 4.1. NOMA-MIMO wiretap channel with a friendly jammer

The jammer is assumed to have more antennas than the BS [81]. Having more antennas enables the friendly jammer to utilize null space beamforming (NSBF), concentrating jamming power towards the eavesdroppers while creating spatial nulls on the BS. The received signals at the BS and eavesdropper can be expressed as follows:

$$\mathbf{y}_B = \overbrace{\underbrace{\mathbf{H}_{BN}\mathbf{x}_N}_{\text{Near User Signal}} + \underbrace{\mathbf{H}_{BF}\mathbf{x}_F}_{\text{Far User Signal}}}^{\text{Legitimate Signals}} + \underbrace{\mathbf{H}_{BJ}\mathbf{x}_J}_{\text{Jamming Signal}} + \underbrace{\mathbf{n}_B}_{\text{Noise}}, \tag{4.1a}$$

$$y_{E_p} = \overbrace{\underbrace{\mathbf{h}_{EN_p}\mathbf{x}_N}_{\text{Near User Signal}} + \underbrace{\mathbf{h}_{EF_p}\mathbf{x}_F}_{\text{Far User Signal}}}^{\text{Overheard Signals}} + \underbrace{\mathbf{h}_{EJ_p}\mathbf{x}_J}_{\text{Jamming Signal}} + \underbrace{n_{E_p}}_{\text{Noise}}, \tag{4.1b}$$

where $\mathbf{H}_{BN} \in \mathbb{C}^{N_B \times N_N}$, $\mathbf{H}_{BF} \in \mathbb{C}^{N_B \times N_F}$ and $\mathbf{H}_{BJ} \in \mathbb{C}^{N_B \times N_J}$ are complex Gaussian flat-fading MIMO channel matrices from the near user, the far user and the jammer to the BS, respectively. $\mathbf{h}_{EN_p} \in \mathbb{C}^{1 \times N_N}$, $\mathbf{h}_{EF_p} \in \mathbb{C}^{1 \times N_F}$ and $\mathbf{h}_{EJ_p} \in \mathbb{C}^{1 \times N_J}$ are the Rayleigh flat-fading MISO channel vectors from the near user, the far user and the jammer to the p^{th} $(p \in \{1, ..., P\})$ eavesdropper, respectively. $\mathbf{x}_N \in \mathbb{C}^{N_N \times 1}$ and $\mathbf{x}_F \in \mathbb{C}^{N_F \times 1}$ are the precoded information bearing vectors, $\mathbf{x}_J \in \mathbb{C}^{1 \times N_J}$ is the precoded jamming vector and $\mathbf{n}_B \in \mathbb{C}^{N_B \times 1}$ and $n_{E_p} \in \mathbb{C}^{1 \times 1}$ are the additive Gaussian noises with variance $\sigma_{\mathbf{B}}$ and σ_{E_P}, respectively. The channel coefficient vectors and matrices can be expressed as $\frac{\mathbf{g}}{\sqrt{d^{-\alpha}}}$, where $\mathbf{g} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the small scale Rayleigh fading gain, α is the path-loss exponent, and d is the distance between transmitter and receiver.

It is assumed that the the perfect CSI of $\mathbf{H}_{BJ}$ and $\mathbf{h}_{EJ_p}$ are known by the friendly jammer, and the CSI of the $\mathbf{H}_{BN}$ and $\mathbf{H}_{BF}$ are known by the legitimate users and the BS[2] [79]. By applying the GSVD method, $\mathbf{H}_{BN}$ and $\mathbf{H}_{BF}$ can be decomposed as follows.

Theorem 4.2.1. *[82]: Let consider that* $\mathbf{C}$ *consists of* $(\mathbf{H}_{BN}^H, \mathbf{H}_{BF}^H)$*, k is the rank of* $\mathbf{C}$*,* r *is equal to* $k - rank(\mathbf{H}_{BF}^H)$*, and* s *is found by* $rank(\mathbf{H}_{BN}^H) + rank(\mathbf{H}_{BF}^H) - k$*. There exists* $\mathbf{U}_N \in \mathbb{C}^{N_N \times N_N}$ *and* $\mathbf{U}_F \in \mathbb{C}^{N_F \times N_F}$ *are unitary matrices,* $\mathbf{V} \in \mathbb{C}^{N_B \times N_B}$ *is a non-singular matrix,* $\mathbf{M} \in \mathbb{C}^{k \times k}$ *and* $\mathbf{0}$ *is a zero matrix of order* $k \times (n-k)$ *such that*

$$\mathbf{V}^{-1}\mathbf{H}_{BN}\mathbf{U}_N = \mathbf{\Lambda}_N(\mathbf{M}, \mathbf{0}_{BN}), \tag{4.2}$$

$$\mathbf{V}^{-1}\mathbf{H}_{BF}\mathbf{U}_F = \mathbf{\Lambda}_F(\mathbf{M}, \mathbf{0}_{BF}), \tag{4.3}$$

where $\mathbf{\Lambda}_N \cong BlkDiag(\mathbf{I_N}, \mathbf{D_N}, \mathbf{0_N})$*,* $\mathbf{\Lambda}_F \cong BlkDiag(\mathbf{I_F}, \mathbf{D_F}, \mathbf{0_F})$*,* $\mathbf{I_F}$ *and* $\mathbf{I_N}$ *are* $r \times r$ *and* $(k-r-s) \times (k-r-s)$ *identity matrices.* $\mathbf{D_N}$ *denotes the diagonal coefficients of* $\mathbf{H}_{BN}$*, i.e,* $\mathbf{D_N} = diag(\lambda_{n,r+1}, ..., \lambda_{n,r+s})$ *and* $\mathbf{D_F}$ *denotes the coefficients of* $\mathbf{H}_{BF}$*, i.e,* $\mathbf{D_F} = diag(\lambda_{f,r+1}, ..., \lambda_{f,r+s})$ *are* $s \times s$ *diagonal matrices which include the generalised singular values of* $\mathbf{H}_{BN}$ *and* $\mathbf{H}_{BN}$*, respectively.* $\mathbf{0_N}$ *and* $\mathbf{0_F}$ *are* $(N_N - r - s) \times (k - r - s)$ *and* $(N_F - r - s) \times (k - r - s)$ *zero matrices, respectively.*

As an uplink scenario is investigated in this chapter, the signals from the legitimate users are precoded first and decoded at the BS. Therefore, the precoding matrix becomes $\mathbf{U}_N$ for the near user, and $\mathbf{U}_F$ for the far user, and the decoding matrix at the BS becomes $\mathbf{V}^{-1}$. The precoded signals transmitted by the near user, the far user, and the jammer can be presented as follows:

$$\mathbf{x}_N = \mathbf{U}_N\sqrt{\mathbf{P}_N}\mathbf{s}_N,\ \mathbf{x}_F = \mathbf{U}_F\sqrt{\mathbf{P}_F}\mathbf{s}_F,\ \text{and}\ \mathbf{x}_J = \mathbf{w}_J\mathbf{z}_J, \tag{4.4}$$

where $\mathbf{P}_N = diag(p_1^N, .., p_{N_N}^N)$ and $\mathbf{P}_F = diag(p_1^F, .., p_{N_F}^F)$ respectively denote the power allocation matrices of the near user and the far user, and $\mathbf{w}_J \in \mathbb{C}^{N_J \times 1}$ denotes the beamforming vector of the jammer. The power of the transmitted symbols $\mathbf{s}_N \in \mathbb{C}^{N_N \times 1}$, $\mathbf{s}_F \in \mathbb{C}^{N_F \times 1}$ and artificial noise vector $\mathbf{z}_J \in \mathbb{C}^{1 \times N_J}$ are normalized with one and the transmit power of the jammer is P_{jam}. Moreover, NSBF is utilized at the jammer to gen-

[2]It's essential to note that estimating the eavesdropper's CSI can be challenging, especially when the eavesdropper is intentionally trying to remain hidden or when the channel conditions are highly dynamic. However, advanced signal-processing techniques can estimates the eavesdropper's channel information.

erate $\mathbf{w}_J$, which is in null space of $\mathbf{H}_{BJ}$ such that $\mathbf{H}_{BJ}\mathbf{w}_J = 0$. Therefore, after applying the GSVD, the received signal at the BS can be written as follows:

$$\begin{aligned}\underbrace{\mathbf{V}^{-1}\mathbf{y}_B}_{\tilde{\boldsymbol{y}}_B} &= \mathbf{V}^{-1}\mathbf{H}_{BN}\mathbf{U}_N\sqrt{\mathbf{P}_N}\mathbf{s}_N + \mathbf{V}^{-1}\mathbf{H}_{BF}\mathbf{U}_F\sqrt{\mathbf{P}_F}\mathbf{s}_F \\ &\quad + \mathbf{V}^{-1}\mathbf{H}_{BJ}\mathbf{w}_J\mathbf{z}_J + \underbrace{\mathbf{V}^{-1}\mathbf{n}_B}_{\tilde{\boldsymbol{n}}_B}, \\ \tilde{\boldsymbol{y}}_B &= \boldsymbol{\Lambda}_N\sqrt{\mathbf{P}_N}\mathbf{s}_N + \boldsymbol{\Lambda}_F\sqrt{\mathbf{P}_F}\mathbf{s}_F + \tilde{\boldsymbol{n}}_B. \end{aligned} \tag{4.5}$$

Even though MIMO channels cannot be compared directly, it is assumed that the channel gain of the near user is stronger than that of the far user due to a lower large-scale path loss effect. Hence, according to the CSI-based SIC principle, the BS first decodes $\mathbf{x_N}$ by treating $\mathbf{x_F}$ as interference. After $\mathbf{x_N}$ is decoded and reconstructed, it is subtracted from $\mathbf{y_B}$. The remaining residue consists of $\mathbf{x_F}$ and $\mathbf{n_B}$. Finally, from the residue signal, $\mathbf{x_F}$ can be retrieved. The instantaneous signal-to-interference-plus-noise ratios (SINR) of the near user and far user at the BS can be expressed as follows.

$$\begin{aligned} SINR^N_{B,i} &= \frac{\lambda^2_{n,i}p^N_i}{\lambda^2_{f,i}p^F_i + (V\mathbf{H}_{BJ}\mathbf{w}_J\mathbf{w}^H_J\mathbf{H}^H_{BJ}\mathbf{V})_{i,i} + \sigma^2_{B,i}} \\ &= \frac{\lambda^2_{n,i}p^N_i}{\lambda^2_{f,i}p^F_i + \sigma^2_{B,i}}, \end{aligned} \tag{4.6a}$$

$$\begin{aligned} SINR^F_{B,i} &= \frac{\lambda^2_{f,i}p^F_i}{(\mathbf{V}\mathbf{H}_{BJ}\mathbf{w}_J\mathbf{w}^H_J\mathbf{H}^H_{BJ}\mathbf{V})_{i,i} + \sigma^2_{B,i}} \\ &= \frac{\lambda^2_{f,i}p^F_i}{\sigma^2_{B,i}}. \end{aligned} \tag{4.6b}$$

The received signal at the eavesdropper is given by

$$y_{E_p} = \mathbf{h}_{EN_p}\mathbf{U}_N\sqrt{\mathbf{P}_N}\mathbf{s}_N + \mathbf{h}_{EF_p}\mathbf{U}_F\sqrt{\mathbf{P}_F}\mathbf{s}_F + \mathbf{h}_{EJ_p}\mathbf{w}_J\mathbf{z}_J + n_{E_p}. \tag{4.7}$$

Additionally, the assumption is made that the eavesdropper is capable of employing SIC. Considering the discourse in Chapter 2, where the SIC order's influence on the sum capacity in uplink NOMA networks remains unaltered, it can be deduced that the change between (4.8a) and (4.8b) do not influence the sum capacity. The SINR for the near user

and the far user at the eavesdropper can be respectively formulated as follows:

$$SINR_{E_p}^{N} = \frac{\|\mathbf{h}_{EN_p}\mathbf{U}_N\sqrt{\mathbf{P}_N}\|^2}{\|\mathbf{h}_{EF_p}\mathbf{U}_F\sqrt{\mathbf{P}_F}\|^2 + P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2}, \tag{4.8a}$$

$$SINR_{E_p}^{F} = \frac{\|\mathbf{h}_{EF_p}\mathbf{U}_F\sqrt{\mathbf{P}_F}\|^2}{P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2}. \tag{4.8b}$$

The SSR maximization problem is formulated to be solved at the BS as follows:

$$R_s = \max_{\mathbf{P}_N,\mathbf{P}_F} \quad [R_B - \sum_{p=1}^{P} R_{E_p}]^+ \tag{4.9a}$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{P}_F) + \text{Tr}(\mathbf{P}_N) \leq P_{max}, \tag{4.9b}$$

$$\mathbf{P}_F \geq 0, \mathbf{P}_N \geq 0, \tag{4.9c}$$

$$R_B^i \geq Q_i, \quad i \in \{F, N\}, \tag{4.9d}$$

$$R_{E_p} \geq 0, \quad (p = 1....P), \tag{4.9e}$$

where R_B and R_{E_p} denote the capacity of the BS and the p^{th} eavesdropper, respectively. R_s is the SSR of the proposed system. P_{max} is the power budgets and Q_F and Q_N are the QoS requirements for the far and the near user, respectively. (4.9b) indicates that the users cannot exceed the power budget, (4.9c) makes sure that the transmit powers are non-negative, and (4.9d) defines the QoS constraints for the legitimate users. (4.9e) guarantees that the capacity of the p^{th} eavesdropper cannot be negative.

4.3 Problem Solution and Performance Analysis

The sum capacity at the BS and the p^{th} eavesdropper are given by (4.10a) and (4.10b), respectively.

$$\begin{aligned} R_B =& \sum_{i=1}^{N_N} \log_2(1 + SINR_{B,i}^{N}) + \sum_{j=1}^{N_F} \log_2(1 + SINR_{B,j}^{F}), \\ =& \sum_{i=1}^{L} \log_2\left(1 + \frac{\lambda_{n,i}^2 p_i^N + \lambda_{f,i}^2 p_i^F}{\sigma_{B,i}^2}\right), \end{aligned} \tag{4.10a}$$

$$R_{E_p} = \log_2(1 + SINR_{E_p}^{N}) + \log_2(1 + SINR_{E_p}^{F}),$$

$$= \log_2 \left(1 + \frac{\|\mathbf{h}_{EN_p} \mathbf{U}_N \sqrt{\mathbf{P}_N}\|^2 + \|\mathbf{h}_{EF_p} \mathbf{U}_F \sqrt{\mathbf{P}_F}\|^2}{P_{jam} \|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2} \right), \tag{4.10b}$$

where L denotes the number of parallel virtual SISO channels between the legitimate users and the BS. Also, L is the minimum rank number of the legitimate channels [22]. In order to guarantee the quality of service (QoS) of legitimate users, the following constraints should be satisfied by the far and near users, respectively.

$$R_B^N = \sum_{i=1}^{L} \log_2 \left(\frac{\lambda_{n,i}^2 p_i^N + \lambda_{f,i}^2 p_i^F + \sigma_{B,i}^2}{\sigma_{B,i}^2} \right) \geq Q_N + Q_F, \tag{4.11a}$$

$$R_B^F = \sum_{i=1}^{L} \log_2 \left(\frac{\lambda_{f,i}^2 p_i^F + \sigma_{B,i}^2}{\sigma_{B,i}^2} \right) \geq Q_F. \tag{4.11b}$$

In order to solve (4.9), which is a non-concave and NP-hard problem [83], the Taylor approximation method is provided to convert it into a suboptimal concave problem. To do this end, some mathematical manipulations are applied first as follows:

$$\begin{aligned} \|\mathbf{h}_{EN_p} \mathbf{U}_N \sqrt{\mathbf{P}_N}\|^2 &= \left(\sqrt{\sum_{i}^{N_N} (\mathbf{h}_{EN_p} \mathbf{U}_N \sqrt{\mathbf{P}_N})_i^2} \right)^2, \\ &= \sum_{i}^{N_N} (\mathbf{h}_{EN_p} \mathbf{U}_N \sqrt{\mathbf{P}_N})_i^2 = \sum_{i}^{N_N} (\mathbf{h}_{EN_p} \mathbf{U}_N)_i^2 p_{N,i}, \\ &= \mathrm{Tr}(|\mathrm{diag}(\mathbf{h}_{EN_p} \mathbf{U}_N)|^2 \mathbf{P}_N). \end{aligned} \tag{4.12}$$

By following similar steps above, $\|\mathbf{h}_{EF_p} \mathbf{U}_F \sqrt{\mathbf{P}_F}\|^2$ can be equivalently written as $\mathrm{Tr}(|\mathrm{diag}(\mathbf{h}_{EF_p} \mathbf{U}_F)|^2 \mathbf{P}_F)$. However, (4.9) is still not a concave problem due to the subtraction of concave functions. Hence, the first-order Taylor approximation is applied on the sum of R_{E_p} in order to convert (4.9) into subtraction of concave and linear functions. As a result, (4.9) becomes a concave optimization problem, as presented in following

$$\begin{aligned} \max_{\mathbf{P}_N, \mathbf{P}_F} \quad & [f(\mathbf{P}_N, \mathbf{P}_F) - \sum_{p=1}^{P} (g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)}) - \frac{\partial g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)})}{\partial \mathbf{P}_N} (\mathbf{P}_N - \mathbf{P}_N^{(K)}) \\ & - \frac{\partial g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)})}{\partial \mathbf{P}_F} (\mathbf{P}_F - \mathbf{P}_F^{(K)}))]^+ \end{aligned} \tag{4.13a}$$

$$\text{s.t.} \quad (4.9b), (4.9c), (4.9d), \text{ and } (4.9e). \tag{4.13b}$$

where $\mathbf{P}_F^{(K)}$ and $\mathbf{P}_N^{(K)}$ are the optimal points of $R_s^{(K)}$ at the K^{th} iteration.

$f(\mathbf{P}_N, \mathbf{P}_F)$ and $g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)})$ are given by (4.10a) and (4.10b), respectively. The partial gradients of g_p with respect to $\mathbf{P}_N$ and $\mathbf{P}_F$ are given by (4.14a) and (4.14b), respectively.

$$\begin{aligned}
&\frac{\partial g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)})}{\partial \mathbf{P}_N}(\mathbf{P}_N - \mathbf{P}_N^{(K)}) = \frac{1}{\ln(2)} \frac{\partial \left(\ln \left(1 + \frac{\|\mathbf{h}_{EN_p}\mathbf{U}_N\sqrt{\mathbf{P}_N}\|^2 + \|\mathbf{h}_{EF_p}\mathbf{U}_F\sqrt{\mathbf{P}_F}\|^2}{P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2} \right) \right)}{\partial \mathbf{P}_N}, \\
&= \frac{1}{\ln(2)} \frac{\partial \left(\ln \left(1 + \frac{\text{Tr}(|\text{diag}(\mathbf{h}_{EN_p}\mathbf{U}_N)|^2\mathbf{P}_N) + \text{Tr}(|\text{diag}(\mathbf{h}_{EF_p}\mathbf{U}_F)|^2\mathbf{P}_F)}{P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2} \right) \right)}{\partial \mathbf{P}_N}, \\
&= \frac{\frac{1}{\ln(2)} \text{Tr}(|\text{diag}(\mathbf{h}_{EN_p}\mathbf{U}_N)|^2 (P_{N,j} - P_N^{(K)}))}{\text{Tr}(|\text{diag}(\mathbf{h}_{EN_p}\mathbf{U}_N)|^2\mathbf{P}_N^{(K)}) + \text{Tr}(|\text{diag}(\mathbf{h}_{EF_p}\mathbf{U}_F)|^2\mathbf{P}_F^{(K)}) + P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2},
\end{aligned} \tag{4.14a}$$

$$\begin{aligned}
&\frac{\partial g_p(\mathbf{P}_N^{(K)}, \mathbf{P}_F^{(K)})}{\partial \mathbf{P}_F}(\mathbf{P}_F - \mathbf{P}_F^{(K)}) = \\
&\frac{\frac{1}{\ln(2)} \partial \left(\ln \left(1 + \frac{\text{Tr}(|\text{diag}(\mathbf{h}_{EN_p}\mathbf{U}_N)|^2\mathbf{P}_N) + \text{Tr}(|\text{diag}(\mathbf{h}_{EF_p}\mathbf{U}_F)|^2\mathbf{P}_F)}{P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2} \right) \right)}{\partial \mathbf{P}_F}, \\
&= \frac{\frac{1}{\ln(2)} \text{Tr}(|\text{diag}(\mathbf{h}_{EF_p}\mathbf{U}_F)|^2 (P_{F,j} - P_F^{(K)}))}{\text{Tr}(|\text{diag}(\mathbf{h}_{EN_p}\mathbf{U}_N)|^2\mathbf{P}_N^{(K)}) + \text{Tr}(|\text{diag}(\mathbf{h}_{EF_p}\mathbf{U}_N)|^2\mathbf{P}_F^{(K)}) + P_{jam}\|\mathbf{h}_{EJ_p}\|^2 + \sigma_{E_P}^2}.
\end{aligned} \tag{4.14b}$$

To solve $(P2)$, a successive convex approximation (SCA) programming-based iterative algorithm is presented in Algorithm 2.

Algorithm 2 GSVD-based SCA algorithm for solving $(P2)$

1: Initialize $\mathbf{P}_N^{(0)}$ and $\mathbf{P}_F^{(0)}$, set $K = 0$ and ϵ
2: **while** $|R_s^{(K+1)} - R_s^{(K)}| \leq \epsilon$ **do**
3: Update $\mathbf{P}_N^{(K)}$ and $\mathbf{P}_F^{(K)}$ by solving $(P2)$
4: Update $K = K + 1$
5: **end while**

Accordingly, $\mathbf{P}_N^{(0)}$ and $\mathbf{P}_F^{(0)}$ are equally power allocated at the beginning. ϵ is the stopping criteria. In each iteration, power allocation coefficients converge to a suboptimal solution.

4.3.1 Convergence Discussion

R_{E_p} is linearized at the K^{th} iteration of the first-order Taylor approximation. Therefore, (4.9a) can be defined as the subtraction of a linear function from a concave function, which is a concave function. As the $P_N^{(K)}$ and $P_F^{(K)}$ are at initial points, the solution can

be derived by solving $(P2)$ at the $(K+1)^{th}$ iteration if $|R_s^{(K+1)} - R_s^{(K)}| \leq \epsilon$ is satisfied. The first-order Taylor approximation-based method is a non-decreasing expression [66]. Therefore, the optimal solution is bounded by the error tolerance. This guarantees the convergence of the proposed solution.

