Resource Allocation Techniques for Non-Orthogonal Multiple Access in Beyond 5G

Xinchen Wei

Doctor of Philosophy

University of York
Electronic Engineering

September 2022

I dedicate this thesis to my mom.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Some of the work presented in this thesis have been published in, or planned to submit to conferences, and journals, which are listed as follows:

1. **X. Wei**, H. Al-Obiedollah, K. Cumanan, M. Zhang, J. Tang, W. Wang, and O. A. Dobre, "Resource Allocation Technique for Hybrid TDMA-NOMA System with Opportunistic Time Assignment," in *Proc. IEEE ICC Workshops*, 2020, pp. 1-6.

2. **X. Wei**, H. Al-Obiedollah, K. Cumanan, W. Wang, Z. Ding and O. A. Dobre, "Spectral-Energy Efficiency Trade-Off Based Design for Hybrid TDMA-NOMA System," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 3377-3382, March 2022.

3. **X. Wei**, H. Al-Obiedollah, K. Cumanan, Z. Ding and O. A. Dobre, "Energy Efficiency Maximization for Hybrid TDMA-NOMA System With Opportunistic Time Assignment," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8561-8573, Aug. 2022.

4. **X. Wei**, K. Cumanan, H. Al-Obiedollah, G. Chen, Z. Ding and O. A. Dobre, "Energy Efficiency Maximization for IRS-Assisted NOMA System with Imperfect CSI," submitted in *IEEE Transactions on Vehicular Technology*.

5. J. Liu, F. Zeng, W. Wang, Z. Sheng, **X. Wei** and K. Cumanan, "Trajectory Design for UAV-enabled Maritime Secure Communications: A Reinforcement Learning Approach," *China Communications*, vol. 19, no. 9, pp. 26-36, Sept. 2022.

Xinchen Wei
September 2022

Acknowledgements

Foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Kanapathippillai Cumanan, for his continuous support, encouragement, and unbounded enthusiasm in my Ph.D research. Without his guidance and constant feedback, this Ph.D would not have been achievable. It is my distinct honour to have such a gracious supervisor during my PhD journey. This gratitude also goes to my research collaborators: Prof. Zhiguo Ding and Prof. Octavia A. Dobre, for their kind support and helpful comments on my research. I would like to thank Haitham Al-Obiedollah for his invaluable advice and feedback on my research in the first year of my Ph.D and for always being so supportive of my work.

I am also grateful to Jane Irisa and Francoise Siganos for their friendship and the warmth they extended to me during my time in the villages and for always making me feel so welcome. Very special thanks go to Libby Redman for being hospitality during the Christmas festival. I would also like to thank Xiaoxuan Wang, Qiming Ni, Hangyu Zhou, Yi Wang, and Xinchen Xu for their kind help and for always being such a friend ready to listen.

Last and most importantly, I would like to extend my heartfelt gratitude to my beloved family, especially my mom, who always believed in me and encouraged me to follow my dreams at any time.

Abstract

To support the wide range of envisioned applications, including autonomous vehicles, augmented reality, holographic communication, and Internet of Everything (IoE), future wireless networks must meet demanding requirements for higher spectral and energy efficiency, lower end-to-end latency and massive connectivity. This requires a vast upgrade in the technologies of the sixth-generation (6G) wireless networks. Non-orthogonal multiple access (NOMA) has been advocated as a prospective effective multiple access technique for future wireless networks due to the wide range of its potential benefits, including superior spectral efficiency (SE), energy efficiency (EE), compatibility, user fairness, and flexibility. To exploit additional degrees of freedom and address the computational complexity with massive connectivity, NOMA has been recently combined with different types of multiple access techniques and appropriate optimization designs. Hence, this thesis attempts to utilize the combination of NOMA with different key technologies, including multiple antenna techniques, conventional OMA techniques, and intelligent reflecting surface (IRS). In particular, different resource allocation techniques have been developed for such integrated NOMA systems, from the downlink (DL) single-input single-output (SISO)-NOMA system, to DL multiple-input single-output (MISO)-NOMA system, as well as the IRS-assisted NOMA system.

Firstly, a hybrid time division multiple access (TDMA)-NOMA system is considered, where both the available time slots and the available transmit power are jointly allocated to maximize the global EE. To further exploit the promising advantages of this hybrid system, the SE-EE trade-off based design and max-min fairness based design are presented in this thesis. By utilizing different convex relaxation and approximation techniques, the non-convexity of the formulated optimization problems are transformed into convex problems. Finally, this thesis investigates a worst-case robust design for an IRS-assisted NOMA multi-user MISO system to maximize the EE with a set of quality of service (QoS) constraints. In particular, an iterative algorithm based on alternating optimization (AO) is proposed to design the transmit beamforming vectors at the base station (BS) and reflection coefficient matrix for IRS. The effectiveness advantages of all the proposed schemes are demonstrated through numerical simulation results.

Table of contents

List of figures

List of tables

List of symbols

$\mathbf{E}\{\cdot\}$	Statistical Expectation
$\mathbb{C}^N$	N-dimensional complex vectors
$\mathbb{R}^N$	N-dimensional real vectors
$\|x\|$	Norm of complex number x
$\mathbf{x}$	Vector $\mathbf{x}$
$\mathbf{X}$	Matrix $\mathbf{X}$
$(\cdot)^T$	Transpose
$(\cdot)^H$	Hermitian Transpose
$\mathrm{tr}(\mathbf{X})$	Trace of the matrix $\mathbf{X}$
$\mathbf{x} \succeq 0$	Each element in $\mathbf{x}$ is greater than zero
$\mathbf{X} \succeq 0$	Positive semi-definite matrix
$\|\| \cdot \|\|_2$	Euclidian vector norm
$\min\{\}$	Minimum of function
$\max\{\}$	Maximum of function
$\mathcal{CN}(\mu, \sigma^2)$	Complex gaussian random variable with mean μ and variance σ^2
$\mathbf{dom} f$	Domain of a function f
∇f	First derivative of
$\nabla^2 f$	Second derivative of

List of abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Project Partnership
4G	Forth Generation
5G	Fifth Generation
6G	Six Generation
AMPS	Advanced Mobile Phone System
AI	Artificial Intelligence
AO	Alternating Optimization
AoI	Age of Information
AR	Augmented Reality
AWGN	Additive White Gaussian Noise
BS	Base Station
CAeC	Contextually Agile eMBB Communications
CDMA	Code Division Multiple Access
COC	Computation Oriented Communications
CSI	Channel State Information
DL	Downlink
EDuRLLC	Event Defined URLLC
EE	Energy Efficiency
EE-Max	EE maximization
eMBB	Enhanced Mobile Broadband
FDMA	Frequency Division Multiple Access
GEE-Max	Global Energy Efficiency Maximization
GPRS	General Packet Radio Services
HCS	Harmonized Communication and Sensing
HSDPA	High-Speed Downlink Packet Access