4.3.2 Complexity Analysis

Algorithm 2 runs the GSVD-based precoding within the first-order Taylor approximation-based loop. The complexity of the GSVD based precoding is $\mathcal{O}\left(\frac{L^3}{\sigma} + \frac{L}{\sigma}\log(\frac{1}{\epsilon})\right)$ [84], where L is found by $\min(N_N + N_F, N_B)$ and σ is the search step. In the outer loop, K is the number of iterations, and ϵ is the error tolerance. Using the first-order approximation, R_{E_p} is linearized, and the objective becomes a concave problem. Thus K has the complexity of $\mathcal{O}(\log(\frac{1}{\epsilon}))$. In addition, the first order derivative of R_{E_p} is taken with respect to $\mathbf{P}_N$ and $\mathbf{P}_F$. The complexity of these two processes is given by $\mathcal{O}(\max\{N_N, N_F\})$. Finally, the time complexity of the proposed algorithm can be expressed by

$$\mathcal{O}(\frac{L^3}{\sigma} + \frac{L}{\sigma}\log(\frac{1}{\epsilon})) \times \log(\frac{1}{\epsilon})\max\{N_N, N_F\}).$$

In addition, the proposed method is applicable to the MIMO-OMA scheme, where the complexity of the near and the far users becomes $\mathcal{O}((\frac{L^3}{\sigma} + \frac{L}{\sigma}\log(\frac{1}{\epsilon})) \times \log(\frac{1}{\epsilon})N_N)$ and $\mathcal{O}((\frac{L^3}{\sigma}+\frac{L}{\sigma})\log(\frac{1}{\epsilon}))\times\log(\frac{1}{\epsilon})N_F)$. The proposed method can be extended to the multi-user scenario using a user pairing algorithms, i.e., near-far user pairing with a linear complexity.

4.3.3 Extension to the Case of More Than Two Users

Even though this chapter investigates the SSR performance of the two-user GSVD-based MIMO-NOMA system, this work can be easily extended to multi-user scenarios by exploiting hybrid multiple access methods. This makes the design compatible with the 3GPP specifications. The idea behind the hybrid multiple access system is that the BS assigns a maximum of two users to one group as standardized in the 3GPP-LTE system by using user pairing methods [85]. The pairs can be separated by orthogonal subcarriers or time slots, while the users in the same pair can use the proposed NOMA scheme.

Therefore, the hybrid multiple access technique improves the SSR performance of the current OMA system with existing fundamental resource blocks [86].

4.3.4 A Cost Comparison between MIMO-NOMA and MIMO-OMA

When a cost comparison is made between MIMO-OMA and MIMO-NOMA systems, there are four important factors: spectral efficiency, power consumption, computational complexity, and equipment cost. MIMO-NOMA systems are proven to be more spectral efficient in [62], [63]. [87] reveals that multi-user NOMA achieves better joint energy and spectral efficiency than OMA systems. Regarding the equipment cost, the SIC receiver has similar cost and hardware complexity compared with the traditional multi-user receivers [88]. These results support the idea that MIMO-NOMA-based systems are more cost-effective than MIMO-OMA systems.

4.4 Numerical Results

In this section, the performance of SSR achieved by MIMO-NOMA with a friendly jammer is compared with some benchmark results. The power of the noise at the i^{th} sub-channel of the BS (denoted by $\sigma^2_{B,i}$) and the p^{th} eavesdropper (denoted by $\sigma^2_{E_P}$) can be calculated as $\sigma^2 = N_0 \times B$. The simulation parameters are listed in Table 4.1.

Table 4.1. Simulation parameters

Parameter	Value
Noise spectral efficiency, N_0	-174 dBm/Hz
Bandwidth, B	10 MHz
Path loss component, α	3.2
Error tolerance, ϵ	10e-3
The distance between the equipment	80 m
The number of legitimate users	2
The number of eavesdroppers, P	2
QoS requirements, Q_N and Q_F	2 BPCU

In NOMA, sharing CSI information between nodes causes longer signaling overhead compared to OMA; however, the performance degradation of NOMA can be compensated by superior spectral efficiency and higher secrecy rate. Furthermore, the number of assigned users in a group is limited to two to cope with the signaling overhead and high receiver complexity problems.

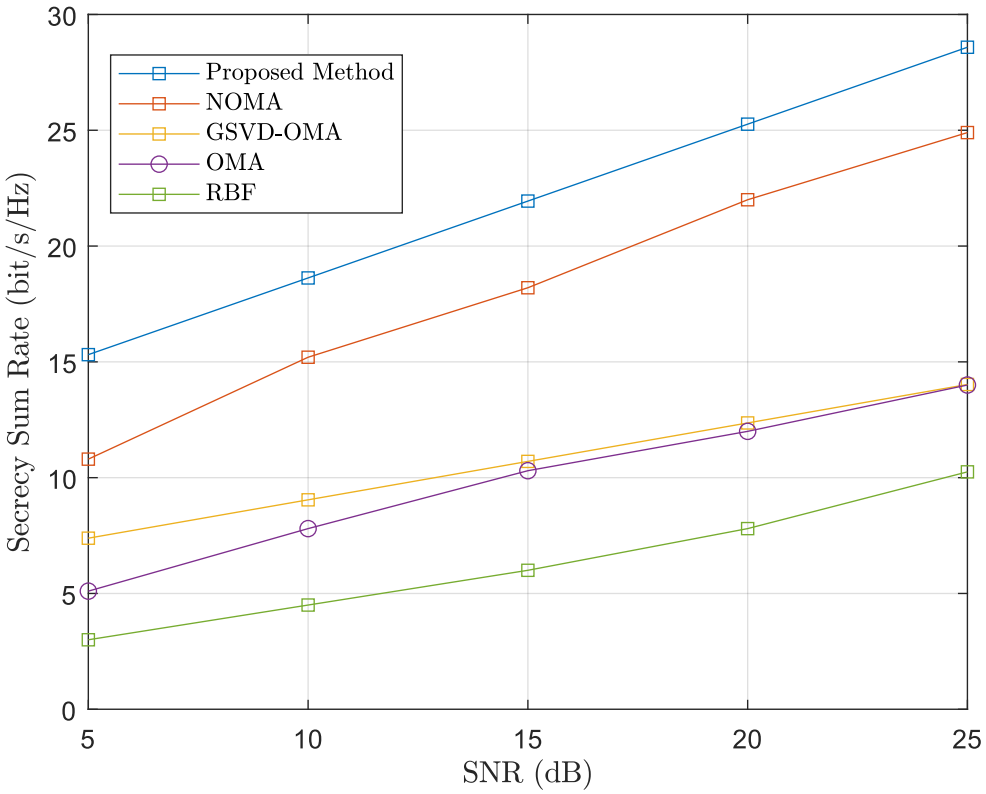

Figure 4.2. SSR comparison in MIMO systems

In Fig. 4.2, the SSR comparison of the proposed method (GSVD-NOMA) with an MMSE-based iterative NOMA, conventional TDMA (OMA), and random beamforming (RBF) methods in [69] are shown in a MIMO scenario ($N_B = 6,\ N_N = 2,\ N_F = 2$), where P_J is 20 dBmW. The covariances of H_{BF} and H_{BN} are ten times higher than the variance of the h_{EB} and h_{EN}. To make a fair comparison between the systems, we randomly deploy two eavesdroppers instead of an eavesdropper with two antennas because the capacity of MIMO systems is proportional to the multiplication of the receiver and the transmitter antennas, while the capacity of the MISO system is proportional with the number of transmitter antennas. Also, the nodes are placed equally distant from the BS, i.e., 80 m. Hence, $\mathbf{h}_{BF}$ has similar large-scale fading with $\mathbf{h}_{BN}$, which is the lower bound for the proposed algorithm. The figure shows that the proposed method has better SSR than

other methods. It can also be seen that the proposed algorithm can be applied to OMA systems, and it (denoted by "GSVD-OMA") archives better SSR compared to conventional OMA systems. Another important insight from Fig. 4.2 is that the performance gap between NOMA and OMA schemes is increased with the increasing SNR. This is because NOMA-MIMO has better channel capacity compared with OMA.

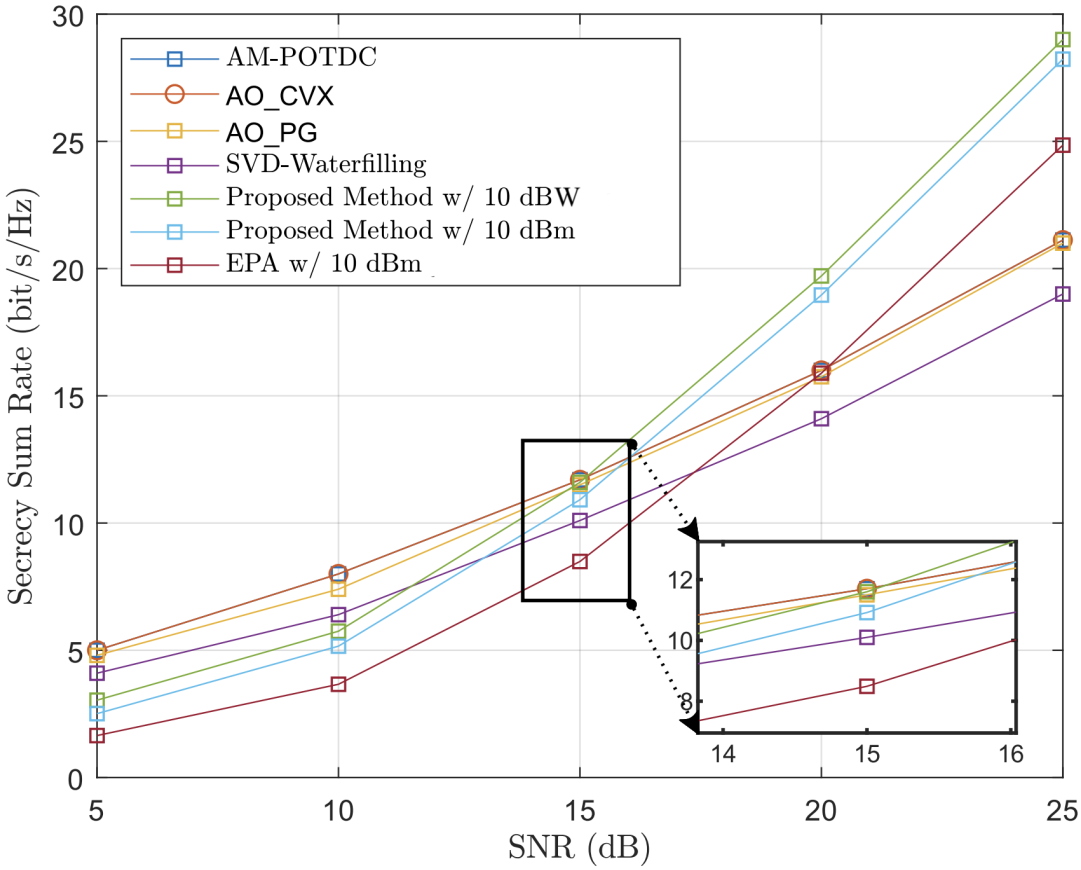

Figure 4.3. SSR comparison NOMA with OMA

In Fig. 4.3[3], the SSR performance of the proposed method is compared with various schemes in [75], namely alternating optimization-based (AM-POTDC), AO-CVX, AO-PG algorithms, the singular value decomposition (SVD) based waterfilling algorithm, and the equal power allocated NOMA. The main difference of this setup ($N_B = 6,\ N_N = 6,\ N_F = 6,\ P = 3$) from the previous figure is the number of eavesdroppers. It can be found that the jammer has increased the SSR performance of the proposed method. However, it is clear that in low SNR, the NOMA system is more vulnerable to wiretapping than the OMA-based system. This is because NOMA suffers from co-channel interference along with multiple eavesdroppers. However, NOMA becomes more advantageous with high SNR as the legitimate users share the whole bandwidth. Therefore,

[3]It should be noted that in Figures 3.3 and 3.4, with a data rate of 1 Gbit for N, the corresponding delay is notably elevated.

the most striking conclusion from this figure is that OMA or NOMA schemes are suitable for different SNRs. To maximize the SSR over SNR, a hybrid system should be considered. Please note that the reason why the 1000-fold increase does not affect the performance of the proposed method is the following. The reason the enhancement in the secrecy sum rate was not substantial, despite a thousandfold increase in jamming power, can be attributed to the base station's utilization of the GSVD technique. This technique facilitates the targeted transmission of downlink signals to authorized users, while the jammer employs omnidirectional jamming except towards the base station. This is achieved through null spacing beamforming. Consequently, the power of the jamming signal does not notably impact the received power from the base station.

Fig. 4.4 compares the SSR performance of NOMA with OMA with two different QoS requirements. It can be seen from the figure that OMA users need less power to start secure transmission compared to NOMA. This is because the far user in NOMA suffers from inter-user interference severely, especially in low SNR. Moreover, as the Q_N constraint increases, the far user must utilize more transmit power to meet the QoS criteria. Hence, the interference becomes more severe. As a result, if any of the NOMA users can not satisfy the QoS constraints, the SSR becomes zero as there is no transmission. On the other hand, once QoS requirements are met, NOMA provides better SSR performance than OMA due to higher spectral efficiency.

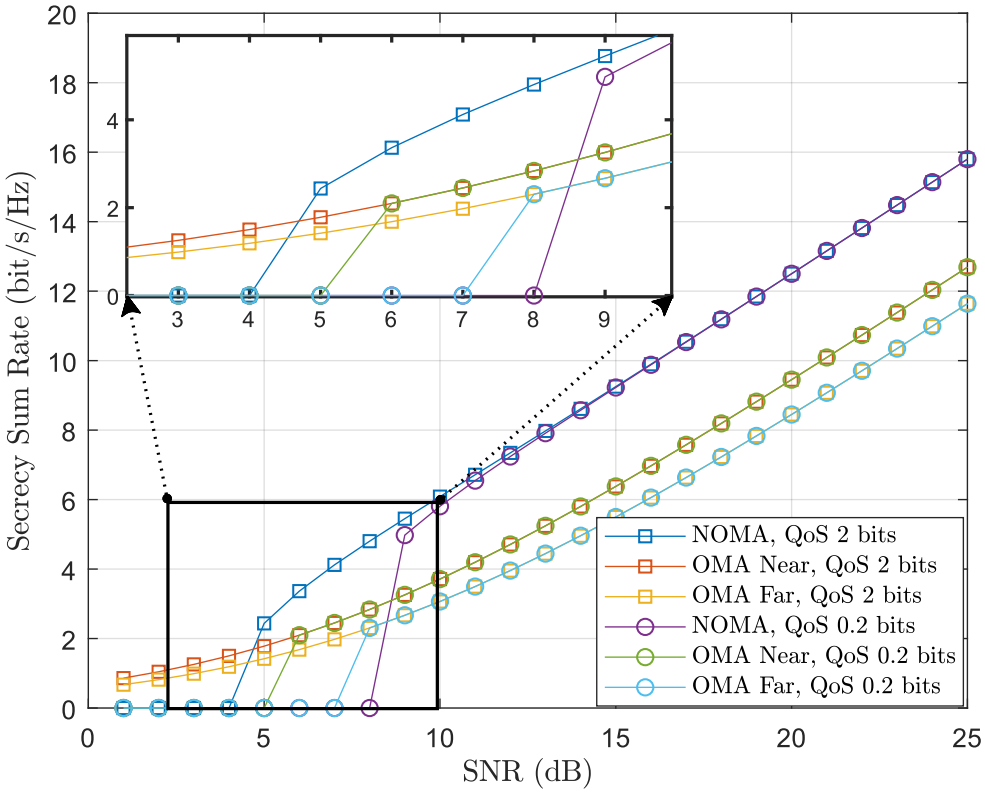

Figure 4.4. The effect of the QoS on the SSR in NOMA and OMA systems

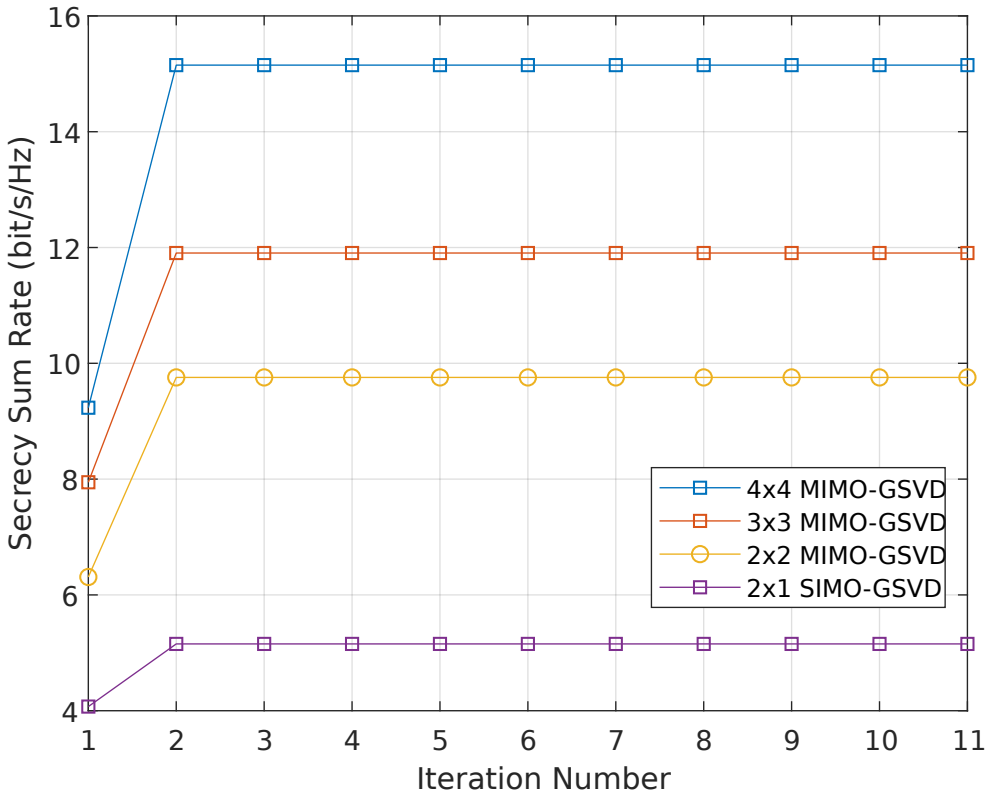

Figure 4.5. The SSR versus the iteration number

Fig. 4.5 presents the convergence performance of the proposed algorithm. This is be-

cause, SCA based algorithm has linearized the sum of R_{E_p} at the point of $(\mathbf{P_N^K}, \mathbf{P_F^K})$. Therefore, the objective function becomes a convex optimization problem. Hence, the problem can be solved in two iterations. The figure also proves the convergence of the proposed method.

4.5 Conclusion of Chapter 4

Security is one of the fundamental aspects of wireless networks. In this chapter, the sum secrecy rates of a NOMA-MIMO uplink network have been maximized under total transmit power and QoS constraints in different scenarios. The SCA-based algorithm has been exploited as the objective function was not convex. Numerical results showed that NOMA had achieved better SSR than some benchmarks, especially in high SNR. Further research might explore eliminating noise floor in NOMA to increase SSR. Also, combining OMA and NOMA, a hybrid multiple access system could be a good research direction to enhance SSR performance.

Chapter 5

Green NOMA-Based MU-MIMO Transmission for MEC in 6G Networks

5.1 Introduction

Chapter 3 and Chapter 4 focused on the optimization of non-orthogonal multiple access (NOMA)-based multiple-input multiple-output (MIMO) networks, primarily aiming to maximize data rate while considering specific objectives such as secrecy rate or delay. In this chapter, our investigation shifts towards examining the energy consumption of NOMA-based MIMO-NOMA networks.

NOMA is a promising candidate for 6G networks [25]. Unlike the traditional orthogonal multiple access (OMA) schemes, NOMA enables multiple users to be served at the same communication resource (i.e., time, frequency, or code) [89]. As a result of this approach, NOMA was proven to have higher spectral efficiency compared to orthogonal frequency-division multiple access (OFDMA) [90]. In recent years, there has been an increasing interest in the power-domain NOMA, which superposes the transmitted signals over the power domain and applies the successive interference cancellation (SIC) technique to decode the received signals [90]. MIMO is also one of the critical technologies for next-generation multiple access schemes by exploiting multiplexing gain or spatial diversity. Comparing channel matrices in MIMO-NOMA is not as straightforward as in SISO-NOMA [86]. To circumvent this issue, [86] proposed using the generalized singular value decomposition (GSVD) technique to decompose two MIMO channels into parallel SISO channels; consequently, GSVD enables the users to combine the benefits

of NOMA with those of MIMO effectively [27].

Computational-intensive, delay-sensitive, and energy-hungry applications, such as virtual/augmented reality and online gaming, require telecommunication operators to bring cloud computing-like facilities to the edge of the networks. Mobile edge computers (MEC) are small-size cloud computers that can meet different needs such as low delay, high privacy, and context-awareness in mobile networks [91]. Most research on NOMA-MEC has been carried out using single-antenna configuration [92]-[93]. In [92], energy efficiency for a delay-constraint massive Internet of Things (IoT) network was maximized by jointly optimizing sub-channels, transmission power, and sub-carriers. In [94], different from [92], an energy minimization problem for multiple single-antenna users was jointly optimized by considering the binary offloading model. The common assumption of these two studies was that the users have the same data length; however, in a massive-user scenario, the same length causes decoding order inconsistency and co-channel interference. To solve this problem, users' average delay was minimized by optimizing offloading task size and computation resources in [95]. In [96], different from the previous studies, a multi-user multi-base station scenario was investigated to minimize energy consumption. The authors optimized this scenario's binary task assignment, power allocation, and user assignment variables. In [93], a wireless-powered NOMA-MEC system was introduced. In the two-user system model, the users harvest the ambient energy and offload their data to the MEC server. The time slots for wireless power transfer (WPT), power allocation variables, and beamforming vectors were optimized [93].

Besides the SISO-NOMA-based studies above, OMA-based approaches can be applied to NOMA systems for energy minimization. The authors of [97] proposed a slotted ALOHA technique for wireless-powered IoT devices accessing the MEC server. This work optimized a time interval for energy harvesting and data offloading. On the other hand, co-channel interference is one of the problems of the NOMA system. This technique can reduce co-channel interference and save energy by allowing the paired NOMA users to access the MEC at a slotted time. In [98], the authors proposed a Lyapunov stochastic optimization-based solution for energy minimization of MEC under the accuracy and latency constraints. Unlike [98], to minimize long-term energy consumption, modulation, and coding schemes, power and CPU frequencies of mobile devices and a base station is

optimized in a discontinuous method [99].