IMT	International Mobile Telecommunications
IoE	Internet of Everything
IoTs	Internet of Things
IRS	Intelligent Reflecting Surfaces
ISAC	Integrated Sensing and Communications
LDS-CDMA	Low-Density Spreading CDMA
LDS-OFDM	Low-Density Spreading-based Orthogonal Frequency Division Multiplexing
LMIs	Linear Matrix Inequalities
LP	Linear Programming
LTE	Long-Term Evolution
M2M	Machine-to-Machine
MEC	Mobile Edge Computing
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
mMTC	Massive Machine Type Communications
mmWave	Millimeter-Wave
MOO	Multi-Objective Optimization
MUSA	Multi-User Shared Access
NMT	Nordic Mobile Telephone
NTT	Nippon Telephone and Telegraph
NOMA	Non-Orthogonal Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PDMA	Pattern Division Multiple Access
QCQP	Quadratic Constrained Quadratic Programming
QoS	Quality of Service
QP	Quadratic Programming
RTBC	Real-Time Broadband Communication
SC	Superposition Coding
SCA	Sequential Convex Approximations
SCMA	Sparse Code Multiple Access
SDMA	Space Division Multiple Access
SDP	Semidefinite Programming
SDU	Service Data Unit
SE	Spectral Efficiency

SIC	Successive Interference Cancellation
SINR	Signal-to-Interference and Noise Ratio
SISO	Single-Input Single-Output
SOC	Second-Order Cone
SOCP	Second-order Cone Programming
SOO	Single-objective Optimization
SWIPT	Simultaneous Wireless Information and Power Transfer
TDD	Time-Division Duplex
TDMA	Time Division Multiple Access
THz	Terahertz
UCBC	UL Centric Broadband Communication
UL	Uplink
URLLC	Ultra-Reliable and Low Latency Communications
VR	Virtual Reality
WCDMA	Wideband Code-Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access

Chapter 1

Introduction

1.1 Overview

In the last few decades, mobile wireless communications have been changing enormously, which is considered upgrading to a new generation every ten years [4]. The applications of wireless communications are growing exponentially. They include augmented reality (AR), autonomous vehicles, virtual reality (VR), mixed-reality (MR), and Internet of Things (IoTs) [4, 5]. In 1979, the first cellular system in the world was launched by Nippon Telephone and Telegraph (NTT) in Tokyo, Japan. Based on the the launch of Nordic Mobile Telephone(NMT) system and the Advanced Mobile Phone System (AMPS), the first generation (1G) mobile system was introduced and used analog transmission for voice services [6]. This analog system employed a multiple access method called frequency division multiple access (FDMA). In FDMA, the available channel bandwidth is divided into many sub-frequency bands, where each band is dynamically assigned to a specific user to access the network. However, there was only limited bandwidth available, which means that only a very small group of users can communicate simultaneously with a base station (BS).

Compared to 1G system with analog technology, the second generation (2G) system used digital signals to provide better quality and capacity, which enabled users to deliver text messages, picture messages at a low speed [7]. In particular, the two important multiple access techniques, such as time division multiple access (TDMA) and code division multiple access (CDMA) were utilized to develop 2G systems. Before moving to the next generation, 2.5G was introduced by combining 2G cellular system and General Packet Radio Services (GPRS) [6]. Due to the demands of a higher number of users, the requirements of high speed data rate and multimedia connectivity, the mobile communication industries started working on a system that can provide

better services, leading to the emerging of the third generation (3G) cellular system in late 1980s. In such 3G system, one key feature was support of the air-interface technology, that is, Wideband code-division multiple access (WCDMA), which included high-speed downlink (DL) packet access (HSDPA), uplink (UL) high-speed data, UL high-speed access for time-division duplex (TDD) [8]. The development of the fourth generation (4G) services in 2010 was a major step forward both in terms of the quality of service (QoS) and data rate requirements for applications, as well as providing far more rapid access to the Internet due to its reliability and capability of delivering larger amounts of data. The goal of 4G is to provide wireless services in anywhere and at anytime [6]. 4G gives more integrity through orthogonal frequency division multiple access (OFDMA) with Long-Term Evolution (LTE)-Advanced, third generation project partnership (3GPP) and mobile Worldwide Interoperability for Microwave Access (WiMAX). Moreover, it is capable of reaching up between 100 Mbit/s and 1 Gbit/s speeds in both indoors and outdoors environments, while satisfying premium quality and high security [9].

As of today, the fifth generation (5G) mobile communication system has been rolled out in many countries and brought 1000 times of the capability compared to that of the 4G mobile communication systems [10]. For example, 5G has been deployed in 50 cities in the United States at the end of 2019. The number of mobile base stations providing 5G services has been to around 3,000 across the UK at the end of 2020. The Chinese mainland currently has 1.425 million installed 5G base stations that support more than 500 million 5G subscribers [11–13]. Specifically, 5G is able to achieve higher spectral efficiency (SE), energy efficiency (EE), and massive connectivity as well as lower cost and energy consumption [14]. Meanwhile, driven by the exponential growth of mobile traffic and newly emerging use-cases and applications, the development of a next-generation system, the sixth generation (6G), is expected to achieve performance superior to 5G and meet the demanding requirements of wireless networks in 2030 [1].

1.2 Towards 5G and Beyond

1.2.1 Requirements of 5G and Beyond

Based on the framework of International Mobile Telecommunications (IMT) for 2020 and beyond, an overview of the main features of 5G and 6G is shown in Fig. 1.1. There are eight key performance metrics for evaluating the 5G and 6G networks, including peak data rate, user-experienced data rate, SE, mobility, latency, connectivity density,

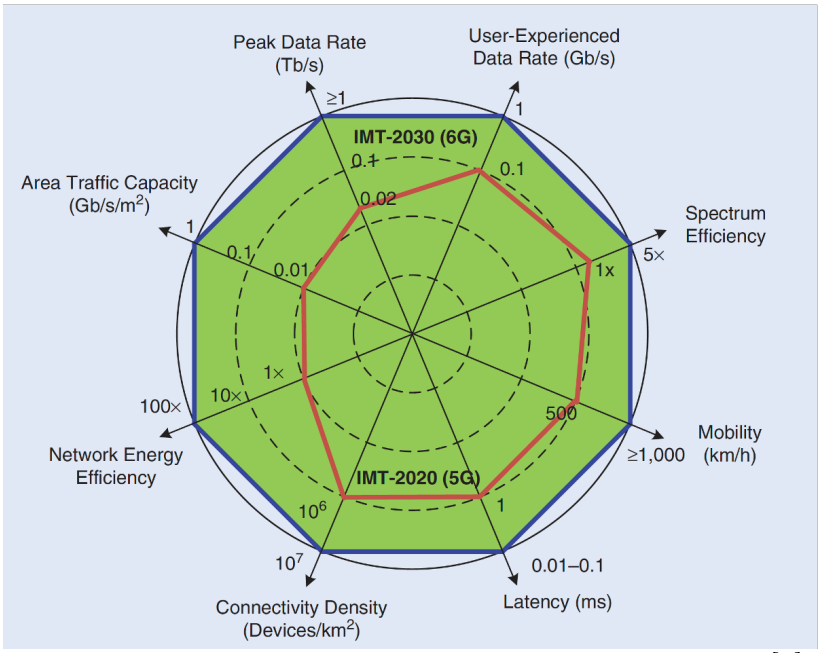

Fig. 1.1 Key requirements of 6G networks, copied from [1].

EE, and area traffic capacity [2]. It can be seen that the vertices of the inner polygon, (in red lines), represent the requirements of 5G, and the vertices of the outer octagon, (in blue lines), represent the requirements of 6G. In the following, the details of eight key performance metrics are briefly presented.

- Peak data rate

  Peak data rate is the highest achievable data rate under ideal conditions. It is expected to support up to 20 Gbit/s in DL and 10 Gbps for UL in 5G, while the peak rate of 6G is envisioned to be 1 Tbps that is 50 times that of 5G [15]. This design requirement should consider supporting the much higher data rate services, such as the extended reality and the holographic communications [16].

- User-experienced data rate

  User-experienced data rate is defined as the 5% point of the cumulative distribution function of the user throughput. In particular, individual users can get at least this minimum achievable data rate at anytime and in anywhere with a possibility of 95% [17]. 5G should be able to provide 100Mbps for DL and 50Mbps for UL. In 6G, the requirement of this rate is 1Gbps, which is 10 times that of 5G.

- Spectral efficiency (SE)

SE refers to the peak SE performance metric, which refers to the highest data throughput per unit of spectrum resource under error-free conditions. The target for user peak SE is 30bps/Hz for DL, and 15bps/Hz for UL. In contrast to 5G, 6G will reach up to five times higher SE than that of 5G networks.

- Mobility

  Mobility is the maximum moving speed of a mobile station with the acceptable QoS. In 5G, due to different use-cases, the requirements of mobility are defined as follows. For higher speed user mobility, it should be supported at 500 km/h, while the requirements for the dense urban and indoor hotspot are 30km/h and 10km/h, respectively. In 6G, the target for high speed, such as airline system and hyperloop tube, is up to 1000km/h.

- Latency

  Latency is defined as the network delay between the transmitter and receiver to successfully deliver a packet. The minimal latency of 5G is defined as 1ms compared with 0.01-0.1 ms expected in 6G, amounting to 10 times better. To overcome this issue, automatic control systems and digital twin technology can be introduced to decrease latency [16].

- Connectivity density

  Connectivity density is the total number of connected devices per unit area. In 5G, it aims to support 10^6 devices per square kilometers with different QoS requirements, while it is expected to support 10 times more connected devices to 10^7 per square kilometers in 6G.

- Energy efficiency (EE)

  EE is defined as the amount of information that can be reliably transmitted per Joule of consumed energy, and which is a key performance indicator for 5G networks. In 6G networks, this performance metrics would be 10-100 times better over that of 5G to improve the overall throughput while reducing the total power consumption of the networks.

- Area traffic capacity

  Area traffic capacity is defined as the total mobile traffic that a network can accommodate for a given traffic speed. For the indoor hotspot case, the requirement of the minimal area traffic capacity for 5G is 10 Mbps/m^2, which is expected to reach 1 Gbps/m^2 for 6G.

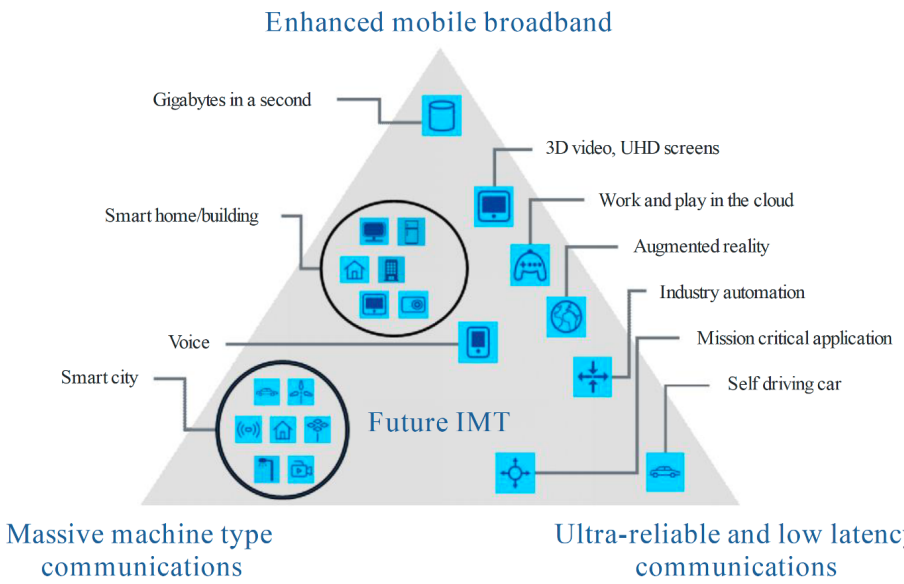

Fig. 1.2 Use-cases and applications for 5G and beyond, copied from [2].

1.2.2 Use-Cases for 5G and Beyond

These key performance metrics are essential elements to enable different envisioned use-cases and applications in 5G and beyond which include automotive communications, remote control with haptic style feedback, immersive AR/VR/MR applications, as well as the very low data rate applications like remote sensors and what is being termed as the IoTs [18]. Furthermore, a broad range of capabilities would be tightly coupled with these intended different use-cases and applications in 5G and beyond [2]. Firstly, three group of use-cases recommended by IMT for 5G and beyond are introduced, which is depicted in Fig. 1.2. These use-cases can be summarized as follows:

- Enhanced mobile broadband (eMBB)

  The eMBB is an extension of 4G broadband services, mainly aiming to fulfill users' demand for an increasingly digital lifestyle, such as ultra-high definition video, augmented reality, and virtual reality [19]. To support a vast amount of applications, the eMBB can be grouped into three main categories namely, broadband access in dense areas, broadband access everywhere, and higher user mobility [20]. Based on this wide variety of usage scenarios, the eMBB has different requirements. For example, for the case of broadband access in dense areas, i.e., multi-storey buildings, urban hospitals, theaters, and stadiums, the requirements of the network will include a higher area traffic capacity and connection density but lower mobility. For the case of broadband access everywhere,

i.e., suburban and rural areas, the requirements of the network will be seamless coverage and a lower data rate. However, in other cases for higher user mobility, i.e., trains, planes, and aircraft, there will be greater demand for mobility but at a lower capacity [21].

- Ultra-reliable and low latency communications (URLLC)

  Unlike the eMBB, the URLLC is a new service category and the most challenging goal in 5G and beyond communication systems in terms of two conflicting metrics, namely, low latency and ultra-high reliability.