It is more realistic to assume that the users have multiple antennas rather than one. Nevertheless, a relatively small body of literature is concerned with the combination of the NOMA, MIMO, and MEC technologies [25], [100]. In [25], the total offloading delay was minimized for a hybrid NOMA-based MIMO-MEC system under power budget and quality of services constraints. In [100], a millimeter wave (mmWave) MIMO-based MEC offloading scenario was studied to maximize the weighted sum rate. In this work, multi-antenna users were associated with a base station, and optimal precoding design and power allocation were proposed.

Research has not yet investigated the combination of multiple-antenna-equipped NOMA users and MEC technologies for the total energy minimization problem. For this reason, the motivation of this chapter is to bring MEC's computation, MIMO's communication, and NOMA's connectivity capabilities together. To this end, we formulate the total energy minimization problem for a GSVD-based NOMA-MIMO-MEC and provide a low complexity algorithm to solve the formulated problem. The main contributions of this chapter can be listed as follows:

- We consider multiple antenna-equipped two NOMA users offloading their data to the MEC server by cooperatively sharing the same communication resource to reduce total energy consumption. The energy minimization problem is formulated by considering offloading time limitations, power allocation, task assignment, and losses in RF chains.
- To combine NOMA, MIMO, and MEC technologies, we exploit the GSVD linear beamforming method. As the formulated problem is non-convex, we used the alternating optimization method by dis-jointly optimizing the task assignment coefficients and the power allocation vectors. The task assignment problem was solved in the proposed solution on the outer layer. The successive convex approximation method was applied in the inner layer to convert the non-convex power allocation problem into a first-order linear form. And the power allocation vectors were obtained.
- Numerical results are provided to analyze the impact of offloading time, power

budget, users' locations, and data rate on energy consumption. We use OMA as a benchmark scheme. According to the findings, NOMA based MIMO-NOMA network saves a significant amount of energy compared to OMA, especially at a high SINR rate with multiple antennas. Also, the findings reveal that optimal user pairing is crucial to reducing energy consumption for NOMA.

5.1.1 Network Model

We consider a two-user NOMA-based MIMO-MEC offloading scenario where the base station has M antennas, and the users (UE_s) are equipped with K antennas as illustrated in Fig. 5.1. The users can simultaneously offload their time-limited, CPU-intensive, or memory-demanding tasks to the base station.

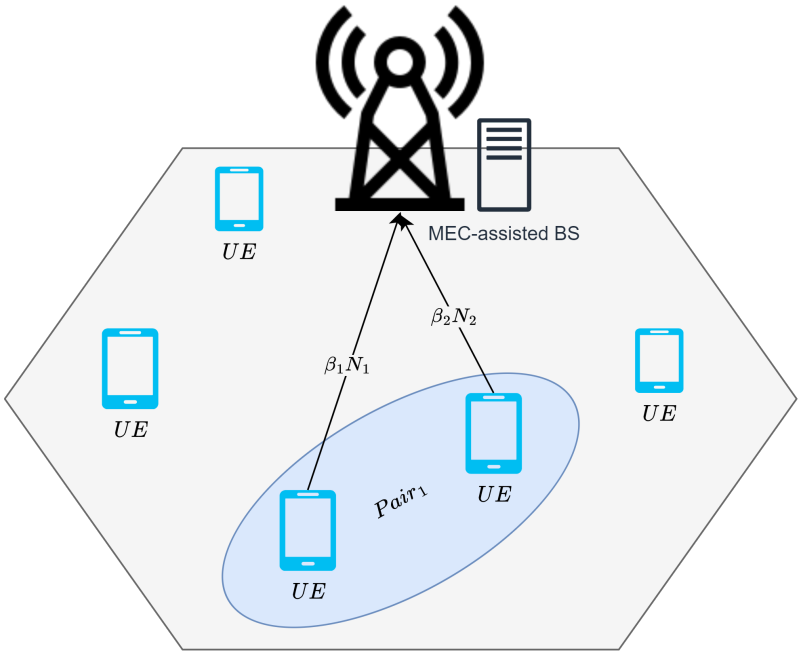

Figure 5.1. NOMA-based MIMO-MEC network

5.1.2 Channel Model

The wireless channel between the i^{th} user (UE_i) and the base station is formulated as a quasi-static Rayleigh fading channel, i.e., $\mathbf{H}_i \in \mathbb{C}^{M \times K} \sim \mathcal{CN}(0, 1)$ with a large scale fading $d^{-\alpha}$, where d is the distance between the user and the base station and α is the path-loss component. The users and the base station are assumed to know full channel state information (CSI).

5.1.3 Application of the GSVD Technique on MIMO Systems

GSVD is a linear precoding and decoding scheme widely proposed in NOMA-based MIMO systems [86], [27]. The GSVD technique converts two MIMO channels into virtual multiple orthogonal SISO channels so that GSVD simplifies communication resource allocation problems in MIMO systems. Even though the GSVD cannot provide optimal performance, it was proven to be highly efficient and reduce the complexity of MIMO systems [86]. $\mathbf{H}_i \in \mathbb{C}^{M \times K}$ denotes a complex Gaussian channel matrix between the i^{th} user and the base station. Two MIMO channels can be decomposed as follows [86]:

$$\mathbf{H}_i = \mathbf{U}\mathbf{\Lambda}_i\mathbf{V}_i^H,\ i \in \{1, 2\}, \tag{5.1}$$

where $\mathbf{U}$ is an $M \times M$ square invertible matrix, $\mathbf{V}_i^H$ is an $M \times M$ unitary matrix and $\mathbf{\Lambda}_i$ is a diagonal matrix. The diagonal elements of the $\mathbf{\Lambda}_i$ are made up from the singular values, i.e., $\mathbf{\Lambda}_i = \{\sigma_{i,1}...\sigma_{i,j}\}$ where j is the number of orthogonal subchannels. It should be noted that $\sigma_{1,j}$ is in increasing order, and $\sigma_{2,j}$ is in decreasing order. In this way, pairing $\sigma_{1,j}$ with $\sigma_{2,j}$ is ideal for combining MIMO with NOMA [86]. Therefore, $\mathbf{V}_i$ and $\mathbf{U}^{-1}$ become the precoding and the detection matrices to diagonalize $\mathbf{H}_i$.

5.1.4 Multiple Access Model

The key idea of the power domain NOMA is to allow multiple UEs to share the same orthogonal resource blocks such as time, frequency, or codeword by exploiting the power domain. The received signal at the base station is

$$\mathbf{y} = \sum_{i=1}^{2} \mathbf{H}_i\mathbf{x}_i + \mathbf{n}, \tag{5.2}$$

where $\mathbf{x}_i$ is the information vector of the i^{th} user. For the uplink transmission, the application of GSVD to NOMA can be given by

$$\mathbf{y} = \sum_{i=1}^{2}\sum_{j=1}^{L} \sigma_{i,j}\sqrt{P_{i,j}}x_{i,j} + \hat{n}_j, \tag{5.3}$$

where L is the number of orthogonal sub-channels, $P_{i,j}$ denotes the power allocation coefficient for the i^{th} user's j^{th} subchannel. $\mathbf{x}_i$ represents the i^{th} user's information bearing signal. $\hat{n}_j$ is complex additive white Gaussian noise at the j^{th} orthogonal channel. The achievable maximum data rates for UE_1 and UE_2 can be given as follows:

$$R_1 = B\sum_{j=1}^{L}\log_2\left(1+\frac{\sigma_{1,j}^2 P_{1,j}}{\sigma_{2,j}^2 P_{2,j}+\hat{n}_j^2}\right),\ bits/s/Hz, \tag{5.4a}$$

$$R_2 = B\sum_{j=1}^{L}\log_2\left(1+\frac{\sigma_{2,j}^2 P_{2,j}}{\hat{n}_j^2}\right),\ bits/s/Hz, \tag{5.4b}$$

where B denotes the transmission bandwidth for NOMA users.

5.1.5 Mobile Edge Computing

This chapter considers partial task offloading. Accordingly, UEs can compute their tasks locally, partially or completely remotely. Note that we omit downloading the results because the task might be just storing data to the MEC or the result consists of the insignificant size of bits [96].

Local Computing

Local computing is to execute tasks locally on the UE's processor. The variables β_1 and β_2 are introduced as the task assignment coefficients for UE_1 and UE_2 such that $(1-\beta_1)N_1$ and $(1-\beta_2)N_2$ bits are assigned to UE_1 and UE_2 for local computing, respectively. The local computing times for these processes can be found by

$$T_1^{loc} = \frac{(1-\beta_1)N_1 C_{l1}}{f_{l1}} \text{ and } T_2^{loc} = \frac{(1-\beta_2)N_2 C_{l2}}{f_{l2}}, \tag{5.5}$$

where f_{l1} and f_{l2} are the CPU frequencies and C_{l1} and C_{l2} are the required CPU cycles to execute one bit for UE_1 and UE_2, respectively. The energy consumption for UE_1 and UE_2 for local computing are [96]

$$E_1^{loc} = \zeta_{l1}(1-\beta_1)N_1 C_{l1} f_{l1}^2 \text{ and } E_2^{loc} = \zeta_{l2}(1-\beta_2)N_2 C_{l2} f_{l2}^2, \tag{5.6}$$

where ζ_{l1} and ζ_{l2} are the energy consumption coefficients for UE_1 and UE_2.

Energy Consumption in RF Chains

There is no single formula for calculating the optimal power consumption of an antenna system. It is equal to the superposition of different components, i.e., amplifier, filter, circuitry, cooling, and operation parameters (frequency, bandwidth, modulation technique, etc.). Therefore, these parameters should be optimized separately. Increasing the array size to a certain point can allow more data to be transmitted simultaneously. Yet, as the system grows in size, the high power consumption and complexity of the RF chains may offset the benefits of MIMO. To compare the MIMO systems fairly, the energy consumed on each antenna element must be considered [101]. [102] shows that the power an antenna consumes is proportional to the number of optimized antenna sizes. This is because each antenna element requires its power amplifier and other components. The radiated power P_{rad} can be written as [103]:

$$P_{rad} = \nu(P_{bud} - 2K \times P_{RFc}^{T} - M \times P_{RFc}^{R}), \tag{5.7}$$

where ν is the efficiency of the power amplifier and P_{bud} is the system's power budget. K and M are the number of antennas, and P_{RFc}^{T} and P_{RFc}^{R} are the power consumed in an RF chain at the transmitter and receiver. We make the assumption, without loss of generality, that there are two identical transmitters, and the value "2" in Eq. (5.7) corresponds to these transmitters.

Task Offloading

There are $\beta_1 N_1$ and $\beta_2 N_2$ bits that need to be offloaded to the MEC for remote execution of UE_1 and UE_2's tasks, respectively. The offloading times for UE_1 and UE_2 are given by

$$T_1^{off} = \frac{\beta_1 N_1}{R_1} \text{ and } T_2^{off} = \frac{\beta_2 N_2}{R_2}. \tag{5.8}$$

The corresponding total energy consumption for UE_1 and UE_2 is [103]

$$E_1^{off} = \frac{\beta_1 N_1}{R_1}\left(\sum_{j=1}^{L} P_{1,j} + KP_{RFc}^{T} + MP_{RFc}^{R}\right) \tag{5.9}$$

$$E_2^{off} = \frac{\beta_2 N_2}{R_2}\left(\sum_{j=1}^{L} P_{2,j} + KP_{RFc}^{T} + MP_{RFc}^{R}\right). \tag{5.10}$$

MEC Computing

After the tasks are offloaded, remote computation at the MEC can start. The computation times at the MEC server for UE_1 and UE_2 are

$$T_1^{mec} = \frac{\beta_1 N_1 C_m}{f_m} \text{ and } T_2^{mec} = \frac{\beta_2 N_2 C_m}{f_m}, \tag{5.11}$$

where f_m is the CPU frequency and C_m is the required CPU cycles to execute one bit at the MEC. The energy consumption for remote execution at the MEC server for UE_1 and UE_2, respectively, can be given by [96]

$$E_1^{mec} = \zeta_m \beta_1 N_1 C_m f_m^2 \text{ and } E_2^{mec} = \zeta_m \beta_2 N_2 C_m f_m^2, \tag{5.12}$$

where ζ_m is the energy consumption coefficient for the MEC.

5.2 Problem Definition and Solution

The total energy minimization problem for task offloading, MEC computing, and local computing with time constraints is formulated as follows:

$$\begin{aligned}
(P1) \quad \min_{\beta_1,\beta_2,P_{1,j},P_{2,j}} \quad & T_1^{off}\left(\sum_{j=1}^{L} P_{1,j} + KP_{RFc}^{T} + MP_{RFc}^{R}\right) \\
& + T_2^{off}\left(\sum_{j=1}^{L} P_{2,j} + KP_{RFc}^{T} + MP_{RFc}^{R}\right) \\
& + \zeta_m \beta_1 N_1 C_m f_m^2 + \zeta_m \beta_2 N_2 C_m f_m^2 \\
& + \zeta_{l1}(1-\beta_1) N_1 C_{l1} f_{l1}^2 + \zeta_{l2}(1-\beta_2) N_2 C_{l2} f_{l2}^2 && \text{(5.13a)} \\
\text{s.t.} \quad & T_1^{off} = \frac{\beta_1 N_1}{R_1} \leq T_{max}^{off} && \text{(5.13b)} \\
& T_2^{off} = \frac{\beta_2 N_2}{R_2} \leq T_{max}^{off} && \text{(5.13c)} \\
& T_1^{loc} = \frac{(1-\beta_1) N_1 C_{l1}}{f_{l1}} \leq 2T_{max}^{off} && \text{(5.13d)}
\end{aligned}$$

$$T_2^{loc} = \frac{(1-\beta_2)N_2C_{l2}}{f_{l2}} \leq 2T_{max}^{off} \tag{5.13e}$$

$$T_1^{mec} = \frac{\beta_1 N_1 C_m}{f_m} \leq T_{max}^{off} \tag{5.13f}$$

$$T_2^{mec} = \frac{\beta_2 N_2 C_m}{f_m} \leq T_{max}^{off} \tag{5.13g}$$

$$0 \leq \sum_{j=1}^{L} P_{i,j} \leq P_{rad},\ i \in \{1,2\},\ j \in \{1,..,L\} \tag{5.13h}$$

$$P_{i,j} \geq 0 \tag{5.13i}$$

$$0 \leq \beta_i \leq 1, \tag{5.13j}$$

where (5.13b) and (5.13c) require that offloading time has to be less than or equal to an offloading time for the UE_1 and UE_2, respectively. (5.13d) and (5.13e) indicates that the deadline for local computing is less than or equal to the sum of offloading time and computing time at the MEC, i.e., $T_{max}^{loc} \leq 2T_{max}^{off}$. (5.13f) and (5.13g) are to avoid congestion at the MEC server, the deadline for the MEC computing is less than or equal to offloading time, i.e., $T_1^{mec} \leq T_1^{off}$ and $T_2^{mec} \leq T_2^{off}$ as illustrated in Fig. 5.2 [95]. (5.13h) makes sure that the users cannot exceed the power budget. (5.13i) indicates that the power allocation coefficients are non-negative. (5.13j) guarantees that β_i can get any value between zero and one.

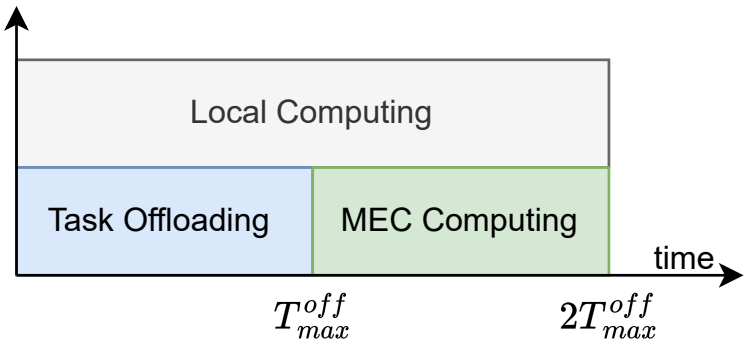

Figure 5.2. Execution times for local computing, task offloading, and MEC computing

$(P1)$ is not a convex optimization problem due to the fractional expressions; however, $(P1)$ can be written as an alternating optimization problem as in $(P2)$.

$$(P2) \quad \min_{\beta_1,\beta_2} g(\beta_1,\beta_2) \triangleq \min_{P_{1,j},P_{2,j}} E(P_{1,j},P_{2,j}) = \frac{\beta_1 N_1}{R_1}\left(\sum_{j=1}^{L} P_{1,j} + KP_{RFc}^T + MP_{RFc}^R\right)$$
$$+ \frac{\beta_2 N_2}{R_2}\left(\sum_{j=1}^{L} P_{2,j} + KP_{RFc}^T + MP_{RFc}^R\right) + \zeta_m \beta_1 N_1 C_m f_m^2 + \zeta_m \beta_2 N_2 C_m f_m^2$$

$$+ \zeta_{l1}(1-\beta_1)N_1 C_{l1} f_{l1}^2 + \zeta_{l2}(1-\beta_2)N_2 C_{l2} f_{l2}^2 \tag{5.14a}$$

$$\text{s.t.} \quad (5.13b), (5.13c), (5.13d), (5.13e), (5.13f), (5.13g), (5.13h) \text{ and } (5.13j), \tag{5.14b}$$

where $g(\beta_1, \beta_2)$ is the outer problem and $E(P_{1,j}, P_{2,j})$ is the inner problem. Therefore, β_1 and β_2 are fixed to deal with the inner problem $E(P_{1,j}, P_{2,j})$.After sub-optimal values for $E(P_{1,j}, P_{2,j})$ are found, we fix them to find out $g(\beta_1, \beta_2)$. Yet $(P2)$ is not a convex problem; the optimal solution can be found by exhaustive search. However, this is a computationally complex and time-consuming solution. Because the computational complexity of $(P2)$ grows exponentially with the number of antenna sizes and power levels. Therefore, the successive convex approximation method can be applied to $(P3)$ to solve $P_{1,j}$ and $P_{2,j}$ iteratively. The inner problem becomes

$$\begin{aligned}
(P3) \quad \min_{P_{1,j}, P_{2,j}} \quad & E(P_{1,j}, P_{2,j}) \approx E(P_{1,j}^K, P_{2,j}^K) + \frac{\partial E(P_{1,j}^K, P_{2,j}^K)}{\partial P_{1,j}}(P_{1,j} - P_{1,j}^K) \\
& + \frac{\partial E(P_{1,j}^K, P_{2,j}^K)}{\partial P_{2,j}}(P_{2,j} - P_{2,j}^K) && (5.15a) \\
\text{s.t.} \quad & \frac{\beta_1 N_1}{R_1} \leq T_{max}^{off} && (5.15b) \\
& \frac{\beta_2 N_2}{R_2} \leq T_{max}^{off} && (5.15c) \\
& 0 \leq \sum_{j=1}^{L} P_{i,j}, \ i \in \{1,2\}, \ j \in \{1,..,L\} && (5.15d) \\
& P_{i,j} \geq 0, && (5.15e)
\end{aligned}$$

where the partial derivative of $E(P_{1,j}^K, P_{2,j}^K)$ with respect to $P_{1,j}$ and $P_{2,j}$ can be written respectively as follows:

$$\frac{\partial E(P_{1,j}^K, P_{2,j}^K)}{\partial P_{1,j}} = \frac{\beta_1 N_1 \log(2)}{B} \times \left[\frac{\sum_{j=1}^{L} \ln(1 + \frac{\sigma_{1,j}^2 P_{1,j}^K}{\sigma_{2,j}^2 P_{2,j}^K + n_j}) - \frac{\sigma_{1,j}^2 \left(\sum_{j=1}^{L} P_{1,j} + K P_{RFc}^T + M P_{RFc}^R \right)}{\sigma_{1,j}^2 P_{1,j}^K + \sigma_{2,j}^2 P_{2,j}^K + n_j}}{\left[\sum_{j=1}^{L} \ln\left(1 + \frac{\sigma_{1,j}^2 P_{1,j}^K}{\sigma_{2,j}^2 P_{2,j}^K + n_j}\right) \right]^2} \right], \tag{5.16a}$$

$$\frac{\partial E(P_{1,j}^K, P_{2,j}^K)}{\partial P_{2,j}} = \frac{\beta_1 N_1 \log(2)}{B} \times \frac{\left(\sum_{j=1}^{L} P_{1,j} + K P_{RFc}^T + M P_{RFc}^R \right) P_{1,j}^K \sigma_{1,j}^2 \sigma_{2,j}^2}{(\sigma_{2,j}^2 P_{2,j}^K + n_j)(\sigma_{2,j}^2 P_{2,j}^K + \sigma_{1,j}^2 P_{1,j}^K + n_j)}$$
$$\times \left[\ln\left(1 + \frac{\sigma_{1,j}^2 P_{1,j}^K}{\sigma_{2,j}^2 P_{2,j}^K + n_j}\right) \right]^{-2}$$

$$+\frac{\beta_2 N_2 \log(2)}{B}\left[\frac{\sum_{j=1}^{L}\ln(1+\frac{\sigma_{2,j}^2 P_{2,j}^K}{n_j})-\frac{\sigma_{2,j}^2}{\sigma_{2,j}^2 P_{2,j}^K+n_j}\left(\sum_{j=1}^{L}P_{2,j}+KP_{RFc}^T+MP_{RFc}^R\right)}{\left[\sum_{j=1}^{L}\ln\left(1+\frac{\sigma_{2,j}^2 P_{2,j}^K}{n_j}\right)\right]^2}\right]. \tag{5.16b}$$

Proposition 1. *After finding sub-optimal $P_{1,j}^*$ and $P_{2,j}^*$, the outer problem, $g(\beta_1,\beta_2)$, becomes a linear optimization problem.*

$$g(\beta_1,\beta_2)=T_1^{off}\left(\sum_{j=1}^{L}P_{1,j}^*+KP_{RFc}^T+MP_{RFc}^R\right)+T_2^{off}\left(\sum_{j=1}^{L}P_{2,j}^*+KP_{RFc}^T+MP_{RFc}^R\right)$$
$$+\zeta_m\beta_1 N_1 C_m f_m^2+\zeta_m\beta_2 N_2 C_m f_m^2+\zeta_{l1}(1-\beta_1)N_1 C_{l1} f_{l1}^2+\zeta_{l2}(1-\beta_2)N_2 C_{l2} f_{l2}^2. \tag{5.17}$$

Finally, the problem can be rewritten as follows:

$$(P4)\quad \min_{\beta_1,\beta_2}\quad T_1^{off}\left(\sum_{j=1}^{L}P_{1,j}^*+KP_{RFc}^T+MP_{RFc}^R\right)+T_2^{off}\left(\sum_{j=1}^{L}P_{2,j}^*+KP_{RFc}^T+MP_{RFc}^R\right)$$
$$+\zeta_m\beta_1 N_1 C_m f_m^2+\zeta_m\beta_2 N_2 C_m f_m^2+\zeta_{l1}(1-\beta_1)N_1 C_{l1} f_{l1}^2+\zeta_{l2}(1-\beta_2)N_2 C_{l2} f_{l2}^2 \tag{5.18a}$$

$$T_1^{loc}=\frac{(1-\beta_1)N_1 C_{l1}}{f_{l1}}\le 2T_{max}^{off} \tag{5.18b}$$

$$T_2^{loc}=\frac{(1-\beta_2)N_2 C_{l2}}{f_{l2}}\le 2T_{max}^{off} \tag{5.18c}$$

$$T_1^{mec}=\frac{\beta_1 N_1 C_m}{f_m}\le T_{max}^{off} \tag{5.18d}$$

$$T_2^{mec}=\frac{\beta_2 N_2 C_m}{f_m}\le T_{max}^{off} \tag{5.18e}$$

$$\beta_1 N_1\le R_1 T_{max}^{off} \tag{5.18f}$$

$$\beta_2 N_2\le R_2 T_{max}^{off} \tag{5.18g}$$

$$0\le\beta_i\le 1,\ i\in\{1,2\}. \tag{5.18h}$$

5.2.1 Complexity Analysis

Algorithm 3 runs the GSVD-based iterative method inside the inner and outer loops. As the task assignment problem $(P4)$ is formulated as a linear optimization problem, and

Algorithm 3 Alternating optimization (AO) based algorithm to find β_1, β_2, $P_{1,j}$, and $P_{2,j}$

```
Initialize β1 and β2
while Δg(β1, β2) → 0 do
    Initialize P1,j and P2,j
    while (ΔE(P1,j, P2,j) → 0) do
        Find out the optimal P1,j and P2,j by using (P3)
    end while
    Solve (P4) to find out the optimal β1 and β2
    Update β1 and β2
end while
```

the power allocation problem $(P3)$ is linearized in the inner loop; consequently, the algorithm converges in the second iteration. The GSVD based precoding has the complexity of $\mathcal{O}\left(\frac{L^3}{\sigma} + \frac{L}{\sigma}\log(1/\epsilon)\right)$, where L is the minimum number of antenna of the paired UE_s and σ is the search step [77]. Consequently, the complexity of the proposed algorithm is $\mathcal{O}\left(\frac{2L^3}{\sigma} + \frac{2L}{\sigma}\log(1/\epsilon)\right)$.