  Generally, latency is a measure of network delay, which should meet multiple related requirements for URLLC systems, including user plane latency, control plane latency, and end-to-end latency. In particular, user plane latency is defined as the one-way transmission time to successfully deliver a packet from the radio protocol layer 2/3 service data unit (SDU) entry point to the radio protocol layer 2/3 SDU exit point of the radio interface in either UL or DL, assuming the user equipment is in active state [21]. The target for user plane latency is 0.5ms for UL, and 0.5ms for DL [22]. Control plane latency is defined as the transition time from the most battery-efficient state (e.g., idle state) to the start of continuous data transfer (e.g., active state), which should satisfy the target of 10ms [22]. End-to-end latency refers to the duration between the transmission of a small data packet from source to destination, including transmitter and receiver processing time, over-the-air latency, core network latency, queuing delay, retransmission time, and so on [23]. The URLLC system should provide 10ms end-to-end latency in general and 1ms end-to-end latency for the use cases with extremely low latency [20].

  The reliability is defined as the success probability of transmitting a packet within the given time constraint required by the targeted service [23]. For a general URLLC system, the reliability should meet the requirement that the target packet failure rate of 10^{-5} for 32 bytes within 1 ms over-the-air latency [22].

- Massive machine type communications (mMTC)

  With the exponential growth of connected devices in future wireless networks, mMTC, or so-called massive machine-to-machine (M2M) communication are introduced and designed for IoTs-based services. The main requirement for the mMTC system is to develop massive communication links for a huge number of low-cost devices with limited radio resources [24]. Motivated by the crucial

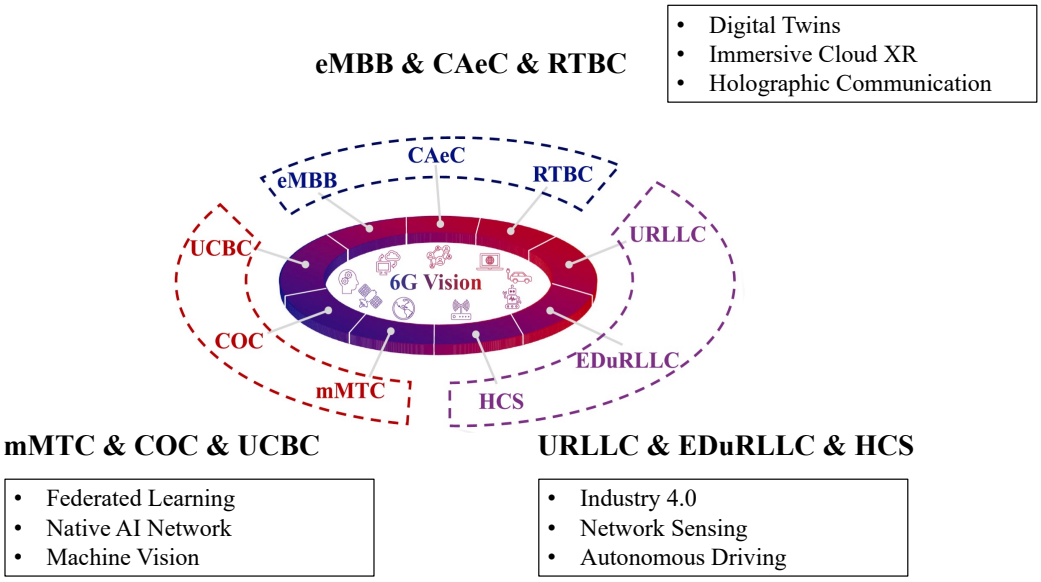

Fig. 1.3 Use-cases and applications of 6G, copied from [3].

role of mMTC in IoT, some applications have already been developed in relevant industries, including structural control of smart buildings, transport and logistics industry, environmental monitoring, and automated health monitoring, and so on.

Based on these 5G usage scenarios, another six new use-cases and applications have been envisioned for 6G as shown in Fig. 1.3, including contextually agile eMBB communications (CAeC), real-time broadband communication (RTBC), event defined URLLC (EDuRLLC), harmonized communication and sensing (HCS), computation oriented communications (COC), and UL centric broadband communication (UCBC) [3]. 6G is expected to go beyond the mobile internet to support ubiquitous artificial intelligence (AI) services and the Internet of Everything (IoE) applications. For example, by applying emerging technologies to reshape our life, holographic communications for tactile and haptic applications to realize multi-dimensional senses exchange (i.e., vision, hearing, touch, smell, and taste) [16]. 6G will also realize the industry 4.0 revolution, overcoming the boundaries between the real factory and the cyber computational space through smart robots and guaranteeing a cost-effective, flexible and efficient way [25].

1.2.3 Key Enabling Techniques for 6G and Beyond

To meet the unprecedented requirements of 6G systems, some key enabling technologies have been identified with unique features. These disruptive technologies include millimeter-wave (mmWave), Terahertz (THz) communications, massive multiple-input multiple-output (massive MIMO), intelligent reflecting surfaces (IRS), non-orthogonal multiple access (NOMA), and mobile edge computing (MEC). Although some of these innovations have already been discussed in the context of 5G, there are still many associated challenges that need to be addressed in terms of technological limitations and practical deployment.

- Millimeter-wave (MmWave)

  The mmWave is one of the key 5G enabling technologies and also remains an essential component in 6G networks, which refers to the range of frequency band between 30 GHz to 300 GHz and the wavelength is between 1 mm and 10 mm [26]. Using much higher frequencies spectrum opens up more spectrum and also provides the possibility of having a much large channel bandwidth. With the utilization of mmWave communications, it can provide us with an extremely broad bandwidth that none of the existing cellular communications have ever seen as well as limited inter-cell interference and a significantly reduced transmission latency [27].

- Terahertz (THz) communications

  Although mmWave communications have the potential advantages of increasing the dimension of antenna arrays and narrow the beams, many of their associated challenges need to be addressed to meet the 6G requirements. Compared to mmWave communications, THz communications with 0.1-10 THz frequency band have a richer spectrum resources and ultra-wideband frequency bands for ultra-high-speed communication [17], [25]. Specifically, they can take advantage of both electromagnetic and light waves as well as reduce the spectrum scarcity and capacity limitations of current wireless communication systems [28].

- Massive multiple-input multiple-output (Massive MIMO)

  Massive MIMO technique has attracted much attention since it is able to provide the flexibility and spatial degrees of freedom by deploying a large number of antennas at the BS [29]. Compared with the MIMO system, massive MIMO use antenna arrays with a few hundred antennas to simultaneously serve many tens

of terminals in the same time-frequency resource. Therefore, the advantage of massive MIMO is not only to reap all the benefits of conventional MIMO, but also on a much greater scale [30].

- Intelligent reflecting surface (IRS)

  To address the signal propagation impairments, IRS has been envisioned as an innovative technology for significantly improving the transmission performance of the networks [3], [28]. In particular, IRS is composed of numerous low-cost passive reflecting elements, where each element is capable of adjusting the phase of the incident signal [31], [32]. In addition to these advantages, an IRS can be deployed between the BS and users to create a new set of high quality transmission environments through improving the signal strength when the line-of-sight communication link is blocked by obstacles (e.g., walls in outdoor buildings) without consuming much transmit power.