5.2.2 Extension to multi-user scenario

This chapter focuses on the two-user scenario. However, this scenario can be extended to a multi-user scenario by using hybrid multiple access protocol [86] as depicted in Figure 5.3. According to the hybrid multiple access method, the users are paired first. Then, the paired users are allocated to orthogonal resource blocks (e.g., time, frequency) as in OMA systems. User pairing is a key factor in NOMA as it increases spectral efficiency, optimizes resource usage, ensures fairness, manages interference, and facilitates massive connectivity. The users in each orthogonal block exploit the proposed scheme. Therefore, the proposed scheme can reduce the energy consumption of the existing OMA-based systems.

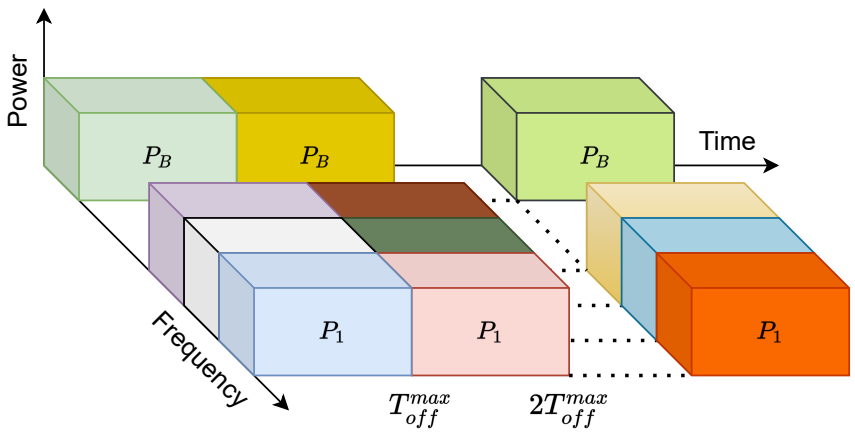

Figure 5.3. Extension from two-user to multi-user scenario

5.3 Numerical Results

This section provides numerical results to validate the superior performance of the NOMA-based MIMO-MEC compared with OMA. The simulation parameters are given in Table 5.1.

Table 5.1. Simulation parameters

Parameter	Value
Bandwidth, B	10 MHz
Cell radius	500 m
Noise Power, N_o	-94 dBm
Path loss component	3
Error tolerance, ϵ	10e-5
Energy consumption coefficient for MEC, ζ_m	10e-29
Energy consumption coefficient for UEs, ζ_{l1},ζ_{l2}	10e-27
MEC CPU frequency, f_m	2.5e10 cycles/sec
UEs CPU frequency, f_{l1} and f_{l2}	3.23e9 cycles/sec
CPU cycles per bit in MEC, C_m	500 cycles/bit
CPU cycles per bit in UEs, C_{l1} and C_{l2}	1000 cycles/bit
The efficiency of the power amplifier, ν	0.68
The power expenditure on the transmitter, P_{RFc}^{T}	24.1 mW
The power expenditure on the receiver, P_{RFc}^{R}	31.25 mW

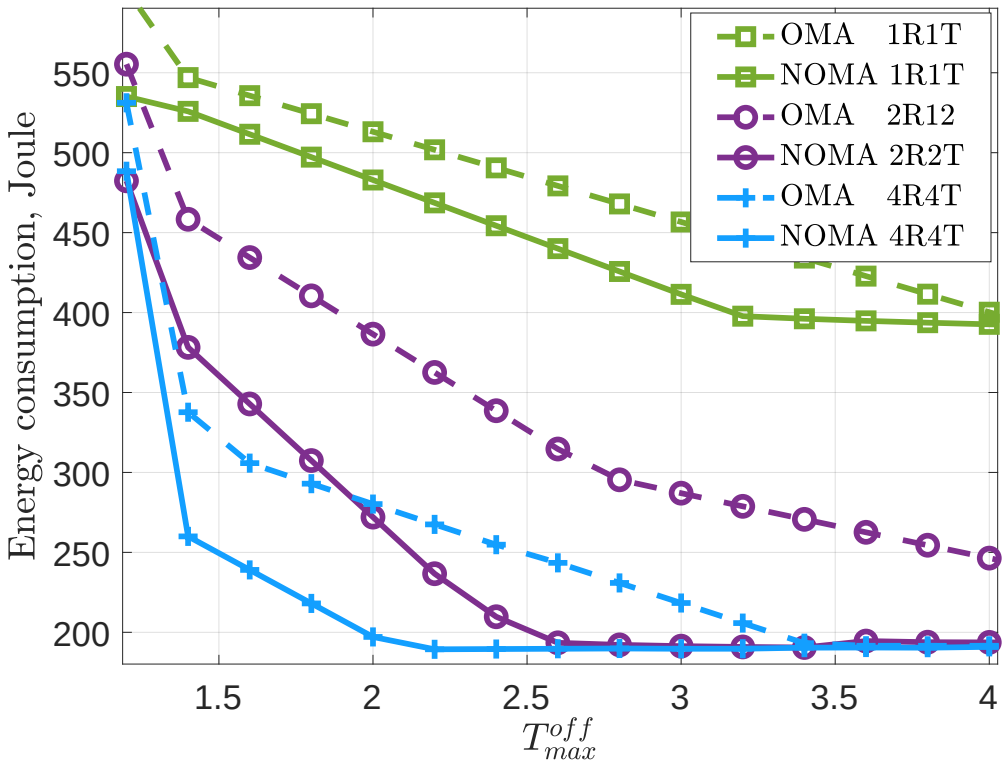

Figure 5.4. Total energy consumption for different antenna configurations with respect to offloading time

Figure 5.4 shows the energy consumption of NOMA and OMA-based MIMO-MEC systems with respect to offloading time. As shown in Figure 5.4, when the number of antennas increases, energy consumption decreases significantly in both OMA and NOMA schemes. This is because, as the offloading time is relaxed, a user can offload more data to the MEC server. And energy consumption of MEC servers is lower than local computing. Therefore, offloading data to the MEC is preferable from the energy perspective. Another intuition from Figure 5.4 is that the energy consumption for data offloading is low compared to the need for processing. Therefore the lines become almost equal when there is sufficient time for offloading. Figure 5.4 also shows that the NOMA-based MIMO-MEC system outperforms the OMA-based system when the users have stringent time constraints.

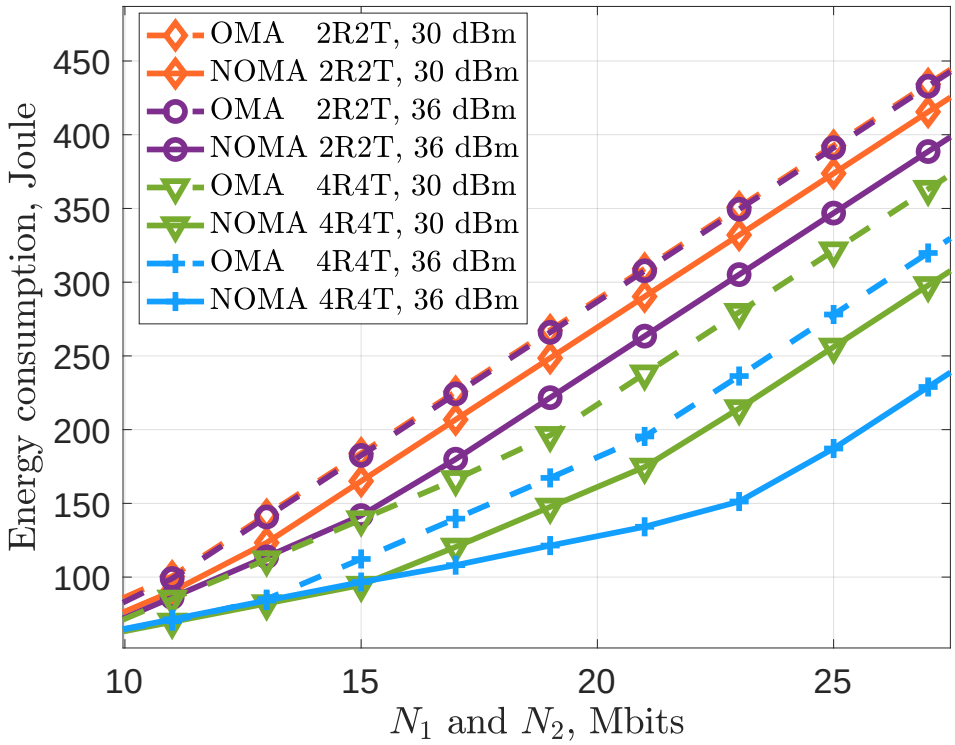

Figure 5.5. Total energy consumption for different antenna configurations and power levels

Figure 5.5 presents the impact of data sizes on energy consumption. As it is expected that the energy consumption will grow with the size of the offloaded data. However, after a point for each MIMO configuration, there is a rapid increase in energy consumption. This increase is because users can offload their data until they reach their power budgets. Once the users reach that point, they should process the remaining data locally, leading to a power consumption jump. The NOMA users can offload more data than OMA, so the performance range increases, especially when the data rate is high.

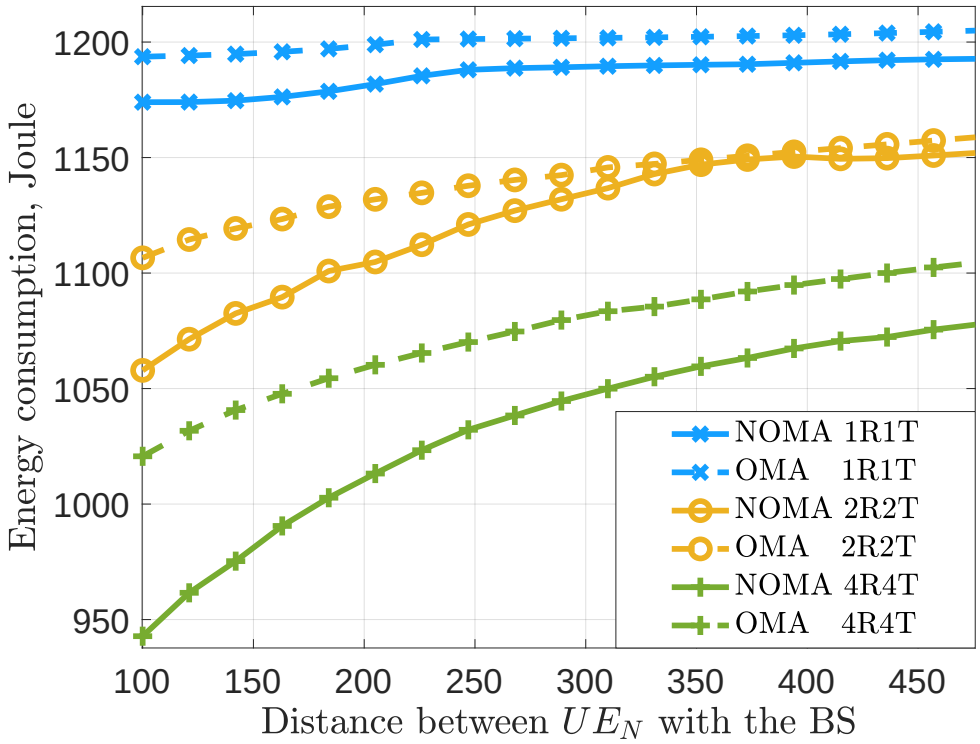

Figure 5.6. Total energy consumption concerning the near user's position

In Figure 5.6, the x-axis shows the distance of the near users from the base station when the far user is located 500 meters away from the base station. Due to co-user interference, the energy consumption significantly increases when the near user approaches the cell edge in the NOMA scheme. It is clear that the $4R4T$ NOMA performs best, but also it is the most affected configuration by co-user interference. It can be concluded from Figure 5.6 that user pairing can significantly improve the energy efficiency of the NOMA-MEC systems.

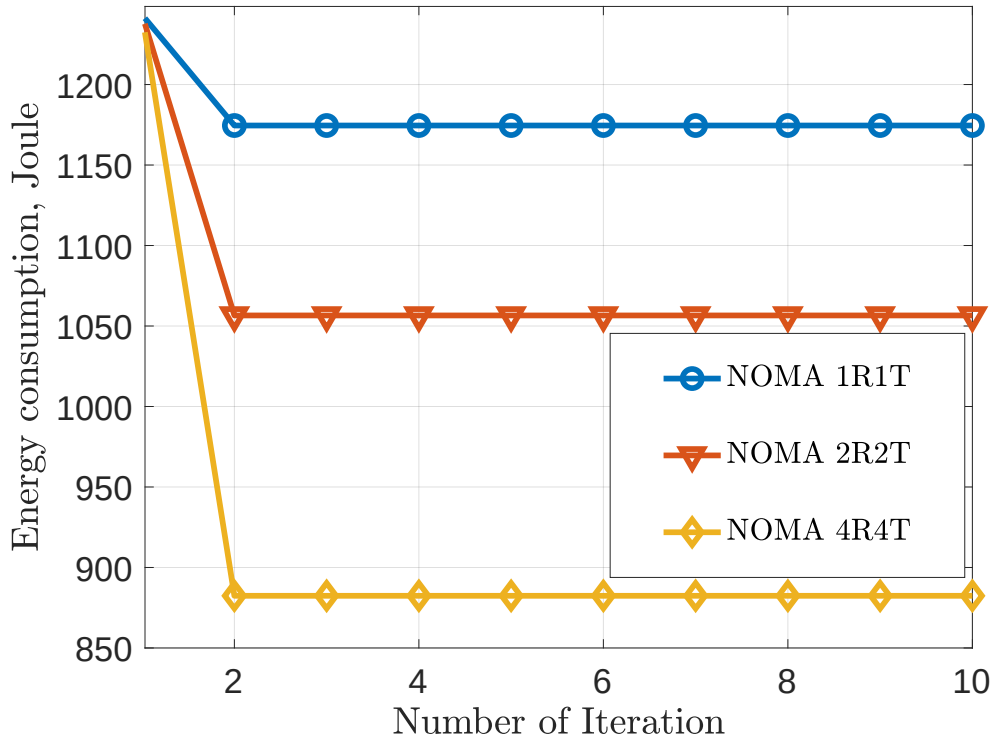

Figure 5.7. Total energy consumption concerning iteration number

Figure 5.7 shows the convergence performance of the proposed algorithm. The inner problem in (5.14) is linearized by the successive convex approximation method, and the outer layer is a linear problem. As a result, the proposed algorithm converges in the second iteration.

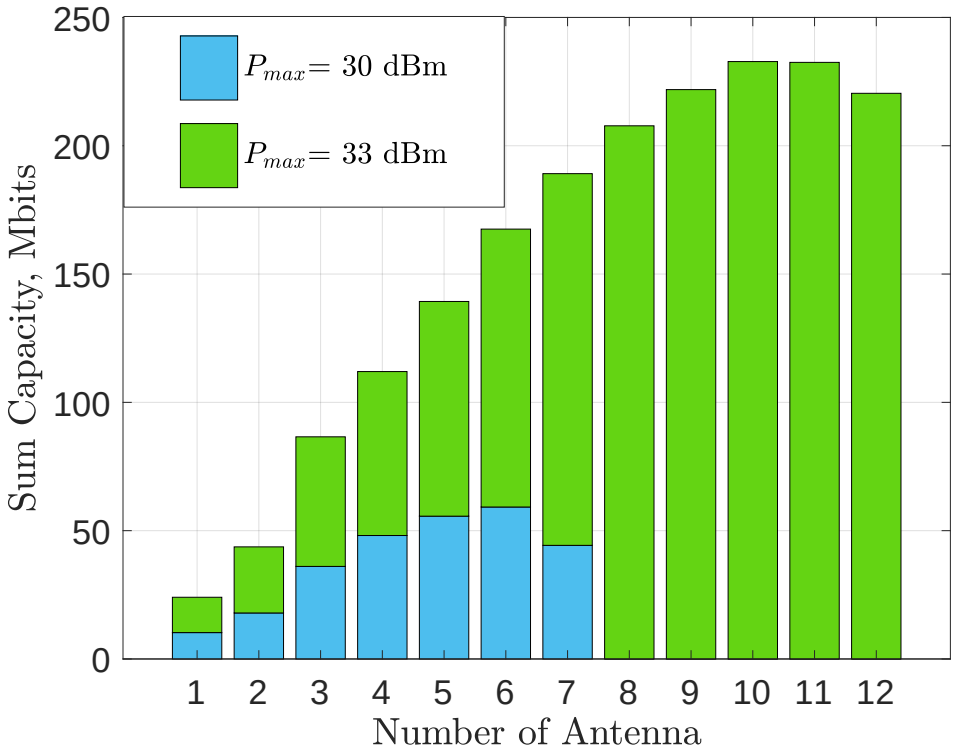

Figure 5.8. The effect of the antenna number over sum capacity

Figure 5.8 compares the sum capacity of different antenna configurations at various total power budgets. It can be seen from the figure that there is a trade-off between the number of antennas and the sum capacity. Sum capacity increases with the number of antennas up to an optimal point due to multiplexing gain. However, adding more antennas consumes more power in the RF chains. Therefore, less power is emitted from the antennas, and this reduces the sum capacity.

5.4 Conclusion of Chapter 5

NOMA, MIMO, and MEC are regarded as key technologies for 6G networks. Due to the computational intensive, delay-sensitive and energy-hungry applications, the energy consumption for NOMA-MIMO-MEC networks was an open research question. In this chapter, an energy minimization problem has been formulated, and a low complexity algorithm has been proposed. The numerical results have shown that the proposed NOMA-based MIMO-MEC system outperforms the conventional OMA-based scheme, especially when the users need to offload high and time-constrained data. It can be concluded from the simulation results that user pairing may significantly reduce the energy consumption of NOMA-MIMO-MEC Networks. This chapter has also found that increasing the antenna number does not necessarily mean higher sum capacity; therefore, when a MIMO system is designed, the optimal antenna number should be considered. A further study could assess the effects of user pairing, task scheduling, and other communication parameters, i.e., CPU frequencies, especially when the users are equipped with multiple antennas.

Chapter 6

Wireless Powered NOMA-Based Cognitive Radio for 6G Networks

It is envisioned that sixth-generation (6G) mobile networks will provide support for high capacity with peak rates of 1 Tbit/s and experienced rates of $10 - 100$ Gbit/s, as well as massive connectivity of up to 10 million/km^2, comprehensive coverage of at least 10 dB, ultra-reliability of 99.99999%, higher energy efficiency by a factor of 100 compared to 5G, and low latency of 0.1 ms [104]. However, there are also potential challenges that come with 6G networks. One such challenge is the potential for spectrum scarcity due to the increasing number of connected devices and the high demand for data, which could limit the available radio frequency spectrum. Powering connected mobile devices is another challenge. However, wireless power transfer improves their battery life and energy efficiency, contributes to sustainability efforts, and ensures their reliability and availability.

NOMA is a strong candidate to meet 6G requirements thanks to its spectral efficiency [105]-[106], improved secrecy [27], broad connectivity [107], high data capacity [108], and low latency [25] compared with orthogonal frequency division multiple access (OFDMA). Two main types of NOMA schemes exist code-domain NOMA and power-domain NOMA. This chapter focuses explicitly on power-domain NOMA, which employs superposition coding at the transmitter and successive interference cancellation (SIC) techniques at the receiver to increase the data capacity of orthogonal multiple access schemes [109].

Cognitive radio (CR) technology is another promising solution to tackle the issue of spectral scarcity, allowing unlicensed users to share the licensed spectrum for wireless

communication [110]. This chapter explores the CR-assisted MIMO-NOMA network, following the investigation of single-cell MIMO NOMA networks in Chapter 3, Chapter 4, and Chapter 5. There are three primary implementation approaches for CR: interweave, overlay, and underlay CR [111]. Interweave cognitive radio functions by transmitting signals within the gaps or unused spectrum of the primary wireless system without interfering with it [112]. In contrast, the cognitive radio device selects the best available frequency band in overlay CR to avoid interference with the primary system [113]. In this chapter, we optimize an underlay CR network, where secondary users operate concurrently with primary users in the same frequency band. Underlay CR necessitates careful power level management and interference control to prevent primary users' communication disruption.

Recently, a growing body of literature has explored the combination of NOMA-based underlay CR with other technologies to improve system performance. However, energy and spectral efficiency are critical parameters that often conflict with each other in these systems. Wireless energy harvesting (EH) is a promising technology for CR-NOMA networks, as it offers sustainability, cost-effectiveness, long-term reliability, energy efficiency, and independence. To address this issue, the use of reflected integrated surfaces (RIS) was proposed in [114], formulating the system model as a multi-objective optimization problem. An iterative block coordinate descent-based algorithm was developed to find an optimal balance between energy and spectral efficiency. Another approach suggested in [115] was to use an unmanned aerial vehicle (UAV) to improve the total throughput of a NOMA-based underlay CR network. A K-means and traveling salesman-based iterative algorithm were used to optimize the UAV's time, transmit power, and location. In [116], a mobile edge computer-assisted underlay CR-NOMA network was examined to enhance computation capacity while considering total weighted energy and physical layer security. In [117], the authors analyzed an overlay CR-NOMA network's throughput and outage probability under imperfect successive interference cancellation (SIC). Meanwhile, [118] proposed a non-linear EH model and optimized the power splitting factor and beamforming vector to minimize total power.

This chapter aims to tackle spectrum scarcity and energy supply issues by combining NOMA, CR, and EH technologies in an innovative, dynamic, and effective way. The key

contributions of this chapter are summarized below:

- Our proposal involves an underlay network that utilizes cognitive radio, MIMO and NOMA technologies. In this network, both the primary and secondary transmitters harvest wireless energy from a power beacon and utilize a time-switching protocol to transmit information signals to primary and secondary users. Despite possible interference experienced by the far user from the primary user, we ensure that the quality of service constraint is satisfied. Meanwhile, considering this constraint, the secondary transmitter optimizes the secondary network sum data rate.
- During the first time slot, wireless power is transferred using a beamforming method based on generalized singular value decomposition. In the second time slot, we use semi-definite programming and the first-order Taylor series expansion to optimize the split time variable (α) and the beamforming vectors for the primary ($\mathbf{w_p}$) and secondary ($\mathbf{w_s}$) networks.
- In the simulations, we study the impact of the QoS and power constraints on the sum rate of the secondary users. Based on the simulation results, the proposed model demonstrated superior performance compared to the TDMA-based benchmark scheme, primarily attributed to its higher spectral efficiency. This higher spectral efficiency allows the primary users to operate with lower power, affecting the energy consumption coefficient set by the secondary network and ultimately reducing interference for the secondary users. However, as the QoS requirements of the primary users become more stringent, the available transmission time for the secondary users decreases. As a result, the performance gap between the proposed method and the benchmark widens as the QoS requirements of the primary users increase.

6.1 System Model

We consider a NOMA-based heterogeneous cognitive network comprising a wireless-powered primary ultra-reliable low-latency communication (URLLC) network and a secondary massive machine-type communication (mMTC) network. The system is designed to accommodate multiple users with different requirements in the same resource

block, with the primary network allocating two NOMA users in a resource block while the secondary network accommodates multiple mMTC devices in the same resource block. The power beacon ($PBea$) simultaneously transmits energy signals to the primary and secondary networks and the harvested energy is utilized for downlink transmission. The primary transmitter (P_T) and secondary transmitter (S_T) are equipped with N_P and N_S antennas, respectively, while the power beacon is equipped with N_B antennas. The primary and secondary users are equipped with a single antenna, and the wireless channels are considered quasi-static, meaning that the channels remain constant between consecutive codewords. The underlay principle ensures that the secondary network does not interfere with the primary users' quality of service constraints. The energy transfer in the network occurs during αT, while the harvested energy is utilized during $(1-\alpha)T$, as depicted in Figure 6.2, where T represents the transmission time interval.

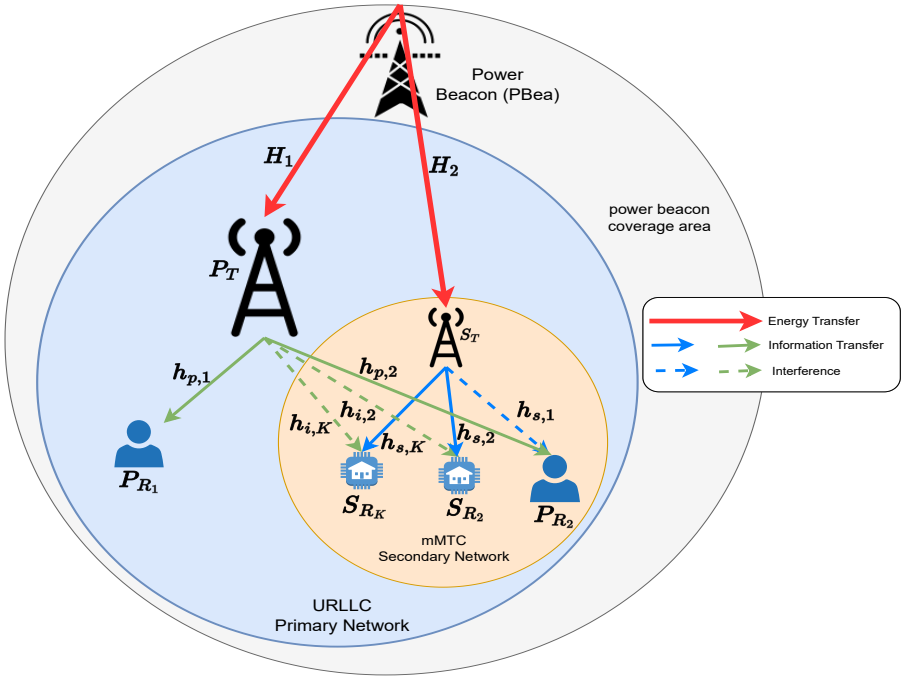

Figure 6.1. System model

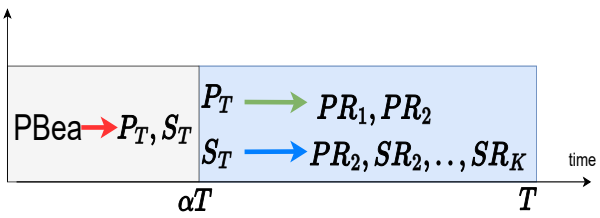

Figure 6.2. Transmission time frame

The received power signals can be expressed as follows at P_T and S_T, respectively:

$$\mathbf{y_P} = \mathbf{D_1 H_1 P x_{pow}} + \mathbf{n_P} \tag{6.1a}$$

$$\mathbf{y_S} = \mathbf{D_2 H_2 P x_{pow}} + \mathbf{n_S}, \tag{6.1b}$$

where the channel matrices between the power beacon and the primary and secondary base stations are denoted as $\mathbf{H_1} \in \mathbb{C}^{N_P \times N_B}$ and $\mathbf{H_2} \in \mathbb{C}^{N_S \times N_B}$, respectively. The precoding matrix of the power beacon is represented by $\mathbf{P} \in \mathbb{C}^{N_B \times N_B}$, while the decoding matrices for the primary and secondary users are $\mathbf{D_1}$ and $\mathbf{D_2}$, respectively. The power signal is denoted by $\mathbf{x_{pow}} \in \mathbb{C}^{N_B \times 1}$.

The signals transmitted by the primary and secondary networks, represented by $\mathbf{x_p}$ and $\mathbf{x_s}$, respectively, are transmitted for a duration of $(1-\alpha)T$. The signal received by the primary near, primary far, and the k^{th} secondary receiver are expressed as $y_{P_{R_1}}$, $y_{P_{R_2}}$, and $y_{P_{S_k}}$, respectively:

$$\mathbf{x_p} = \mathbf{w_{p,1}} s_{p,1} + \mathbf{w_{p,2}} s_{p,2} \tag{6.2a}$$

$$\mathbf{x_s} = \overbrace{\mathbf{w_{s,1}} s_{p,1}}^{underlay\ signal\ for\ PR_2} + \sum_{k=2}^{K} \mathbf{w_{s,k}} s_{s,k} \tag{6.2b}$$

$$y_{P_{R_1}} = \mathbf{h_{p,1}^H x_p} + n_{P_{R_1}} \tag{6.2c}$$

$$y_{P_{R_2}} = \mathbf{h_{p,2}^H x_p} + \mathbf{h_{s,1}^H x_s} + n_{P_{R_2}} \tag{6.2d}$$

$$y_{S_{R_k}} = \mathbf{h_{s,k}^H x_s} + \mathbf{h_{i,k}^H x_p} + n_{S_{R_k}}, \tag{6.2e}$$

where $s_{p,1}$, $s_{p,2}$, and $s_{s,k}$ are the information-bearing signals for the primary near, primary far, and k^{th} secondary users, respectively, and that $\|s_{p,v}\|^2 = 1$ for $v \in 1, 2, k$. If the power signal is normalized to one, i.e., $\|\mathbf{x_{pow}}\|^2 = 1$, then the maximum received power at P_T and S_T are denoted as P_1 and P_2, respectively.

$$P_1 = \mathbf{tr}(\mathbf{D_1 H_1 P P^H H_1^H D_1^H}) \tag{6.3a}$$

$$P_2 = \mathbf{tr}(\mathbf{D_2 H_2 P P^H H_2^H D_2^H}). \tag{6.3b}$$

Assuming that $PBea$, P_T, and P_S are each equipped with K antennas without loss of

generality. As a result, the SINR at P_{R_1}, P_{R_2}, and S_{R_K} is:

$$SNR_{P_1} = \frac{\|\mathbf{h_{p,1}^H w_{p,1}}\|^2}{\sigma_p^2} \tag{6.4a}$$

$$SINR_{P_2} = \frac{\|\mathbf{h_{p,2}^H w_{p,2}} + \mathbf{h_{s,1}^H w_{s,1}}\|^2}{\|\mathbf{h_{p,2}^H w_{p,1}}\|^2 + \sigma_p^2} \tag{6.4b}$$

$$SINR_{S_k} = \frac{\|\mathbf{h_{s,k}^H w_{s,k}}\|^2}{\sum_{j=1}^{k-1} \|\mathbf{h_{s,k}^H w_{s,j}}\|^2 + \|\mathbf{h_{i,k}^H w_p}\|^2 + \sigma_{n_{P_{S_k}}}^2}. \tag{6.4c}$$

6.2 Problem Definition and Solution

We assume that the UEs on the secondary network are arranged in descending order of $\|\mathbf{h_{s,1}^H w_{s,1}}\|^2 \geq \|\mathbf{h_{s,2}^H w_{s,2}}\|^2 \geq ... \geq \|\mathbf{h_{s,k}^H w_{s,K}}\|^2$. Additionally, we assume that the variances of $n_{P_{R_1}}$, $n_{P_{R_2}}$, and $n_{S_{R_K}}$ are all equal to BN_0. Following SIC, the achievable data rates for $R_{P_{R_1}}$, $R_{P_{R_2}}$, and $R_{P_{S_k}}$ are respectively provided as follows:

$$R_{P_{R_1}} = \log_2\left(1 + \frac{\|\mathbf{h_{p,1}^H w_{p,1}}\|^2}{BN_0}\right), bits/s/Hz \tag{6.5a}$$

$$R_{P_{R_2}} = \log_2\left(1 + \frac{\|\mathbf{h_{p,2}^H w_{p,2}} + \mathbf{h_{s,1}^H w_{s,1}}\|^2}{\|\mathbf{h_{p,2}^H w_{p,1}}\|^2 + BN_0}\right), bits/s/Hz \tag{6.5b}$$

$$R_{P_{S_k}} = \log_2\left(1 + \frac{\|\mathbf{h_{s,k}^H w_{s,k}}\|^2}{\sum_{j=1}^{k-1} \|\mathbf{h_{s,k}^H w_{s,j}}\|^2 + \|\mathbf{h_{i,k}^H w_p}\|^2 + BN_0}\right), bits/s/Hz. \tag{6.5c}$$

Therefore, we formulate the optimization problem as follows:

$$(\mathcal{P}_1) \quad \max_{\mathbf{P}, \mathbf{D_i}, \alpha, \mathbf{w_p}, \mathbf{w_s}} \quad \sum_{k=2}^{K} R_{P_{S_k}} \tag{6.6a}$$

$$\text{s.t.} \quad R_{P_{S_k}} \geq R_{s,k},\ k = \{2, .., K\} \tag{6.6b}$$

$$R_{P_{R_1}} \geq R_{t1} \tag{6.6c}$$

$$R_{P_{R_2}} \geq R_{t2} \tag{6.6d}$$

$$\|\mathbf{w_{p,1}}\|^2 + \|\mathbf{w_{p,2}}\|^2 \leq \frac{(\eta \times P_1)\alpha}{1-\alpha} \tag{6.6e}$$

$$\sum_{k=1}^{K} \|\mathbf{w_{s,k}}\|^2 \leq \frac{(\eta \times P_2)\alpha}{1-\alpha} \tag{6.6f}$$

$$0 < \alpha < 1 \tag{6.6g}$$

where η represents a power harvesting coefficient. The data rate requirement of the primary near, far, and secondary k^{th} users is denoted by R_{t1}, R_{t2}, and $R_{s,k}$, respectively. While $PBea$ sends power signal $\mathbf{x_{pow}}$ to P_T and S_T during αT, the expressions for P_1 and P_2 are non-convex, making optimization challenging. To address this, we utilize a GSVD-based linear beamforming scheme that efficiently exploits the wireless spectrum to define $\mathbf{P}$, $\mathbf{D_1}$, and $\mathbf{D_2}$ [25]. However, the original problem remains non-concave due to the non-convex constraints. To simplify the problem, we apply a low-complexity and highly efficient GSVD method [28] instead.

The GSVD can be briefly defined as follows: Consider $\mathbf{H_i} \in \mathbb{C}^{M_i,N}$, where $i = \{1, 2\}$, and its elements are i.i.d. with zero mean and unit variance. The decomposition of $\mathbf{H_i}$ by GSVD can be expressed as [28]:

$$\mathbf{H_i} = \mathbf{U_i}\boldsymbol{\Sigma_i}\mathbf{Q^{-1}}, \quad i = \{1, 2\}, \tag{6.7}$$

where $\mathbf{U_i}$ is $M_i \times M_i$ unitary matrix, $\mathbf{Q}$ is an $N \times N$ non-singular matrix, and $\boldsymbol{\Sigma_i}$=diag$(\sigma_{i,1}, .., \sigma_{i,\min(M_i)}) \in \mathbb{R}^{M_i \times \min(M_i)}$ is a diagonal matrix. (6.3a) and (6.3b) can be rewritten as follows:

$$P_1 = \mathbf{tr}(\mathbf{D_1U_1\Sigma_1Q^{-1}PP^HQ^{-H}\Sigma_1U_1^HD_1^H}) \text{ and } P_2 = \mathbf{tr}(\mathbf{D_2U_2\Sigma_2Q^{-1}PP^HQ^{-H}\Sigma_2U_2^HD_2^H}). \tag{6.8}$$

The GSVD method dictates that $\mathbf{P}$ should be set equal to $\mathbf{Q}$, and D_1 and D_2 should be equivalent to $\mathbf{U_1}^{-1}$ and $\mathbf{U_2}^{-1}$, respectively. As a result, we can express P_1 and P_2 as follows:

$$P_1 = \mathbf{tr}(\underbrace{\mathbf{U_1^{-1}U_1}}_{\mathbf{I}}\boldsymbol{\Sigma_1}\underbrace{\mathbf{Q^{-1}Q}}_{\mathbf{I}}\underbrace{\mathbf{Q^HQ^{-H}}}_{\mathbf{I}}\boldsymbol{\Sigma_1}\underbrace{\mathbf{U_1^HU_1^{-H}}}_{\mathbf{I}}) \tag{6.9a}$$

$$P_2 = \mathbf{tr}(\underbrace{\mathbf{U_2^{-1}U_2}}_{\mathbf{I}}\boldsymbol{\Sigma_2}\underbrace{\mathbf{Q^{-1}Q}}_{\mathbf{I}}\underbrace{\mathbf{Q^HQ^{-H}}}_{\mathbf{I}}\boldsymbol{\Sigma_2}\underbrace{\mathbf{U_2^HD_2^H}}_{\mathbf{I}}) \tag{6.9b}$$

$$P_1 = \mathbf{tr}(\boldsymbol{\Sigma_1^2}) \quad \text{and} \quad P_2 = \mathbf{tr}(\boldsymbol{\Sigma_2^2}). \tag{6.9c}$$

The original problem can be expressed by taking into account imperfect energy harvest-

ing as follows:

$$(\mathcal{P}_2) \quad \max_{\alpha,\mathbf{w_p},\mathbf{w_s}} \quad \sum_{k=2}^{K} R_{P_{S_k}} \tag{6.10a}$$

$$\text{s.t.} \quad R_{P_{S_k}} \geq R_{s,k},\ k = \{2,..,K\} \tag{6.10b}$$

$$R_{P_{R_1}} \geq R_{t1} \tag{6.10c}$$

$$R_{P_{R_2}} \geq R_{t2} \tag{6.10d}$$

$$\|\mathbf{w_{p,1}}\|^2 + \|\mathbf{w_{p,2}}\|^2 \leq \frac{\left[\frac{P_m}{1+\exp(-a_c(\mathbf{tr}(\Sigma_1^2)-b))-P_m\frac{1}{1+\exp(a_c b)}}\right]^{\frac{\alpha}{1-\alpha}}}{1-\frac{1}{1+\exp(a_c b)}} \tag{6.10e}$$

$$\sum_{k=1}^{K}\|\mathbf{w_{s,k}}\|^2 \leq \frac{\left[\frac{P_m}{1+\exp(-a_c(\mathbf{tr}(\Sigma_2^2)-b))-P_m\frac{1}{1+\exp(a_c b)}}\right]^{\frac{\alpha}{1-\alpha}}}{1-\frac{1}{1+\exp(a_c b)}} \tag{6.10f}$$

$$0 < \alpha < 1 \tag{6.10g}$$

where a_c and b are positive constants associated with the circuit specifications, and P_m represents the maximum quantity of power that can be harvested when the EH circuit is saturated [119]. Let denote $\hat{\mathbf{w}}_\mathbf{p}^\mathbf{H} = \left[\mathbf{w}_{\mathbf{p,1}}^\mathbf{H},\ \mathbf{w}_{\mathbf{p,2}}^\mathbf{H}\right]$, $\hat{\mathbf{w}}_\mathbf{s}^\mathbf{H} = \left[\mathbf{w}_{\mathbf{s,1}}^\mathbf{H},...,\mathbf{w}_{\mathbf{s,K}}^\mathbf{H}\right]$ and $\hat{\mathbf{h}}_\mathbf{i}^\mathbf{H} = \left[\mathbf{h}_{\mathbf{i,1}}^\mathbf{H},\ \mathbf{h}_{\mathbf{i,2}}^\mathbf{H}\right]$. Also, $\hat{\mathbf{h}}_{\mathbf{p,1}}$, $\hat{\mathbf{h}}_{\mathbf{p,2}}$, $\hat{\mathbf{h}}_\mathbf{i}$ and $\hat{\mathbf{h}}_{\mathbf{s,j}}$ are zero-padded vectors such that $\hat{\mathbf{h}}_{\mathbf{p,1}}^\mathbf{H} = [\mathbf{h}_{\mathbf{p,1}}^\mathbf{H},\ \mathbf{0}^\mathbf{H}]$, $\hat{\mathbf{h}}_{\mathbf{p,2}}^\mathbf{H} = [\mathbf{0}^\mathbf{H},\ \mathbf{h}_{\mathbf{p,2}}^\mathbf{H}]$, $\hat{\mathbf{h}}_\mathbf{i} = \left[\mathbf{h}_{\mathbf{i,1}}^\mathbf{H},\ \mathbf{h}_{\mathbf{i,2}}^\mathbf{H}\right]$, $\hat{\mathbf{h}}_{\mathbf{s,j}}^\mathbf{H} = [\mathbf{0}^\mathbf{H},\ \mathbf{h}_{\mathbf{s,j}}^\mathbf{H}]$ and $\mathbf{w}_\mathbf{y}^\mathbf{H} = [\mathbf{w}_\mathbf{p}^\mathbf{H},\ \mathbf{w}_\mathbf{s}^\mathbf{H}]$, as illustrated in Fig. 6.3.