- Non-orthogonal multiple access (NOMA)

  NOMA has been advocated as a prospective effective solution for future wireless networks to improve the performance of networks through its potential benefits, including superior SE, compatibility, fairness, flexibility, and massive connectivity [33, 34]. In contrast to the conventional orthogonal multiple access (OMA) scheme, the distinctive feature of NOMA is to serve multiple users using the same orthogonal radio resources, i.e., time and frequency, by exploiting power-domain superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver [33]. In particular, the SC approach is adopted to encode signals intended to different users by using different power levels, which is referred to as the power-domain multiplexing [35]. At the receiver, the SIC technique is utilized at stronger users to decode the signals intended to the weaker users prior to decoding their own signals [34, 35]. For future wireless networks, NOMA has been identified as a viable technique to support the proliferation of IoTs by offering massive connectivity [36].

- Mobile edge computing (MEC)

  Due to the appearance of a lot of applications with computation-intensive and delay-sensitive computational tasks, the devices will achieve a degraded performance in terms of their limited resources and low computation capabilities [37]. MEC technology allows such devices to move processing and storage tasks to the edge computing servers so that the computational delay can be greatly

reduced. By doing this way, MEC can reduce the delay and energy consumption of computing tasks significantly and improve the resource utilization.

1.3 Towards Non-orthogonal Multiple Access

As discussed in the previous section, the main idea of NOMA is to serve more than one user in the same radio resource block, i.e., the time slot in TDMA, the frequency band in OFDMA, the spreading code in CDMA, and the space in space division multiple access (SDMA). There are two dominant types of NOMA, namely, power-domain and code-domain NOMA. The key feature of power-domain NOMA is to allow different users to share the same time, frequency, and code, but with different power levels. This feature makes NOMA capable of increasing SE over the conventional multiple access schemes. In code-domain NOMA, different user-specific spreading codes are assigned to different users and then multiplexed over the same time-frequency resources. This thesis focuses on power-domain NOMA.

The potential capability of NOMA to serve multiple users in the same resource block addresses the increasing demand for massive access in future wireless networks. On the other hand, NOMA can improve user fairness and be combined with other emerging technologies, including multiple-antenna techniques, conventional OMA techniques, and IRS technology. In particular, the combination of NOMA with different multiple-antenna techniques has been extensively investigated in the literature, such as multiple-input single-output (MISO)-NOMA [38] and MIMO-NOMA [39] design. For example, in [38], two algorithms are proposed for solving an EE maximization problem with the DL beamforming design for the MISO-NOMA system. In [39], the EE design is investigated in a multi-cluster multi-user MIMO-NOMA system with pre-defined QoS requirements.

Furthermore, NOMA can also be combined with the existing conventional OMA techniques, such as hybrid TDMA-NOMA [40] and OFDMA-NOMA [41, 42] systems. For example, in a hybrid TDMA-NOMA system, several users are divided into different groups (i.e., clusters) and each cluster is assigned to a time slot for transmission. In particular, multiple users in each group are served by exploiting power-domain NOMA.

1.4 Thesis Outline and Contributions

In 5G and beyond, NOMA is considered as a promising multiple access technique for significantly improving the transmission performance in terms of SE, EE, as well

as massive connectivity [35]. Motivated by demanding heterogeneous requirements of future wireless networks, combining NOMA with other techniques can bring more degrees of freedom and exploit their mutual benefits and complementary features. Therefore, this thesis studied different resource allocation techniques for such hybrid NOMA systems. This thesis consists of seven chapters and the main contributions of each chapter are summarised as follows:

In Chapter 2, basic concepts of NOMA and related literature review are presented. In the first part of this chapter, the fundamental principles of NOMA, such as SIC and SC techniques, are provided by considering an DL transmission of multi-user scenario. Furthermore, the combination of NOMA with multiple antenna, conventional OMA techniques, and IRS are discussed. In the second part of this chapter, a detailed literature review related to recent resource allocation techniques for different NOMA systems is provided.

In Chapter 3, different resource allocation techniques to optimize the performance metrics in future wireless networks are firstly discussed. Next, mathematical optimization techniques related to the radio resource allocation problems in NOMA systems are presented. In particular, the fundamental concepts of convex optimization techniques are discussed in details. Then, different types of convex optimization problems are provided. These problems include linear programming (LP), quadratic programming (QP), quadratic constrained quadratic programming (QCQP), second-order cone programming (SOCP) and semidefinite programming (SDP). Furthermore, multi-objective optimization (MOO) framework are described with necessary details.

In Chapter 4, a hybrid TDMA-NOMA system is introduced, where a joint global energy efficiency maximization (GEE-Max) design for a DL transmission is considered. In such a hybrid system, the available time for transmission is divided into several sub-time slots, and a sub-time slot is allocated to serve a group of users (i.e., cluster). In this design, both the power levels for users and time slot allocations for the clusters are considered for a hybrid TDMA-NOMA system. A feasibility check is carried out as the formulated GEE-Max problem might turn out to be infeasible due to some constraints. Next, to deal with the non-convexity issues, two iterative algorithms are developed to solve the feasible GEE-Max problem. In the first algorithm, a novel SOC formulation along with sequential convex approximations (SCA) is utilized to realize a feasible solution to the problem. In the second algorithm, the Dinkelbach's algorithm is employed to determine the solution of the original GEE- Max problem. Simulation results demonstrate the superior performance of the proposed GEE-Max design with opportunistic time allocations. Furthermore, results also confirm that the

proposed novel SOC approach with the iterative SIC not only provides the solution to the original GEE-Max optimization problem, but also converges within a few number of iterations.

Motivated by the importance of both SE and EE in future wireless networks, an SE-EE trade-off based resource allocation technique for a hybrid TDMA-NOMA system is considered in chapter 5. The EE and SE are conflicting performance metrics. Optimizing SE degrades the overall EE, provided the available transmit power is more than the green power. Similarly, EE maximization does not offer maximum SE. Unlike the work in previous chapter, which aiming to individually maximize the EE of the system, this design offers an additional degree of freedom in resource allocation. Specifically, the proposed design is formulated as a non-convex MOO problem. The MOO framework is reformulated as a single-objective optimization (SOO) problem by combining the multi-objectives through a weighted-sum objective function. With this, each of the original objectives is assigned with a weight factor to reflect its importance in the design. Then, the SCA and a second-order cone (SOC) approach are jointly utilized to deal with the non-convexity issues of the SOO problem. Simulation results reveal that the proposed trade-off based design strikes a good balance between the objective functions, while meeting the requirements of the system. To fully exploit underlying benefits of this hybrid TDMA-NOMA system and guarantee the user-fairness, a max-min resource allocation problem is also formulated in this chapter. However, this max-min problem is non-convex due to coupled design parameters of time and power allocations. Hence, a novel SOC formulation is exploited to overcome this non-convexity issue and an iterative algorithm is developed to realize a solution to the original max-min problem. Simulation results show that this joint resource allocation technique has a considerable performance enhancement in terms of both minimum achieved rate and overall system throughput compared to that of the conventional resource allocation technique where equal time-slots are assigned to the groups of users.