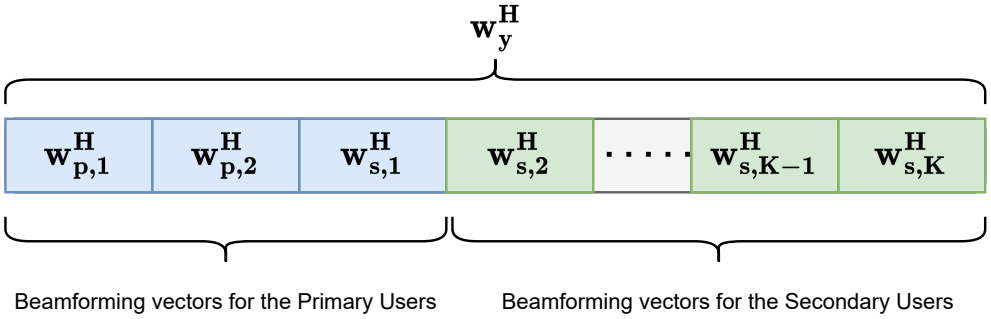

Figure 6.3. Concatenation of the beamforming vectors

Accordingly (6.10) can be written as

$$(\mathcal{P}_3) \quad \max_{\alpha,\mathbf{w_y}} \quad \sum_{k=2}^{K}\log_2\left(1+\frac{\|\hat{\mathbf{h}}_{\mathbf{s,k}}^\mathbf{H}\mathbf{w_y}\|^2}{\sum_{j=1}^{k-1}\|\hat{\mathbf{h}}_{\mathbf{s,k}}^\mathbf{H}\mathbf{w_y}\|^2+\|\hat{\mathbf{h}}_{\mathbf{i,k}}^\mathbf{H}\mathbf{w_y}\|^2+BN_0}\right)\times(1-\alpha) \tag{6.11a}$$

$$\text{s.t.} \quad \log_2\left(1+\frac{\|\hat{\mathbf{h}}_{\mathbf{s,k}}^{\mathbf{H}}\mathbf{w_y}\|^2}{\sum_{j=1}^{k-1}\|\hat{\mathbf{h}}_{\mathbf{s,k}}^{\mathbf{H}}\mathbf{w_y}\|^2+\|\hat{\mathbf{h}}_{\mathbf{i,k}}^{\mathbf{H}}\mathbf{w_y}\|^2+BN_0}\right)\geq\frac{R_{s,k}}{(1-\alpha)},\ k=\{2,..,K\} \tag{6.11b}$$

$$\log_2\left(1+\frac{\|\hat{\mathbf{h}}_{\mathbf{p,1}}^{\mathbf{H}}\mathbf{w_y}\|^2}{BN_0}\right)\geq\frac{R_{t1}}{(1-\alpha)} \tag{6.11c}$$

$$\log_2\left(1+\frac{\|\hat{\mathbf{h}}_{\mathbf{d}}^{\mathbf{H}}\mathbf{w_y}\|^2}{\|\hat{\mathbf{h}}_{\mathbf{p,2}}^{\mathbf{H}}\mathbf{w_y}\|^2+BN_0}\right)\geq\frac{R_{t2}}{(1-\alpha)} \tag{6.11d}$$

$$\|\mathbf{w_y}(\mathbf{1}:\mathbf{4})\|^2\leq\frac{EH_p\alpha}{1-\alpha} \tag{6.11e}$$

$$\|\mathbf{w_y}(\mathbf{5}:\mathbf{20})\|^2\leq\frac{EH_s\alpha}{1-\alpha} \tag{6.11f}$$

$$0<\alpha<1. \tag{6.11g}$$

Due to the fractional expressions in $(6.11a)$, $(6.11b)$ and $(6.11d)$, and coupling variables, $\mathcal{P}_3$ remains a non-convex optimization problem. To address this, we propose a two-layer alternating optimization algorithm, summarized in Algorithm 4. We fix α in the inner layer and find an optimal vector $\mathbf{w_y}$. In the outer layer, we keep $\mathbf{w_y}$ constant and obtain a sub-optimal solution for α. We repeat these steps until the objective value converges.

$$(\mathcal{P}_4)\quad \max_{\mathbf{W_y}}\quad \sum_{k=2}^{K}\log_2\left(1+\frac{\text{Tr}(\mathbf{H_{s,k}W_y})}{\sum_{j=1}^{k-1}\text{Tr}(\mathbf{H_{s,k}W_y})+\text{Tr}(\mathbf{H_{i,k}W_y})+BN_0}\right)\times(1-\alpha) \tag{6.12a}$$

$$\text{s.t.}\quad \text{Tr}(\mathbf{H_{s,k}W_y})\geq(2^{\frac{R_{s,k}}{1-\alpha}}-1)\times\left(\sum_{j=1}^{k-1}\text{Tr}(\mathbf{H_{s,k}W_y})+\text{Tr}(\mathbf{H_{i,k}W_y})+BN_0\right),$$

$$k=\{2,..,K\} \tag{6.12b}$$

$$\text{Tr}(\mathbf{H_{p,1}W_y})\geq BN_0\times(2^{\frac{R_{t1}}{1-\alpha}}-1) \tag{6.12c}$$

$$\text{Tr}(\mathbf{H_dW_y})\geq(2^{\frac{R_{t2}}{1-\alpha}}-1)\times(\text{Tr}(\mathbf{H_{p,2}W_y})+BN_0) \tag{6.12d}$$

$$\text{Tr}(\mathbf{W_y}(\mathbf{1}:\mathbf{4},\mathbf{1}:\mathbf{4}))\leq\frac{\alpha EH_p}{1-\alpha} \tag{6.12e}$$

$$\text{Tr}(\mathbf{W_y}(\mathbf{5}:\mathbf{20},\mathbf{5}:\mathbf{20}))\leq\frac{\alpha EH_s}{1-\alpha} \tag{6.12f}$$

$$0<\alpha<1 \tag{6.12g}$$

$$\mathbf{W_y}\succeq 0 \tag{6.12h}$$

$$\text{rank}(\mathbf{W_y}) = 1. \tag{6.12i}$$

By combining the term inside the logarithm, we reformulate $(\mathcal{P}_4)$ as $(\mathcal{P}_5)$.

$$(\mathcal{P}_5) \quad \max_{\mathbf{W_y}} \quad \sum_{k=2}^{K} \log_2 \left(\frac{\sum_{j=1}^{k} \text{Tr}(\mathbf{H_{s,k}}\mathbf{W_y}) + \text{Tr}(\mathbf{H_{i,k}}\mathbf{W_y}) + BN_0}{\sum_{j=1}^{k-1} \text{Tr}(\mathbf{H_{s,k}}\mathbf{W_y}) + \text{Tr}(\mathbf{H_{i,k}}\mathbf{W_y}) + BN_0} \right) \times (1-\alpha) \tag{6.13a}$$

$$\text{s.t.} \quad (6.12b),\ (6.12c),\ (6.12d),\ (6.12e),\ (6.12f),\ (6.12g),\ (6.12h),\ \text{and}\ (6.12i). \tag{6.13b}$$

SCA based solution

We propose an SCA-based solution for $(\mathcal{P}_5)$ in this subsection because it is still not concave due to the fractional expression. (6.13) can be rearranged as follows:

$$\begin{aligned}(\mathcal{P}_6) \max_{\mathbf{W_y}} \quad & \sum_{k=2}^{K} \log_2 \left(\sum_{j=1}^{k} \text{Tr}(\mathbf{H_{s,k}}\mathbf{W_y}) + \text{Tr}(\mathbf{H_{i,k}}\mathbf{W_y}) + BN_0 \right) \times (1-\alpha) \\ & - \sum_{k=2}^{K} \log_2 \left(\sum_{j=1}^{k-1} \text{Tr}(\mathbf{H_{s,k}}\mathbf{W_y}) + \text{Tr}(\mathbf{H_{i,k}}\mathbf{W_y}) + BN_0 \right) \times (1-\alpha) \end{aligned} \tag{6.14a}$$

$$\text{s.t.} \quad (6.12b),\ (6.12c),\ (6.12d),\ (6.12e),\ (6.12f),\ (6.12g),\ (6.12h),\ \text{and}\ (6.12i). \tag{6.14b}$$

Further, we can use the homogeneity principle as follows

$$\begin{aligned}(\mathcal{P}_7) \max_{\mathbf{W_y}} \quad & \underbrace{\sum_{k=2}^{K} \log_2 \left(\text{Tr}(\mathbf{W_y} \sum_{j=1}^{k} \mathbf{H_{s,k}}) + \text{Tr}(\mathbf{W_y}\mathbf{H_{i,k}}) + BN_0 \right)}_{f1(\mathbf{W_y})} \times (1-\alpha) \\ & - \underbrace{\sum_{k=2}^{K} \log_2 \left(\text{Tr}(\mathbf{W_y} \sum_{j=1}^{k-1} \mathbf{H_{s,k}}) + \text{Tr}(\mathbf{W_y}\mathbf{H_{i,k}}) + BN_0 \right)}_{f2(\mathbf{W_y})} \times (1-\alpha) \end{aligned} \tag{6.15a}$$

$$\text{s.t.} \quad (6.12b),\ (6.12c),\ (6.12d),\ (6.12e),\ (6.12f),\ (6.12g),\ (6.12h),\ \text{and}\ (6.12i). \tag{6.15b}$$

The objective function consists of summation of $f_1(\mathbf{W_y})$ and $f_2(\mathbf{W_y})$ multiplied by $(1-\alpha)$. Where $f_2(\mathbf{W_y})$ breaks concavity in the objective function, it can be approximated using

the first-order Taylor series. Furthermore, $f_2(\mathbf{W_y})$ is not a holomorphic function, which means that complex differentiation of $f_2(\mathbf{W_y})$ in each neighborhood of complex-valued $\mathbf{W_y}$ matrix does not exist. However, it is a real function of a complex matrix $\mathbf{W_y}$. Therefore, the complex gradient of $f_2(\mathbf{W_y})$ is equal to [43]

$$\frac{\partial f_2(\mathbf{W_y})}{\partial(\mathbf{W_y})} = \frac{1}{2}\left(\frac{\partial f_2(\mathbf{W_y})}{\partial \Re(\mathbf{W_y})} - i\frac{\partial f2(\mathbf{W_y})}{\partial \Im(\mathbf{W_y})}\right), \tag{6.16}$$

where $\Re\{\mathbf{W_y}\}$ and $\Im\{\mathbf{W_y}\}$ are the real and imaginary parts of $\mathbf{W_y}$, respectively. Note that $\mathbf{W_y}$ is a symmetric matrix; therefore, taking the derivative of $f_2(\mathbf{W_y})$ with respect to $\mathbf{W_y}$'s lower triangular part is sufficient to calculate the first order Taylor approximation of $f_2(\mathbf{W_y})$ [120]. Finally, (6.11) can be reformulated using DC programming as follows:

$$\begin{aligned}(\mathcal{P}_8) \quad \max_{\mathbf{W_y}} \quad & f_1(\mathbf{W_y}) - f_2(\mathbf{W_y}^{(t-1)}) \\ & - \sum_{z=1}^{N}\sum_{x=1}^{z-1} \Re\{\mathbf{W_y}(z,x) - \mathbf{W_y}(z,x)^{(t-1)}\} \times \frac{\partial f_2(\mathbf{W_y})}{\partial \Re(\mathbf{W_y}(z,x)^{(t-1)})} \qquad (6.17a) \\ & - \sum_{z=1}^{N}\sum_{x=1}^{z-1} \Im\{\mathbf{W_y}(z,x) - \mathbf{W_y}(z,x)^{(t-1)}\} \times \frac{\partial f_2(\mathbf{W_y})}{\partial \Im(\mathbf{W_y}(z,x)^{(t-1)})} \\ \text{s.t.} \quad & (6.12b),\ (6.12c),\ (6.12d),\ (6.12e),\ (6.12f),\ (6.12g),\ (6.12h), \text{ and } (6.12i), \qquad (6.17b)\end{aligned}$$

where (t) denotes the iteration number. The partial derivative of $f_2(\mathbf{W_y})$ with respect to the real and imaginary parts of $\mathbf{W_y}$ can be given by

$$\frac{\partial f_2(\mathbf{W_y})}{\partial \Re(\mathbf{W_y}(z,x))} = \frac{1}{\ln(2)} \times \sum_{k=2}^{K} \frac{\sum_{j=1}^{k-1}\left(\mathbf{H_{s,k}}(z,x) + \mathbf{H_{s,k}^H}(z,x)\right) + \mathbf{H_{i,k}}(z,x) + \mathbf{H_{i,k}^H}(z,x)}{\text{Tr}(\mathbf{W_y}\sum_{j=1}^{k-1}\mathbf{H_{s,k}}) + \text{Tr}(\mathbf{W_y}\mathbf{H_{i,k}}) + BN_0} \tag{6.18a}$$

$$\frac{\partial f_2(\mathbf{W_y})}{\partial \Im(\mathbf{W_y}(z,x))} = \frac{-i}{\ln(2)} \times \sum_{k=2}^{K} \frac{\sum_{j=1}^{k-1}\left(\mathbf{H_{s,k}}(z,x) - \mathbf{H_{s,k}^H}(z,x)\right) + \mathbf{H_{i,k}}(z,x) - \mathbf{H_{i,k}^H}(z,x)}{\text{Tr}(\mathbf{W_y}\sum_{j=1}^{k-1}\mathbf{H_{s,k}}) + \text{Tr}(\mathbf{W_y}\mathbf{H_{i,k}}) + BN_0}, \tag{6.18b}$$

where (z,x) represent the z-th row and x-th column of the matrices.

The fundamental concept of Gaussian randomization involves generating random solutions from a Gaussian distribution. These random solutions undergo a feasibility check for the optimization problem, and if deemed feasible, the best solution within the feasible set is selected as the final solution. The original optimization problem (6.10) was first divided into two sub-problems (6.17) and (6.19). Then, the SCA-based solution was proposed in (6.2). (6.11) is first transformed into a semi-definite programming problem by relaxing the non-convex constraint $\text{rank}(\mathbf{W_y}) = 1$. The rest of the problem is solved efficiently in (6.17). Gaussian randomization method was used to obtain an optimal $\mathbf{w_y^*}$ from $\mathbf{W_y}$ as shown in Algorithm 4.

6.2.1 A Solution for the Time Switching Coefficient in the Outer Layer

After finding an optimal solution for $\mathbf{w_y}$, an optimal solution for α can be found as follows.

$$(\mathcal{P}_9) \quad R_s^{NOMA} = \max_{\alpha} \quad \sum_{k=2}^{K} \log_2 \left(1 + \frac{\|\mathbf{h_{s,k}^H w_y^*}\|^2}{\sum_{j=1}^{k-1} \|\mathbf{h_{s,k}^H w_y^*}\|^2 + \|\mathbf{h_{i,k}^H w_y^*}\|^2 + BN_0} \right) \times (1-\alpha) \tag{6.19a}$$

$$\log_2 \left(1 + \frac{\|\mathbf{h_{s,k}^H w_y^*}\|^2}{\sum_{j=1}^{k-1} \|\mathbf{h_{s,k}^H w_y^*}\|^2 + \|\mathbf{h_{i,k}^H w_y^*}\|^2 + BN_0} \right) \times (1-\alpha) \geq R_{s,k},$$

$$k = \{2, .., K\} \tag{6.19b}$$

$$\log_2 \left(1 + \frac{\|\mathbf{h_{p,1}^H w_y^*}\|^2}{BN_0} \right) \times (1-\alpha) \geq R_{t1} \tag{6.19c}$$

$$\log_2 \left(1 + \frac{\|\mathbf{h_d^H w_y^*}\|^2}{\|\mathbf{h_{p,2}^H w_y^*}\|^2 + BN_0} \right) \times (1-\alpha) \geq R_{t2} \tag{6.19d}$$

$$\|\mathbf{w_y^*(1:4)}\|^2 \leq \frac{\alpha E H_p}{1-\alpha} \tag{6.19e}$$

$$\|\mathbf{w_y^*(5:20)}\|^2 \leq \frac{\alpha E H_s}{1-\alpha} \tag{6.19f}$$

$$0 < \alpha < 1. \tag{6.19g}$$

Algorithm 4 SCA and alternating optimization-based algorithm to find sub-optimal $\mathbf{w_y}$ and α

repeat
  Initialize α
  Find out an optimal $\mathbf{W_y^*}$ by solving (6.17)
  $\mathbf{w_{y,j}} = \text{mvnrnd}(zeros(1, N), \mathbf{W_y^*}, \text{J})$, where $\text{j} = 1, 2, ..., \text{J}$
  repeat
    if $\mathbf{w_{y,j}}$ is feasible **then**
      Calculate $\sum_{k=2}^{K} R_{P_{S_k}}$ using (6.19a)
    end if
    $\text{j} = \text{j} + 1$
  until $\text{j} \leq \text{J}$
  Find $\mathbf{w_{y,j^*}} = \arg\max_{\mathbf{w,j}}(\sum_{k=2}^{K} R_{P_{S_k}}), j = 1, 2, ..., \text{J}$
  Solve (6.19) to find out α^*
until $\Delta \sum_{k=2}^{K} R_{P_{S_k}} \to 0$

6.2.2 A TDMA-Based Benchmark Scheme

This section defines a TDMA-based benchmark scheme that prioritizes the primary users, as in the NOMA case. The aim is to maximize the sum rate of the secondary users considering the QoS constraints of the primary users on time. The system model for the benchmark can be described as follows:

$$(\mathcal{P}_{10}) \quad \max_{\alpha, p_{p,1}, p_{p,2}, p_{s,k}} \quad \sum_{k=2}^{K} \log_2\left(1 + \frac{p_{s,k}\|\mathbf{h_{s,k}^H}\|^2}{BN_0}\right) \times \frac{1-\alpha-t_1-t_2}{K-1} \tag{6.20a}$$

$$\text{s.t} \quad \frac{R_{t,1}}{\log_2\left(1 + \frac{p_{p,1}\|\mathbf{h_{p,1}^H}\|^2}{NB_0}\right)} \leq t_1 \tag{6.20b}$$

$$\frac{R_{t,2}}{\log_2\left(1 + \frac{p_{p,2}\|\mathbf{h_{p,2}^H}\|^2}{NB_0}\right)} \leq t_2 \tag{6.20c}$$

$$p_{p,1} + p_{p,2} \leq \frac{\alpha E H_p}{1-\alpha} \tag{6.20d}$$

$$\sum_{k=2}^{K} p_{s,k} \leq \frac{\alpha E H_s}{1-\alpha} \tag{6.20e}$$

$$p_{p,1}, p_{p,2}, p_{s,k} \geq 0, k = \{2, .., K\} \tag{6.20f}$$

$$0 < \alpha < 1. \tag{6.20g}$$

where t_1 and t_2 denote the transmission time slot for the primary users. We propose an alternating algorithm that finds an optimal value for α and the other parameters as given in Algorithm 5. After fixing α, it becomes clear that $p_{p,1}$ and $p_{p,2}$ are independent of $p_{s,k}$. This allows us to divide Eq. (6.20) into two sub-problems, denoted as $(\mathcal{P}11)$ and $(\mathcal{P}13)$. The objective of $(\mathcal{P}_{11})$ is to meet the QoS requirements of the primary users with minimal latency. Following that, for $(\mathcal{P}_{13})$, our objective is to maximize the total data

Algorithm 5 Alternating optimization-based algorithm to find sub-optimal α, $p_{p,1}$, $p_{p,2}$, and $p_{s,k}$

Initialize α
repeat
  Set $l = 1$
  Initialize $p_{p,1}$ to some value, such as $p_{p,1}^{(0)} = EHp/2$
  repeat
    Calculate the gradient of $f_3(p_{p,1})$ as in (6.23)
    Update $p_{p,1}$ by using (6.24)
    $l = l + 1$
  until $\Delta f_3(p_{p,1}) \to 0$
  Calculate t_1 and t_2 and plug them in (6.26b)
  Update $p_{s,k}$ using (6.27)
  Update α using (6.28)
until $\Delta \sum_{k=2}^{K} R_{P_{S_k}} \to 0$

rate of the secondary users within the remaining time slot.

$$(\mathcal{P}_{11}) \quad \min_{p_{p,1},p_{p,2}} \quad \frac{R_{t,1}}{\log_2\left(1+\frac{p_{p,1}\|\mathbf{h_{p,1}^H}\|^2}{NB_0}\right)} + \frac{R_{t,2}}{\log_2\left(1+\frac{p_{p,2}\|\mathbf{h_{p,2}^H}\|^2}{NB_0}\right)} \tag{6.21a}$$

$$\text{s.t} \quad p_{p,1} + p_{p,2} \leq \frac{\alpha^* EH_p}{1-\alpha^*} \tag{6.21b}$$

$$p_{p,1}, p_{p,2} \geq 0. \tag{6.21c}$$

We observe that $(\mathcal{P}11)$ is not convex. To minimize the delay for the primary users and provide more time to the secondary users, we need to ensure that the allocated power for the primary users sums up to the maximum available power, i.e., $p_{p,1} + p_{p,2} = \frac{\alpha^* EH_p}{1-\alpha^*}$. Thus, we can express $p_{p,2}$ as $(\frac{\alpha^* EH_p}{1-\alpha^*} - p_{p,1})$. Therefore, $(\mathcal{P}_{11})$ can be reorganized as $(\mathcal{P}_{12})$.

$$(\mathcal{P}_{12}) \quad \min_{p_{p,1}} \quad \underbrace{\frac{R_{t,1}}{\log_2\left(1+\frac{p_{p,1}\|\mathbf{h_{p,1}^H}\|^2}{NB_0}\right)} + \frac{R_{t,2}}{\log_2\left(1+\frac{(\frac{\alpha^* EH_p}{1-\alpha^*}-p_{p,1})\|\mathbf{h_{p,2}^H}\|^2}{NB_0}\right)}}_{f_3(p_{p,1})} \tag{6.22a}$$

$$\text{s.t} \quad 0 \leq p_{p,1} \leq \frac{\alpha^* EH_p}{1-\alpha^*}. \tag{6.22b}$$

To solve $(\mathcal{P}_{12})$, the SCA method is applied to (6.22a). The derivative of the objective function can be taken with respect to $p_{p,1}$ as follows:

$$\frac{df_3(p_{p,1})}{dp_{p,1}} = -\frac{5\ln(2)\left\|\mathbf{h_{p,2}^H}\right\|^2}{NB_0\ln^2\left(1-\frac{\left\|\mathbf{h_{p,2}^H}\right\|^2(p_{p,1}-\frac{\alpha^* EH_p}{1-\alpha^*})}{NB_0}\right)\left(\frac{\left\|\mathbf{h_{p,2}^H}\right\|^2(p_{p,1}-\frac{\alpha^* EH_p}{1-\alpha^*})}{NB_0}-1\right)}$$

$$- \frac{R_{t,1} \ln(2) \left\| \mathbf{h}_{\mathbf{p,1}}^{\mathbf{H}} \right\|^2}{NB_0 \ln^2 \left(\frac{p_{p,1} \left\| \mathbf{h}_{\mathbf{p,1}}^{\mathbf{H}} \right\|^2}{NB_0} + 1 \right) \left(\frac{P1 \left\| \mathbf{h}_{\mathbf{p,1}}^{\mathbf{H}} \right\|^2}{NB_0} + 1 \right)} \tag{6.23}$$

An optimal solution for $p_{p,1}$ can be found by $(\mathcal{P}_{13})$ in each iteration as defined in Algorithm 5.

$$(\mathcal{P}_{13}) \quad \min_{p_{p,1}} \quad f_3(p_{p,1}^l) + \frac{df_3(p_{p,1}^l)}{dp_{p,1}}(p_{p,1} - p_{p,1}^l) \tag{6.24a}$$

$$\text{s.t} \quad 0 \leq p_{p,1} \leq \frac{\alpha^* EH_p}{1 - \alpha^*}. \tag{6.24b}$$

where l denotes the iteration number. Using the optimal value for $p_{p,1}^*$, t_1 and t_2 can be given as follows:

$$t_1 = \frac{R_{t,1}}{\log_2 \left(1 + \frac{p_{p,1}^* \|\mathbf{h}_{\mathbf{p,1}}^{\mathbf{H}}\|^2}{NB_0}\right)} \tag{6.25a}$$

$$t_2 = \frac{R_{t,2}}{\log_2 \left(1 + \frac{(\frac{\alpha^* EH_p}{1-\alpha^*} - p_{p,1}^*) \|\mathbf{h}_{\mathbf{p,2}}^{\mathbf{H}}\|^2}{NB_0}\right)} \tag{6.25b}$$

$$(\mathcal{P}_{14}) \quad \max_{p_{s,k}} \quad \sum_{k=2}^{K} \log_2 \left(1 + \frac{p_{s,k} \|\mathbf{h}_{\mathbf{s,k}}^{\mathbf{H}}\|^2}{BN_0}\right) \times \frac{1 - \alpha^* - (t_1 + t_2)}{K - 1} \tag{6.26a}$$

$$\text{s.t} \quad \sum_{k=2}^{K} p_{s,k} \leq \frac{\alpha^* EH_s}{1 - \alpha^*} \tag{6.26b}$$

$$p_{s,k} \geq 0, k = \{2, .., K\}. \tag{6.26c}$$

The closed-form solution for $p_{s,k}$ can be found by using the waterfilling method as follows:

$$p_{s,k}^* = \left[\frac{1}{\mu_k} - \frac{BN_0}{\|\mathbf{h}_{\mathbf{s,k}}^{\mathbf{H}}\|^2} \right]^+, \; s.t. \; \sum_{k=2}^{K} \mu_k = EH_s. \tag{6.27}$$

$$(\mathcal{P}_{15}) \quad \max_{\alpha} \quad \sum_{k=2}^{K} \log_2 \left(1 + \frac{p_{s,k} \|\mathbf{h}_{\mathbf{s,k}}^{\mathbf{H}}\|^2}{BN_0}\right) \times \frac{1 - \alpha - t_1 - t_2}{K - 1} \tag{6.28a}$$

$$\text{s.t} \quad p_{p,1} + p_{p,2} \leq \frac{\alpha EH_p}{1 - \alpha} \tag{6.28b}$$

$$\sum_{k}^{K} p_{s,k} \leq \frac{\alpha EH_s}{1 - \alpha} \tag{6.28c}$$

$$0 < \alpha < 1. \tag{6.28d}$$

Equivalently, the closed-form solution for α is given as follows:

$$\alpha^* = \min\left(\frac{p_{p,1} + p_{p,2}}{p_{p,1} + p_{p,2} + EH_p}, \frac{\sum_{j=2}^{K} p_{s,j}}{\sum_{j=2}^{K} p_{s,j} + EH_s}\right) \tag{6.29}$$

6.2.3 Complexity Analysis

This subsection presents the time complexity analysis of the proposed algorithm. Algorithm 4 consists of two loops: an outer loop that uses a linear program to solve for α, and an inner loop that utilizes semi-definite programming to solve for $\mathbf{w_p}$ and $\mathbf{w_s}$. In the other problem for α, the worst-case time complexity of the simplex method, a linear program with n variables and m constraints, is $\mathcal{O}(2^n)$, where n is the number of variables and m is the number of constraints. Please note that n is equal to 1 in this case. In the inner loop, the worst-case time complexity of solving a semi-definite relaxation using interior point methods is $\mathcal{O}\left((c+v)^{6.5}\ln(1/\epsilon)\right)$, where c is the number of variables, v is the number of constraints, and ϵ is the desired accuracy. Above all, the complexity of the GSVD-based precoding is $\mathcal{O}\left(\frac{L^3}{\sigma} + \frac{L}{\sigma}\right)$, where L is the minimum number of antennas between $PBea$ and P_T and S_T and σ is the search step [26]. Finally, the time complexity of the proposed algorithm is $\mathcal{O}\left((\frac{L^3}{\sigma} + \frac{L}{\sigma}) + 2(c+v)^{6.5}\ln(1/\epsilon)\right)$.

6.3 Numerical Results

In this section, we evaluate the performance of the proposed method. As a benchmark scheme, we use a TDMA-based system defined in Section (6.2.2). Table 6.1 presents the essential simulation parameters.

Table 6.1. Simulation parameters

Parameter	Value
The number of antenna and users at the primary and secondary networks	2 and 4
Distance from $PBea$ to P_T and P_S	10 m
Distance from P_T to $P_{R,1}$ and $P_{R,2}$	15 m and 25 m
Distance from S_T to $P_{S,2}$, $P_{S,3}$ and $P_{S,4}$	10 m, 15 m, 20 m and 25 m
Distance from P_T to secondary users	10 m, 15 m, 20 m and 25 m
Noise power σ_p^2 and $\sigma_{s,k}^2$	-94 dBm/Hz
Path loss components for the primary network α_p and secondary network α_s	3.2 and 3

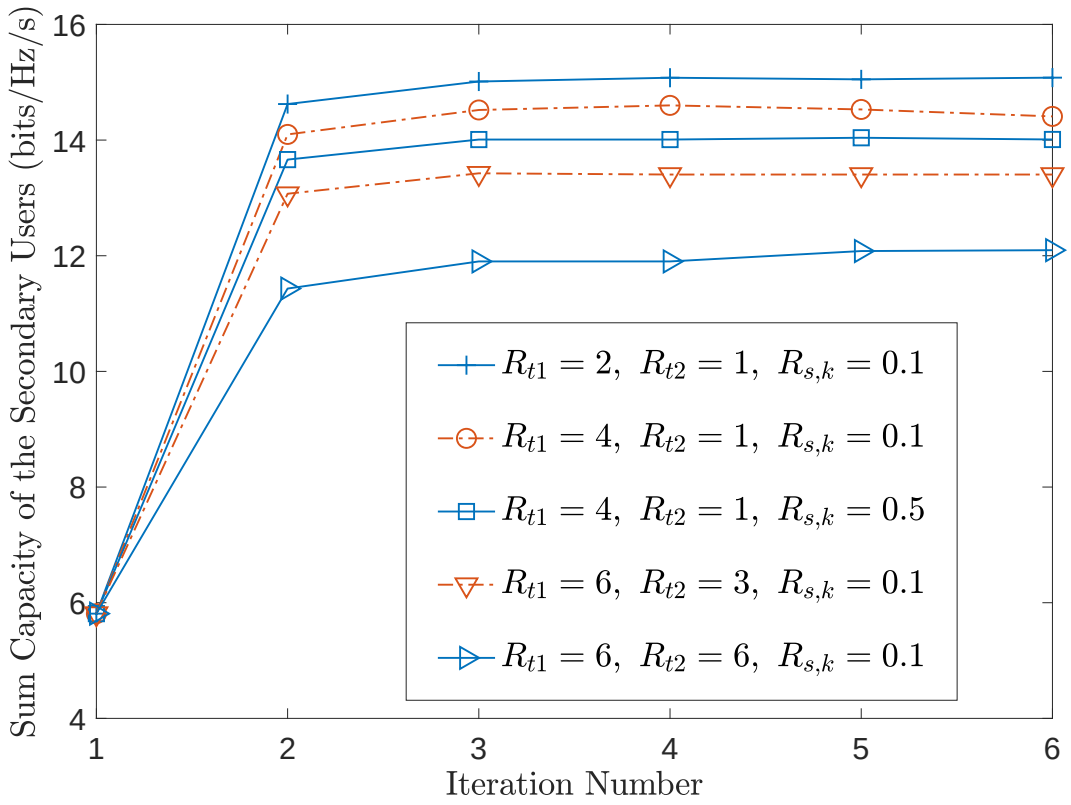

Figure 6.4. Convergence of the proposed method

The convergence performance of the proposed algorithm is shown in Figure 6.4. It can be observed that the algorithm's solution experiences significant changes in the first three iterations, resulting in a rapid increase in the sum rate of the secondary users. However, after the third iteration, the increase stabilizes, suggesting that further iterations may not yield significant improvements.

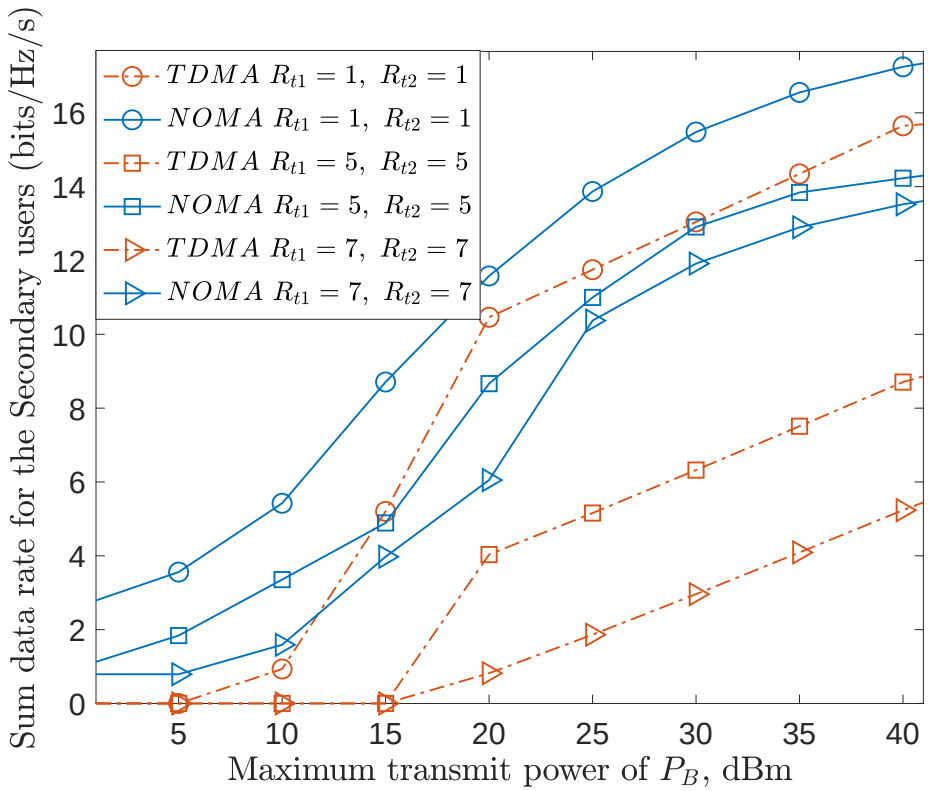

Figure 6.5. The effect of QoS and transmitted power on the sum capacity of the secondary network

Figure 6.5 presents the relationship between the primary network's QoS requirements and the secondary users' sum rate. It can be seen from the figure that the secondary users' sum rate decreases as the primary network's required QoS increases. This can be attributed to two key factors: firstly, the signals intended for the primary users are perceived as interference by the secondary users, and secondly, if the primary network fails to meet its QoS demands, the distant user within the primary network collaborates with the secondary transmitter, resulting in energy loss and interference for the secondary network. The proposed method outperforms the benchmark TDMA model due to its higher spectral efficiency. With superior spectral efficiency, the primary users necessitate less power, effectively determining the energy consumption coefficient established by the secondary network, and reducing interference for the secondary users. However, as the primary users' QoS requirements increase, the secondary users' available transmission time is shortened. Consequently, the performance gap between the proposed method and the benchmark widens.

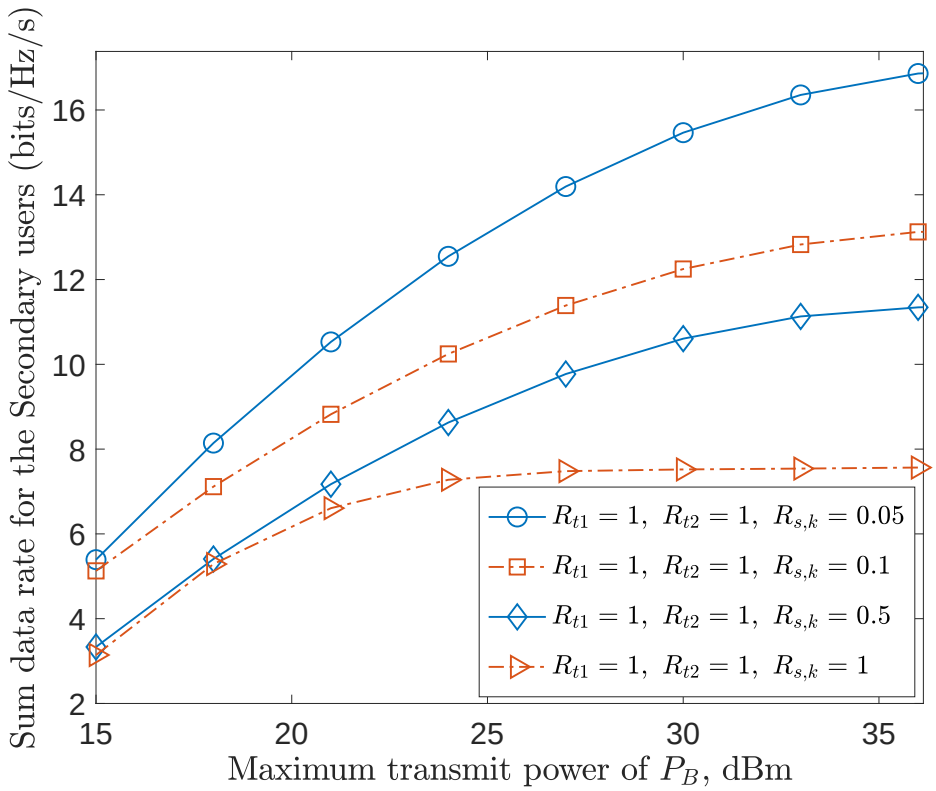

Figure 6.6. The effect of minimum rate constraint for the secondary users on the sum capacity

Figure 6.6 illustrates the variations in the sum rate of the secondary users as the QoS threshold for the secondary users' changes. These constraints are implemented to ensure fairness among the users, but they result in substantial reductions in the sum rate. The decrease in the sum rate is primarily attributed to the minimum SINR rate experienced by the furthest user, caused by interference from the preceding users. Consequently, to meet the QoS constraint specified in Eq. (6.10b), the previous users are compelled to reduce their power, leading to a loss in the overall sum rate.

6.4 Conclusion of Chapter 6

The chapter introduced a new wireless-powered mobile device approach using NOMA and MU-MIMO antenna systems. This solution was tailored for a cognitive underlay radio scenario where a primary network demanded a specific QoS while secondary network users utilized the same spectrum for downloading their data. The proposed method optimized a joint beamforming vector for primary and secondary networks and a time-switching coefficient for energy harvesting and information transfer. The problem formulation was non-convex; therefore, we used semi-definite programming, successive convex approximation, and alternating optimization to develop an effective and low-complexity algorithm. The NOMA-based solution achieved a higher achievable data rate than the TDMA-based benchmark scheme and outperformed it, especially in the low transmit power region.

Chapter 7

Conclusions and Future Works

7.1 Conclusions of the Thesis

6G Networks require massive connectivity, ultra-reliability, secure and high-speed data transfer, and diverse QoS requirements. NOMA-MIMO-based systems are promising to overcome these demands. This thesis has addressed various aspects of wireless communication by combining NOMA and MIMO technologies with GSVD-based linear beamforming to optimize the performance of the proposed systems. The central contributions and key insights of this thesis can be encapsulated as follows:

Chapter 2 presented an explanation and comparison of NOMA with OMA schemes. Furthermore, the required mathematical tools used in the thesis were introduced in this chapter.

Chapter 3 focused on the delay minimization problem in the hybrid-NOMA-based MIMO-MEC offloading scenario. An iterative closed-form solution was obtained using the Dinkelbach method to handle the fractional expression. Simulation results demonstrated that the proposed method significantly improved delay performance and reduced energy consumption in the MIMO-MEC system.

Chapter 4 explored enhancing physical layer security in NOMA-MIMO systems. The objective of maximizing sum secrecy rates under power and quality-of-service constraints was executed using an SCA-based algorithm. The results indicated that NOMA outperformed other benchmarks, particularly in high SINR. From this chapter, it is evident that the presence of multiple antennas at both the transmitter and receiver plays a pivotal role in significantly enhancing the system's delay performance.

Chapter 5 investigated the energy consumption of NOMA-MIMO-MEC networks, considering the rise of computational-intensive and delay-sensitive applications. An energy minimization problem was formulated, and a low-complexity algorithm was proposed. The numerical results demonstrated the superiority of the NOMA-based MIMO-MEC system over conventional OMA-based schemes, particularly for high and time-constrained data offloading. This chapter's findings lead to the conclusion that augmenting the number of antennas within the confines of a constant power budget does not uniformly escalate data rates, mainly owing to power consumption within RF chains. Additionally, the impact of user pairing on energy consumption within the NOMA MEC system has been illuminated.

Chapter 6 proposed a new protocol for wireless-powered mobile users utilizing NOMA and MIMO technologies. The primary network has pre-defined quality of service (QoS) requirements, which coexist with a secondary network. The secondary network operates by utilizing the underlay cognitive radio technique to download data. By optimizing joint beamforming and time-switching coefficients, the NOMA-based approach achieved higher achievable data rates compared to the benchmark TDMA-based scheme, particularly at low transmit power levels. The pivotal discovery of this chapter underscores that NOMA exhibits superior performance particularly as the quality of service constraint escalates for primary users.

In summation, this thesis traversed the intricacies of cutting-edge wireless communication paradigms, culminating in the fusion of NOMA and MIMO technologies bolstered by advanced beamforming strategies. The results presented within each chapter collectively underscore the potency of the proposed methodologies in advancing the domain of 6G Networks, charting a compelling course for the future of wireless communication systems.

7.2 Future Works

1. A promising direction for future research involves addressing the limitation of assuming perfect CSI availability at the transmitters and receivers. One possible extension is to explore the implications and benefits of incorporating imperfect chan-

nel estimation with random error matrices, as demonstrated in [121]. By adopting this approach, the studies conducted in the thesis can be expanded to real-world scenarios where imperfect CSI is more prevalent.

2. Another limitation arises from the constraints of the GSVD technique, which only allows the combination of two users simultaneously. Fortunately, recent advances in the field have shown that the GSVD technique can be extended to more than two users, as demonstrated in [122] and [123]. The results obtained indicate that the proposed methods have considerable potential for expansion beyond two MIMO-NOMA users. This enables assigning more than two users in a resource block.

3. An enticing avenue for future research involves integrating the proposed system models with emerging technologies such as intelligent reflecting surfaces and unmanned aerial vehicles. This integration holds significant potential and offers promising opportunities to enhance the received SINR of the MIMO-NOMA systems.

4. The research conducted in Chapter 5 explores the potential benefits of an optimal user pairing that reduces the energy consumption of a pair of NOMA users. Therefore, the proposed method in Chapter 5 could be further extended to a multi-user scenario considering the total minimization of the whole system. Thus this research idea would provide valuable insights into the efficacy of user pairing in large-scale user deployments.

5. The remarkable performance improvement achieved by hybrid-SIC ordering in a SISO-NOMA network highlights its significance in SISO-NOMA networks [15]. However, this issue is even more crucial in MIMO cases, as elaborated in the context of GSVD-based MIMO-NOMA networks in Section 2.2.2. Notably, to the best of the authors' knowledge, this particular issue has not been thoroughly investigated in the existing literature. Consequently, there is a research gap that necessitates further exploration and analysis.

6. In this thesis, the solutions primarily rely on employing convex approximations to convert non-convex problems into convex. In order to overcome the limitations arising from the conversions, which often result in suboptimal solutions and potential performance loss, a promising direction for future work involves investigat-

ing the application of machine learning-based algorithms. By leveraging machine learning capabilities, such algorithms can provide adaptive and efficient solutions to the discussed system models, handling the complexity introduced by adding more variables and adapting to time-varying environments.

7. Future work for the thesis includes exploring the potential benefits and performance improvements offered by the rate-splitting multiple access techniques compared to NOMA. The thesis has already investigated the GSVD-based MIMO scheme, which decomposes the channel into private and common components. This scheme can be extended to the rate-splitting technique, where the transmit signal incorporates common and private components for each user. This observation suggests a valuable opportunity to transform the proposed studies on NOMA-based systems into rate-splitting-based studies. By analyzing and evaluating the performance of the rate-splitting technique, a comprehensive understanding of its advantages over NOMA can be obtained.

This thesis makes significant contributions to the field of wireless communication systems by integrating MIMO and NOMA technologies through the utilization of the GSVD-based beamforming technique. The proposed solutions have successfully showcased advancements in various key aspects, including reduced offloading delay, enhanced energy efficiency, improved security, and higher achievable data rates. These findings provide a deeper understanding of the potential benefits of MIMO-NOMA integration and pave the way for future research directions. Areas worth exploring include investigating the effects of imperfect CSI, exploring the multi-linear-GSVD approach, combining the proposed techniques with emerging technologies such as intelligent reflecting surfaces and unmanned aerial vehicles, optimizing adaptive SIC ordering strategies, and exploring the transformation of the proposed solutions into GSVD-based RSMA-MIMO networks. These future research avenues hold promise for further improving the performance and capabilities of wireless communication systems.

References

[1] S. A. A. Hakeem, H. H. Hussein, and H. Kim, "Vision and research directions of 6G technologies and applications," *Journal of King Saud University-Computer and Information Sciences*, 2022 (cited on p. 16).

[2] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Transactions on Communications*, vol. 98, no. 3, pp. 403–414, 2015 (cited on p. 17).

[3] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of noma: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, pp. 179–189, 2020. DOI: 10.1109/OJCOMS.2020.2969899 (cited on pp. 17, 23).

[4] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5191–5202, 2017 (cited on pp. 17, 68).

[5] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, "Delay Minimization for NOMA-MEC Offloading," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1875–1879, 2018 (cited on pp. 17, 52–54).

[6] W. Zhao, B. Wang, H. Bao, and B. Li, "Secure Energy-Saving Resource Allocation on Massive MIMO-MEC System," *IEEE Access*, vol. 8, pp. 137 244–137 253, 2020 (cited on pp. 17, 52).

[7] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "Scma codebook design," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5. DOI: 10.1109/VTCFall.2014.6966170 (cited on p. 23).

[8] J. Zeng, B. Li, X. Su, L. Rong, and R. Xing, "Pattern division multiple access (PDMA) for cellular future radio access," in *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, 2015, pp. 1–5. DOI: 10.1109/WCSP.2015.7341229 (cited on p. 23).

[9] Y. Cao, H. Sun, J. Soriaga, and T. Ji, "Resource spread multiple access - a novel transmission scheme for 5G uplink," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5. DOI: 10.1109/VTCFall.2017.8288412 (cited on p. 23).

[10] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5. DOI: 10.1109/VTCSpring.2016.7504361 (cited on p. 23).

[11] S. Hu, B. Yu, C. Qian, *et al.*, "Nonorthogonal interleave-grid multiple access scheme for industrial internet of things in 5G network," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5436–5446, 2018. DOI: 10.1109/TII.2018.2858142 (cited on p. 23).

[12] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising physical layer strategy for LTE evolution," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 98–105, 2016. DOI: 10.1109/MCOM.2016.7470942 (cited on p. 23).