In Chapter 6, another hybrid NOMA system, namely, IRS-assisted multi-user MISO NOMA system is introduced. To take into account the inevitable channel uncertainties, worst-case robust design for such system is studied, where a bounded channel uncertainty model is considered to define the channel state information (CSI) errors. The transmit beamforming and reflecting matrix are designed to maximize the EE under a set of QoS constraints. In particular, the robust EE-Max problem is solved by jointly designing both the transmit beamforming for BS and reflection matrix for IRS based on the definition of the system EE, which is defined as the ratio between the total sum rate and total power consumption. To guarantee the QoS requirement

of each user, the individual minimum data rate is chosen as a performance metric. Moreover, the total transmit power and the phase shift unit-modulus constraints are also incorporated in our formulated problem to guarantee the maximum transmit power at BS and the unit-modulus requirements of the reflection elements at the IRS, respectively. Then, an alternating optimization (AO) algorithm is proposed by applying the *S*-procedure and SDR to address the original non-convex problem. Furthermore, due to the imperfect CSI, the constraints are reformulated in terms of convex linear matrix inequalities (LMIs) that can be easily solved, exploiting the *S*-procedure. Finally, simulation results are provided to demonstrate the performance of the proposed robust IRS-assisted NOMA EE-Max design by comparing that of the design with non-robust scheme in terms of achieved EE. Simulation results further confirm that the proposed AO algorithm can provide the solution to the original non-convex EE-Max optimization problem as well as convergence within a few number of iterations.

Finally, Chapter 7 concludes this thesis and identifies interesting future research directions.

Chapter 2

Fundamentals Concepts and Literature Review

In this chapter, the fundamental concepts of the NOMA systems are introduced. Firstly, the key component techniques of NOMA, namely SC and SIC, are presented, and then the DL transmission of a two-user single-input single-output (SISO)-NOMA system is discussed. Next, the combination of NOMA technology with other 5G and beyond emerging technologies, including multiple-antenna techniques, OMA techniques, and IRS technology, are presented. Finally, the related literature review on EE resource allocation, NOMA with multiple-antenna techniques, NOMA with OMA techniques, and NOMA with IRS technology is shown briefly.

2.1 NOMA Fundamentals

NOMA has been envisioned as a promising multiple access technique to support different use-cases and applications in future wireless networks. This is due to the wide range of potential benefits of NOMA [33, 43, 44]. Different from the conventional OMA technologies, such as TDMA and OFDMA, multiple users in NOMA simultaneously share the same radio resources, namely time and frequency resources. Generally, the NOMA technique can be classified into two categories: power-domain NOMA and code-domain NOMA [35]. In power-domain NOMA, different users are assigned with different power levels according to their channel conditions, which is the type of NOMA investigated in this thesis. Unlike power-domain NOMA, code-domain NOMA allows different users to be assigned with user-specific spreading code sequences over the entire available time-frequency resources. Moreover, code-domain NOMA can be characterized as low-density spreading CDMA (LDS-CDMA), low-density spreading-

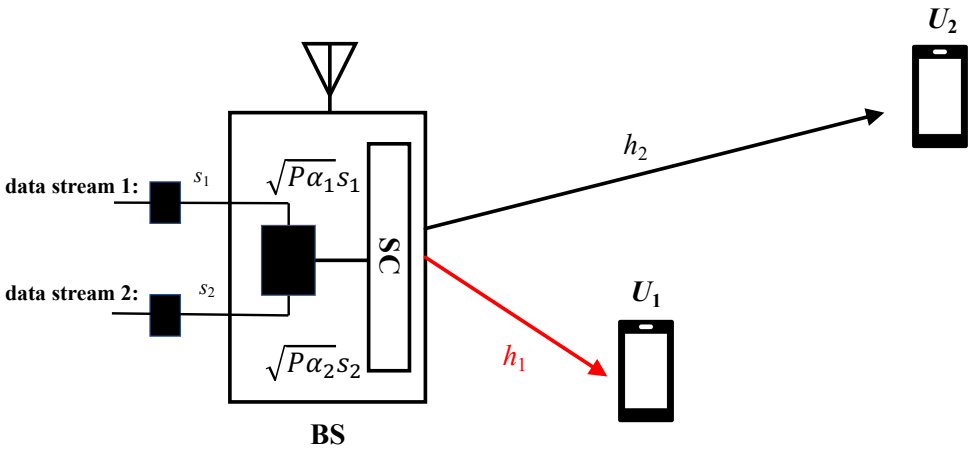

Fig. 2.1 A Two-user SISO NOMA system with SC.

based orthogonal frequency division multiplexing (LDS-OFDM), sparse code multiple access (SCMA), pattern division multiple access (PDMA), and multi-user shared access (MUSA) [45].

At the transmitter, this simultaneous resource sharing is implemented through exploiting power domain multiplexing, which is referred to as a power-domain SC technique. At the receiver, the user with a weaker channel strength (i.e., weaker user) can decode its signal directly by treating the signal of the user with a stronger channel strength (i.e., stronger user) as interference. While stronger user needs to detect the signal intended to weaker users, then subtract this signal from the received signals, and finally decode its own signal without interference. This process is referred to as SIC. In the following subsections, the basic concepts of SC and SIC are introduced.

2.1.1 Superposition Coding and Successive Interference Cancellation

The SC has been considered to improve the capacity of broadcast channels and the SE in wireless systems, which allows the transmitter to transmit signals of multiple users at the same time [35]. In other words, SC is used at the transmitter in order to efficiently multiplex signals from multiple users. For example, as shown in Fig. 2.1, a two-user scenario for the NOMA system is considered, where the BS transmits two