[13] A. Benjebbour, "An overview of non-orthogonal multiple access," *ZTE Communications*, vol. 15, no. S1, pp. 21–30, 2017. DOI: 10.3969/j.issn.1673-5188.2017.S1.003 (cited on pp. 23, 26, 28).

[14] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless Communications and Mobile Computing*, pp. 1–24, 2018. DOI: 10.1155/2018/9713450 (cited on pp. 24, 27).

[15] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the Importance of SIC in NOMA Systems—Part 1: State of the Art and Recent Findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373–2377, 2020. DOI: 10.1109/LCOMM.2020.3012604 (cited on pp. 24, 125).

[16] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9013344 (cited on p. 27).

[17] C. C. Paige, "Computing the generalized singular value decomposition," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1126–1146, 1986. DOI: 10.1137/0907077 (cited on p. 29).

[18] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002. DOI: 10.1109/TSP.2002.801937 (cited on p. 29).

[19] O. Alter, P. O. Brown, and D. Botstein, "Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms," *Proceedings of the National Academy of Sciences*, vol. 100, no. 6, pp. 3351–3356, 2003 (cited on p. 29).

[20] C. Rao, Z. Ding, and X. Dai, "GSVD-based MIMO-noma security transmission," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1484–1487, 2021. DOI: 10.1109/LWC.2021.3071365 (cited on p. 30).

[21] S. A. A. Fakoorian and A. L. Swindlehurst, "Dirty paper coding versus linear GSVD-Based Precoding in MIMO broadcast channel with confidential messages," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011, pp. 1–5. DOI: 10.1109/GLOCOM.2011.6134129 (cited on p. 30).

[22] W. Mei, Z. Chen, and J. Fang, "GSVD-based precoding in MIMO systems with integrated services," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1528–1532, 2016 (cited on pp. 30, 67, 75).

[23] C. Rao, Z. Ding, and X. Dai, "GSVD-based precoding in MIMO systems with integrated services," *IET Communications*, vol. 14, no. 21, pp. 3802–3812, 2020. DOI: 10.1049/iet-com.2020.0555 (cited on p. 30).

[24] M. F. Hanif and Z. Ding, "Robust power allocation in MIMO-NOMA systems," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1541–1545, 2019 (cited on p. 30).

[25] Y. Dursun, F. Fang, and Z. Ding, "Hybrid NOMA based MIMO offloading for mobile edge computing in 6G networks," *China Communications*, vol. 19, no. 10, pp. 12–20, 2022 (cited on pp. 30, 84, 86, 103, 109).

[26] Y. Dursun, M. B. Goktas, and Z. Ding, "Green NOMA based MU-MIMO transmission for mec in 6G networks," *Computer Networks*, vol. 228, p. 109 749, 2023, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2023.109749 (cited on pp. 30, 39, 118).

[27] Y. Dursun, K. Wang, and Z. Ding, "Secrecy sum rate maximization for a MIMO-NOMA uplink transmission in 6G networks," *Physical Communication*, vol. 53, p. 101 675, 2022 (cited on pp. 30, 39, 85, 88, 103).

[28] Z. Chen, Z. Ding, X. Dai, and R. Schober, "Asymptotic performance analysis of GSVD-NOMA systems with a large-scale antenna array," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 575–590, 2019. DOI: 10.1109/TWC.2018.2883102 (cited on pp. 29, 33, 109).

[29] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, "Optimal Resource Allocation for Delay Minimization in NOMA-MEC Networks," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7867–7881, 2020. DOI: 10.1109/TCOMM.2020.3020068 (cited on pp. 36, 39).

[30] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Latency minimization for IRS-Aided NOMA MEC systems with WPT-enabled IoT devices," *IEEE Internet of Things Journal*, pp. 1–1, 2023. DOI: 10.1109/JIOT.2023.3240395 (cited on pp. 36, 39).

[31] S. S. Yılmaz and B. Özbek, "Massive MIMO-NOMA Based MEC in Task Offloading for Delay Minimization," *IEEE Access*, vol. 11, pp. 162–170, 2023. DOI: 10.1109/ACCESS.2022.3232731 (cited on p. 39).

[32] Y. Zhang, Z. Na, Y. Wang, and C. Ji, "Joint power allocation and deployment optimization for HAP-assisted NOMA-MEC system," *Wireless Networks*, pp. 1–13, 2022 (cited on p. 39).

[33] F. Guo, H. Zhang, H. Ji, X. Li, and V. C. Leung, "Joint Trajectory and Computation Offloading Optimization for UAV-assisted MEC with NOMA," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 1–6. DOI: 10.1109/INFOCOMWKSHPS47286.2019.9093764 (cited on p. 39).

[34] C. Zheng and W. Zhou, "Computation bits maximization in backscatter-assisted wireless-powered NOMA-MEC networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 23, 2022 (cited on p. 39).

[35] Y. Xu, T. Zhang, Y. Zou, and Y. Liu, "Reconfigurable Intelligence Surface Aided UAV-MEC Systems With NOMA," *IEEE Communications Letters*, vol. 26, no. 9, pp. 2121–2125, 2022. DOI: 10.1109/LCOMM.2022.3183285 (cited on p. 39).

[36] I. Altin and M. Akar, "A joint resource allocation method for hybrid NOMA MEC offloading," *Physical Communication*, vol. 54, p. 101 809, 2022 (cited on p. 39).

[37] L. Lin, W. Zhou, Z. Yang, and J. Liu, "Deep reinforcement learning-based task scheduling and resource allocation for NOMA-MEC in industrial internet of things," *Peer-to-Peer Networking and Applications*, vol. 16, no. 1, pp. 170–188, 2023 (cited on p. 39).

[38] C. Li, H. Wang, and R. Song, "Mobility-aware offloading and resource allocation in NOMA-MEC systems via DC," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1091–1095, 2022. DOI: 10.1109/LCOMM.2022.3154434 (cited on p. 39).

[39] T. Van Truong and A. Nayyar, "System performance and optimization in NOMA mobile edge computing surveillance network using GA and PSO," *Computer Networks*, p. 109 575, 2023 (cited on p. 39).

[40] M. Chen, Y. Wan, M. Wen, and T. Zhou, "Fairness optimization in IRS-assisted MEC computational offloading," *Physical Communication*, vol. 54, p. 101 855, 2022 (cited on p. 39).

[41] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967 (cited on pp. 44, 53).

[42] K. Shen and W. Yu, "Fractional programming for communication systems—part i: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018 (cited on pp. 44, 45).

[43] K. B. Petersen, M. S. Pedersen, *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, no. 15, p. 510, 2008 (cited on pp. 45, 113).

[44] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004 (cited on p. 47).

[45] Z. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010. DOI: 10.1109/MSP.2010.936019 (cited on p. 48).

[46] M. Grant and S. Boyd, *Cvx: Matlab Software for Disciplined Convex Programming, version 2.1*, 2014 (cited on pp. 49, 57).

[47] Z. Ding, F. Adachi, and H. V. Poor, "The Application of MIMO to Non-Orthogonal Multiple Access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, 2015 (cited on p. 52).

[48] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G systems: Potentials and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2016 (cited on pp. 52, 57).

[49] Y. Wang, R. Chen, and D.-C. Wang, "A survey of Mobile Cloud Computing Applications: Perspectives and Challenges," *Wireless Personal Communications*, vol. 80, no. 4, pp. 1607–1623, 2015 (cited on p. 52).

[50] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint Optimization of Radio and Computational Resources for Multicell Mobile-Edge Computing," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89–103, 2015 (cited on p. 52).

[51] J. Wang, B. Liu, and L. Feng, "Secure MISO Cognitive-based Mobile Edge Computing with Wireless Power Transfer," *IEEE Access*, vol. 8, pp. 15 518–15 528, 2020 (cited on p. 52).

[52] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and J. Wang, "Joint MU-MIMO Precoding and Resource Allocation for Mobile-Edge Computing," *IEEE Transactions on Wireless Communications*, 2020 (cited on p. 52).

[53] J. Yan, S. Bi, Y. J. Zhang, and M. Tao, "Optimal Task Offloading and Resource Allocation in Mobile-Edge Computing with Inter-User Task Dependency," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 235–250, 2019 (cited on p. 52).

[54] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, "Optimal Resource Allocation for Delay Minimization in NOMA-MEC Networks," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7867–7881, 2020 (cited on p. 52).

[55] F. Fang, K. Wang, Z. Ding, and V. C. Leung, "Energy-Efficient Resource Allocation for NOMA-MEC Networks with Imperfect CSI," *IEEE Transactions on Communications*, 2021 (cited on p. 52).

[56] A. Kiani and N. Ansari, "Edge Computing Aware NOMA for 5G Networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018 (cited on pp. 52, 53).

[57] H. Li, F. Fang, and Z. Ding, "Joint Resource Allocation for Hybrid NOMA-assisted MEC in 6G Networks," *Digital Communications and Networks*, vol. 6, no. 3, pp. 241–252, 2020 (cited on p. 53).

[58] Z. Ma, Z. Ding, P. Fan, and S. Tang, "A General Framework for MIMO Uplink and Downlink Transmissions in 5G Multiple Access," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*, IEEE, 2016, pp. 1–4 (cited on p. 53).

[59] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017 (cited on p. 53).

[60] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the Importance of SIC in NOMA Systems—Part 1: State of the Art and Recent Findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373–2377, 2020 (cited on p. 56).

[61] Y. Zhang, Y. Wang, and W. Zhang, "Energy Efficient Resource Allocation for Heterogeneous Cloud Radio Access Networks with User Cooperation and QoS guarantees," in *2016 IEEE Wireless Communications and Networking Conference*, IEEE, 2016, pp. 1–6 (cited on p. 60).

[62] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018 (cited on pp. 68, 78).

[63] Y. Liu, G. Pan, H. Zhang, and M. Song, "On the capacity comparison between MIMO-NOMA and MIMO-OMA," *IEEE Access*, vol. 4, pp. 2123–2129, 2016 (cited on pp. 68, 78).

[64] F. Kara, "Error performance of cooperative relaying systems empowered by SWIPT and NOMA," *Physical Communication*, vol. 49, p. 101 450, 2021 (cited on p. 68).

[65] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Communications Letters*, vol. 20, no. 5, pp. 930–933, 2016 (cited on p. 68).

[66] F. Jia, C. Zhang, C. Jiang, M. Li, and J. Ge, "Guaranteeing positive secrecy rate for NOMA system against internal eavesdropping," *IEEE Communications Letters*, vol. 25, no. 6, pp. 1805–1809, 2021 (cited on pp. 68, 77).

[67] Y. Cao, N. Zhao, Y. Chen, *et al.*, "Secure transmission via beamforming optimization for NOMA networks," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 193–199, 2019 (cited on p. 68).

[68] C. Yin and L. Yan, "Secure beamforming design for the UAV-enabled transmission over NOMA networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1–11, 2020 (cited on p. 68).

[69] Y. Deng, Q. Li, Q. Zhang, L. Yang, and J. Qin, "Secure beamforming design in MIMO NOMA networks for internet of things with perfect and imperfect CSI," *Computer Networks*, vol. 187, p. 107 839, 2021 (cited on pp. 68, 79).

[70] F. Zhou, Z. Chu, Y. Wu, N. Al-Dhahir, and P. Xiao, "Enhancing PHY security of MISO NOMA SWIPT systems with a practical non-linear EH model," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, 2018, pp. 1–6 (cited on p. 68).

[71] N. Nandan, S. Majhi, and H.-C. Wu, "Beamforming and power optimization for physical layer security of MIMO-NOMA based CRN over imperfect CSI," *IEEE*

Transactions on Vehicular Technology, vol. 70, no. 6, pp. 5990–6001, 2021 (cited on p. 68).

[72] M. Zeng, N.-P. Nguyen, O. A. Dobre, and H. V. Poor, "Securing downlink massive MIMO-NOMA networks with artificial noise," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 685–699, 2019 (cited on p. 68).

[73] M. Tian, Q. Zhang, S. Zhao, Q. Li, and J. Qin, "Secrecy sum rate optimization for downlink MIMO nonorthogonal multiple access systems," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1113–1117, 2017 (cited on p. 68).

[74] K. Jiang, T. Jing, Y. Huo, F. Zhang, and Z. Li, "SIC-based secrecy performance in uplink NOMA multi-eavesdropper wiretap channels," *IEEE Access*, vol. 6, pp. 19 664–19 680, 2018 (cited on p. 69).

[75] J. Steinwandt, S. A. Vorobyov, and M. Haardt, "Secrecy rate maximization for MIMO gaussian wiretap channels with multiple eavesdroppers via alternating matrix POTDC," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 5686–5690 (cited on pp. 69, 80).

[76] G. Gomez, F. J. Martin-Vega, F. J. Lopez-Martinez, Y. Liu, and M. Elkashlan, "Physical layer security in uplink NOMA multi-antenna systems with randomly distributed eavesdroppers," *IEEE Access*, vol. 7, pp. 70 422–70 435, 2019 (cited on p. 69).

[77] K. Jiang, W. Zhou, and L. Sun, "Jamming-aided secrecy performance in secure uplink NOMA system," *IEEE access*, vol. 8, pp. 15 072–15 084, 2020 (cited on pp. 69, 95).

[78] M. Vaezi, W. Shin, and H. V. Poor, "Optimal beamforming for gaussian MIMO wiretap channels with two transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6726–6735, 2017 (cited on p. 69).

[79] C. Rao, Z. Ding, and X. Dai, "GSVD-based MIMO-NOMA security transmission," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1484–1487, 2021 (cited on pp. 69, 72).

[80] S. Zhao, J. Liu, Y. Shen, X. Jiang, and N. Shiratori, "Secure and energy-efficient precoding for MIMO two-way untrusted relay systems," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3371–3386, 2021 (cited on p. 69).

[81] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Transactions on wireless communications*, vol. 7, no. 6, pp. 2180–2189, 2008 (cited on pp. 70, 71).

[82] M. F. Hanif and Z. Ding, "Robust power allocation in MIMO-NOMA systems," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1541–1545, 2019 (cited on p. 72).

[83] N. Vucic, S. Shi, and M. Schubert, "Dc programming approach for resource allocation in wireless networks," in *8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, IEEE, 2010, pp. 380–386 (cited on p. 75).

[84] Y. Qi and M. Vaezi, "Secure transmission in MIMO-NOMA networks," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2696–2700, 2020 (cited on p. 77).

[85] Y. Yuan, Z. Yuan, G. Yu, *et al.*, "Non-orthogonal transmission technology in LTE evolution," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 68–74, 2016 (cited on p. 77).

[86] Z. Chen, Z. Ding, X. Dai, and R. Schober, "Asymptotic performance analysis of GSVD-NOMA systems with a large-scale antenna array," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 575–590, 2018 (cited on pp. 78, 84, 88, 95).

[87] W. U. Khan, F. Jameel, T. Ristaniemi, S. Khan, G. A. S. Sidhu, and J. Liu, "Joint spectral and energy efficiency optimization for downlink NOMA networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 645–656, 2019 (cited on p. 78).

[88] M. Mollanoori and M. Ghaderi, "Uplink scheduling in wireless networks with successive interference cancellation," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1132–1144, 2013 (cited on p. 78).

[89] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2017. DOI: 10.1109/COMST.2016.2621116 (cited on p. 84).

[90] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018. DOI: 10.1109/COMST.2018.2835558 (cited on p. 84).

[91] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017 (cited on p. 85).

[92] B. Liu, C. Liu, and M. Peng, "Resource allocation for energy-efficient MEC in NOMA-enabled massive iot networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1015–1027, 2020 (cited on p. 85).

[93] B. Su, Q. Ni, W. Yu, and H. Pervaiz, "Optimizing computation efficiency for NOMA-assisted mobile edge computing with user cooperation," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 858–867, 2021 (cited on p. 85).

[94] L. Lin, J. Liu, D. Zhang, Y. Xie, *et al.*, "Joint offloading decision and resource allocation for multiuser NOMA-MEC systems," *IEEE Access*, vol. 7, pp. 181 100–181 116, 2019 (cited on p. 85).

[95] M. Sheng, Y. Dai, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Delay-aware computation offloading in NOMA MEC under differentiated uploading delay," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2813–2826, 2020 (cited on pp. 85, 92).

[96] F. Fang, K. Wang, Z. Ding, and V. C. Leung, "Energy-efficient resource allocation for NOMA-MEC networks with imperfect CSI," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3436–3449, 2021 (cited on pp. 85, 89, 91).

[97] N. A. Mitsiou, P. N. Gavriilidis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless powered multi-access edge computing with slotted ALOHA," *IEEE Communications Letters*, 2022 (cited on p. 85).

[98] F. Binucci, P. Banelli, P. Di Lorenzo, and S. Barbarossa, "Dynamic resource allocation for multi-user goal-oriented communications at the wireless edge," in *2022 30th European Signal Processing Conference (EUSIPCO)*, IEEE, 2022, pp. 697–701 (cited on p. 85).

[99] M. Merluzzi, N. di Pietro, P. Di Lorenzo, E. C. Strinati, and S. Barbarossa, "Discontinuous computation offloading for energy-efficient mobile edge computing," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 1242–1257, 2021 (cited on p. 86).

[100] G. Niu, Q. Cao, and M.-O. Pun, "Qos-aware resource allocation for mobile edge networks: User association, precoding and power allocation," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 617–12 630, 2021 (cited on p. 86).

[101] E. Vlachos, J. Thompson, A. Kaushik, and C. Masouros, "Radio-frequency chain selection for energy and spectral efficiency maximization in hybrid beamform-

ing under hardware imperfections," *Proceedings of the Royal Society A*, vol. 476, no. 2244, p. 20 200 451, 2020 (cited on p. 90).

[102] A. Kaushik, J. Thompson, E. Vlachos, C. Tsinos, and S. Chatzinotas, "Dynamic RF chain selection for energy efficient and low complexity hybrid beamforming in millimeter wave MIMO systems," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 886–900, 2019 (cited on p. 90).

[103] Y. Pei, T.-H. Pham, and Y.-C. Liang, "How many RF chains are optimal for large-scale MIMO systems when circuit power is considered?" In *2012 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2012, pp. 3868–3873 (cited on p. 90).

[104] B. Rong, "6G: The Next Horizon: From connected people and things to connected intelligence," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 8–8, 2021. DOI: 10.1109/MWC.2021.9615100 (cited on p. 103).

[105] Y. Liu, S. Zhang, X. Mu, *et al.*, "Evolution of NOMA toward next generation multiple access (NGMA) for 6G," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1037–1071, 2022. DOI: 10.1109/JSAC.2022.3145234 (cited on p. 103).

[106] H. Semira, F. Kara, H. Kaya, and H. Yanikomeroglu, "Multi-user joint maximum-likelihood detection in uplink NOMA-IoT networks: Removing the error floor," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2459–2463, 2021 (cited on p. 103).

[107] M. B. Goktas and Z. Ding, "A wireless power transfer assisted NOMA transmission scheme for 5G and beyond mMTC," *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1239–1242, 2022 (cited on p. 103).

[108] Y. Liu, G. Pan, H. Zhang, and M. Song, "On the capacity comparison between MIMO-NOMA and MIMO-OMA," *IEEE Access*, vol. 4, pp. 2123–2129, 2016 (cited on p. 103).

[109] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 174–180, 2019 (cited on p. 103).

[110] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009 (cited on p. 104).

[111] S. Bhattacharjee, T. Acharya, and U. Bhattacharya, "Cognitive radio based spectrum sharing models for multicasting in 5G cellular networks: A survey," *Computer Networks*, vol. 208, p. 108 870, 2022 (cited on p. 104).

[112] A. Srivastava and G. Kaur, "CEAR: A cooperation based energy aware reward scheme for next generation green cognitive radio networks," *Physical Communication*, vol. 56, p. 101 947, 2023 (cited on p. 104).

[113] Y. Zheng, X. Li, H. Zhang, *et al.*, "Overlay cognitive ABCom-NOMA-based ITS: An in-depth secrecy analysis," *IEEE Transactions on Intelligent Transportation Systems*, 2022 (cited on p. 104).

[114] Y. Wu, F. Zhou, W. Wu, Q. Wu, R. Q. Hu, and K.-K. Wong, "Multi-objective optimization for spectrum and energy efficiency tradeoff in IRS-assisted CRNs with NOMA," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6627–6642, 2022 (cited on p. 104).

[115] S. K. Nobar, M. H. Ahmed, Y. Morgan, and S. A. Mahmoud, "Resource allocation in cognitive radio-enabled UAV communication," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 296–310, 2021 (cited on p. 104).

[116] W. Feng, J. Tang, N. Zhao, *et al.*, "NOMA-based UAV-aided networks for emergency communications," *China Communications*, vol. 17, no. 11, pp. 54–66, 2020 (cited on p. 104).

[117] D.-T. Do, A.-T. Le, and B. M. Lee, "Noma in cooperative underlay cognitive radio networks under imperfect SIC," *IEEE Access*, vol. 8, pp. 86 180–86 195, 2020. DOI: 10.1109/ACCESS.2020.2992660 (cited on p. 104).

[118] S. Mao, S. Leng, J. Hu, and K. Yang, "Power minimization resource allocation for underlay MISO-NOMA SWIPT systems," *IEEE Access*, vol. 7, pp. 17 247–17 255, 2019 (cited on p. 104).

[119] K. Xu, M. Zhang, J. Liu, N. Sha, W. Xie, and L. Chen, "SWIPT in mMIMO system with non-linear energy-harvesting terminals: Protocol design and performance optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, pp. 1–15, 2019 (cited on p. 110).

[120] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Energy-efficient design for irs-empowered uplink MIMO-NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9490–9500, 2022 (cited on p. 113).

[121] C.-L. Wang, Y.-C. Ding, Y.-C. Wang, and P. Xiao, "A low-complexity power allocation scheme for MIMO-noma systems with imperfect channel estimation," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2022, pp. 234–239. DOI: 10.1109/PIMRC54779.2022.9977487 (cited on p. 125).

[122] L. Khamidullina, A. L. F. de Almeida, and M. Haardt, "Multilinear generalized singular value decomposition (ml-GSVD) and its application to multiuser MIMO systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2783–2797, 2022. DOI: 10.1109/TSP.2022.3178902 (cited on p. 125).

[123] L. Khamidullina, A. L. F. de Almeida, and M. Haardt, "Rate splitting and precoding strategies for multi-user MIMO broadcast channels with common and private streams," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10095138 (cited on p. 125).