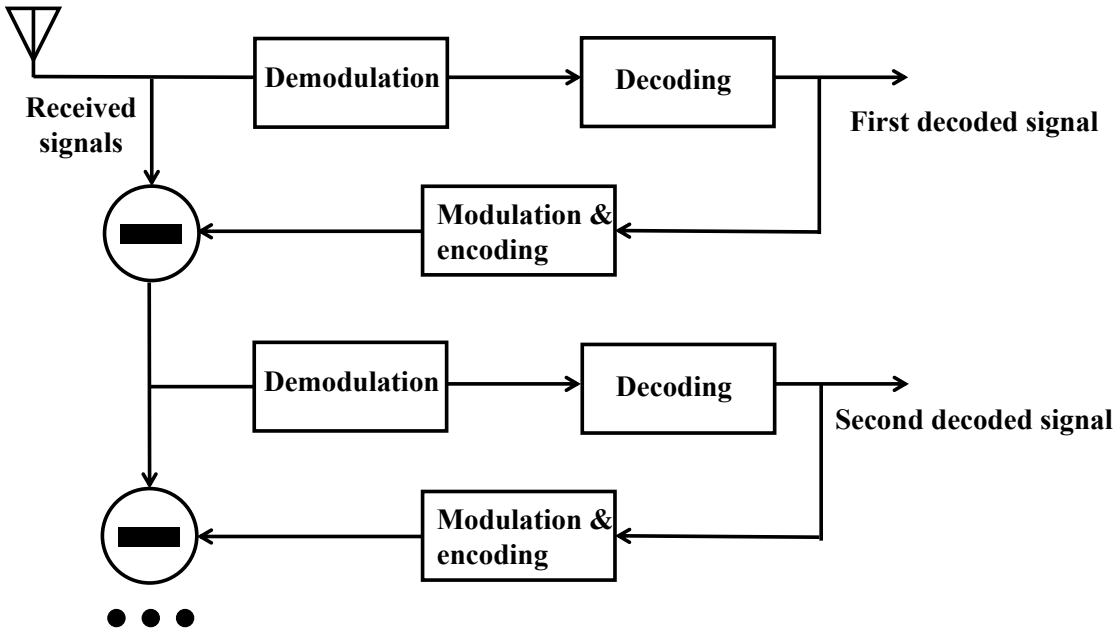

Fig. 2.2 Successive interference cancellation.

signals to both users using SC. The s_1 and s_2 denote the symbols intended to the user 1 (U_1) and the user 2 (U_2), respectively. Let P is the total transmit power and the factor $\alpha_i \in (0, 1)$, $i = 1, 2$, are the fraction of transmit power allocated to user i. Therefore, the transmitted signal from the BS can be defined as

$$x = \sqrt{P\alpha_1}s_1 + \sqrt{P\alpha_2}s_2, \tag{2.1}$$

where $\alpha_1 + \alpha_2 = 1$. It is worth mentioning that stronger users are assigned with lower power levels, while weaker users are assigned with higher power [35]. It is assumed that the U_1 has a better channel condition than that of the U_2 based on the distance between users and BS, which is given by

$$|h_1|^2 \geq |h_2|^2, \tag{2.2}$$

where h_i denotes the channel coefficient between U_i and the BS. Based on the above assumption, it is obvious to realize that $\alpha_1 \leq \alpha_2$ in Fig. 2.1.

During the decoding at the receiver, the user near to the BS can decode other user signals due to stronger channel conditions, while the user far from the BS can only decode its own signals due to weaker channel conditions. The basic idea of the SIC technique at the receiver is to successively decode different user signals and eliminate inter-user interference step by step. As shown in Fig. 2.2, the received signal is initially

demodulated and decoded to detect the signals intended for different users [35]. When one of the user signals is decoded, the corresponding interference is subtracted from the rest of the received signals. The second multiplexed signal can be demodulated and decoded in the absence of interference from the first detected signal. Accordingly, at user U_i, $i = 1, 2$, the received signals for the above SISO system are given by

$$y_i = h_i(\sqrt{P\alpha_1}s_1 + \sqrt{P\alpha_2}s_2) + n_i, \tag{2.3}$$

where n_i denotes the additive white Gaussian noise (AWGN) with zero-mean and variance ${\sigma_i}^2$ at receiver. Specifically, U_1 has better channel conditions than U_2 as defined in (2.2), then U_1 would decode the signal s_2 and remove it from the combined signals to decode s_1 by employing SIC. U_2 decodes its own signal s_2 directly from the received signal with the interference from s_1. Consequently, the received signal at user U_i, $i = 1, 2$ after SIC processing can be expressed as follows:

$$\hat{y}_1 = h_1\sqrt{P\alpha_1}s_1 + n_1, \tag{2.4}$$

$$\hat{y}_2 = h_2\sqrt{P\alpha_2}s_2 + h_2\sqrt{P\alpha_1}s_1 + n_2. \tag{2.5}$$

Therefore, the signal-to-interference and noise ratio (SINR) for U_1 and U_2 can be written by

$$SINR_1^1 = \frac{|h_1|^2 P\alpha_1}{\sigma_1^2}, \tag{2.6}$$

$$SINR_1^2 = \frac{|h_1|^2 P\alpha_2}{|h_1|^2 P\alpha_1 + \sigma_1^2}, \tag{2.7}$$

$$SINR_2^2 = \frac{|h_2|^2 P\alpha_2}{|h_2|^2 P\alpha_1 + \sigma_2^2}, \tag{2.8}$$

where $SINR_1^2$ and $SINR_2^2$ denote the SINR of decoding the message intended for U_2 at the stronger user U_1, the SINR of its own signal of the weaker user, U_2, by treating the signal intended for the stronger user U_1 as interference, respectively. For the weaker user U_2, the SINR of decoding this signal can be defined as

$$SINR_2 = \min\left\{SINR_1^2, SINR_2^2\right\} = SINR_2^2. \tag{2.9}$$

Note that the equation in (2.9) always holds true when $\sigma_1^2 = \sigma_2^2$. Based on the Shannon's capacity definition, the available data rate for U_1 and U_2 can be written by

$$R_1 = \log_2\left(1 + SINR_1^1\right) = \log_2\left(1 + \frac{|h_1|^2 P\alpha_1}{\sigma_1^2}\right), \tag{2.10}$$

$$R_2 = \log_2\left(1 + SINR_2\right) = \log_2\left(1 + \frac{|h_2|^2 P\alpha_2}{|h_2|^2 P\alpha_1 + \sigma_2^2}\right). \tag{2.11}$$

2.1.2 Advantages of NOMA

NOMA can support to meet the unprecedented requirements of future wireless networks due to its different potential benefits and the main advantages of this novel multiple access technique are listed as follows:

- Bandwidth efficiency

  In OMA, such as in OFDMA, a specific frequency resource is assigned exclusively to one user at any time. Then, the overall system may suffer from low SE and throughput since this user cannot utilize the allocated bandwidth efficiently when experiencing a deep channel fading. On the contrary, NOMA is attributed to the fact that it allows each resource block to be shared between multiple users at the same time by completely removing the orthogonality concept in OMA. Hence, the resource assigned for the weaker user is also used by a stronger user, and the interference can be mitigated efficiently through SIC processes at the receiver ends. Therefore, NOMA exhibits a high bandwidth efficiency and improves the overall system throughput.

- Fairness

  In OMA, the user with a good channel condition has a higher priority to be served while the user with a poor channel condition has to wait for access, which opens up naturally fairness problems. However, NOMA always allocates more power to weaker users. By doing so, NOMA can guarantee an attractive trade-off between the fairness among users in terms of their achieved throughput.

- Compatibility

  NOMA is also compatible with the current and future communication systems to offer additional degrees of freedom since it can be invoked as an "add-on" technique for any existing OMA techniques [33]. It does not require significant modifications to the existing architecture. For example, NOMA has been included

in the study of DL multi-user superposition transmission by 3GPP LTE Release 13 and UL NOMA techniques for massive machine type communications by 3GPP LTE Release 14 [46].

- Connectivity

  Based on the definition of OMA, the bandwidth resources occupied by the individual user cannot be shared by other users, which implies that OMA can only support a limited number of users in practical implementation. However, future wireless networks are expected to offer massive connectivity to support billions of smart devices in the IoT era [47]. As such, NOMA can serve multiple users simultaneously with different channel conditions and utilize the available resource blocks efficiently. Therefore, it can support massive number of connections compared to the conventional OMA techniques.

2.2 NOMA with other Techniques

In this section, we present the possibilities of integrating NOMA with different emerging technologies, including multiple-antenna techniques, conventional OMA techniques, and IRS technology.

2.2.1 NOMA with Multiple-Antenna Techniques

To exploit the spatial degrees of freedom, multiple-antenna techniques have gained significant interest in both academia and industry. In particular, multiple antennas have the potential capabilities to enhancing the bit rate, the error performance, and the co-channel interference mitigation of the system [48]. Multiple antennas can be utilized at the transmitters and the receivers. To overcome the path loss due to the propagation environment, massive MIMO can be utilized to generate narrow high-power beams for establishing reliable links [49]. When it is combined with NOMA, the capacity of the networks and system performance, such as SE, EE, and spatial diversity, can be utilized to a greater extent compared to the single antenna systems.

To further demonstrate the combination of NOMA and multiple-antenna techniques, a DL transmission of a multiple-user MISO NOMA system is considered, as shown in Fig. 2.3. This MISO system consists of a BS with M antennas and K single antenna users $U_1, U_2, \ldots, U_K$ located within the coverage area. All signals from BS are mapped onto the antenna array by corresponding beamforming vector $\mathbf{w}_i \in \mathbb{C}^{M\times 1}$,

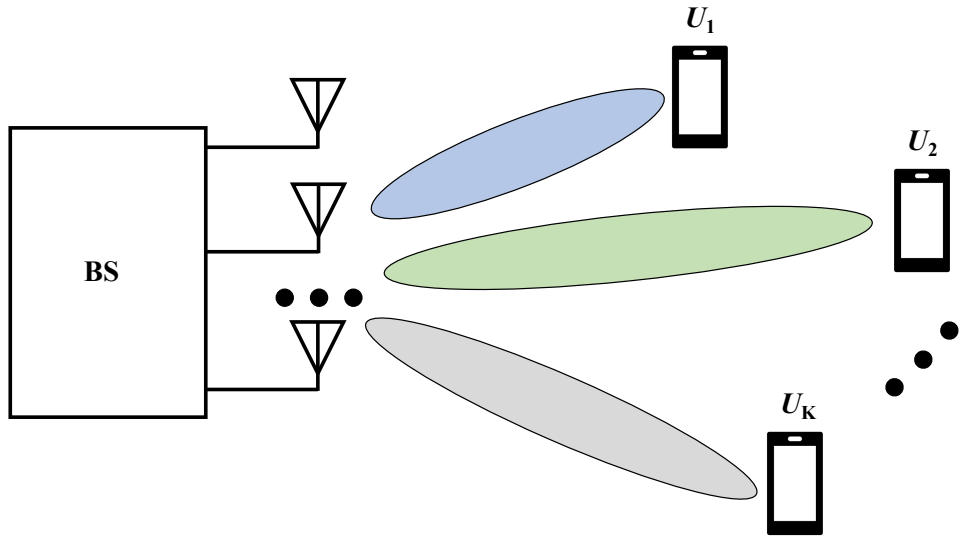

Fig. 2.3 A multiple-user MISO system.

$i = 1, 2, ..., K$. Hence, the transmitted signal $\mathbf{x}$ from the BS is given by

$$\mathbf{x} = \sum_{i=1}^{K} \mathbf{w}_i s_i, \tag{2.12}$$

where s_i denotes the transmitted symbol for user k. Assume that s_i, $i = 1, 2, ..., K$, are independent and uncorrelated, and the signal powers $E[|s_i|^2] = 1$. The channel coefficient between the BS and the ith user is denoted by $\mathbf{h}_i \in \mathbb{C}^{M\times 1}$, $i = 1, 2, ..., K$, and those channels are assumed to be quasi-static fading channels, which means that $\mathbf{h}_i$ remains constant within each symbol period. Therefore, the received signal y_i at user U_i can be expressed as follows:

$$y_i = \mathbf{h}_i^H \mathbf{x} + n_i, i = 1, 2, ..., K. \tag{2.13}$$

By combining both (2.12) and (2.13), the received signal y_i at user U_i can be rewritten as

$$\begin{aligned} y_i &= \mathbf{h}_i^H \left(\sum_{j=1}^{K} \mathbf{w}_j s_j \right) + n_i \\ &= \underbrace{\mathbf{h}_i^H \mathbf{w}_i s_i}_{desired\ signal} + \underbrace{\mathbf{h}_i^H \left(\sum_{j=1, j\neq i}^{K} \mathbf{w}_j s_j \right)}_{interference} + \underbrace{n_i}_{noise}, i = 1, 2, ..., K. \end{aligned} \tag{2.14}$$

In the NOMA scheme, users can be multiplexed in the power domain and the user ordering can be determined according to their channel strengths. As such, the first user (i.e., U_1) has the strongest channel strength while the channel strength of U_K is the weakest. In other words, the channels can equivalently be ordered as follows:

$$||\mathbf{h}_1||^2 \geq ||\mathbf{h}_1||^2 \geq \cdots \geq ||\mathbf{h}_K||^2. \tag{2.15}$$

Based on the users' channel conditions, U_i should be able to decode the signals intended for the users from U_i to U_K and effectively remove the interference from the $\{i+1\}$th to Kth users, whereas the signals intended for the rest of the users, i.e., $U_1, \ldots, U_{i-1}$ are treated as the interference at U_i. Therefore, the received signal at U_i after employing SIC can be expressed as

$$\hat{y}_i = \mathbf{h}_i^H \mathbf{w}_i s_i + \mathbf{h}_i^H \left(\sum_{j=1}^{i-1} \mathbf{w}_j s_j \right) + n_i, i = 1, 2, ..., K, \tag{2.16}$$

where the first term in (2.16) indicates the intended signal for the ith user. The second term represents the interference from the signals intended for the users $U_1, \ldots, U_{i-1}$. The last term is the AWGN. According to Shannon's capacity theorem, the achievable data rate at the ith user can be defined as follows:

$$R_i = B \log_2 \left(1 + \frac{|\mathbf{h}_i^H \mathbf{w}_i|^2}{\sum_{j=1}^{i-1} |\mathbf{h}_i^H \mathbf{w}_j|^2 + {\sigma_i}^2} \right), \tag{2.17}$$

where B represents the available bandwidth in Hz, which is assumed to be one ($B = 1$) throughout this thesis.

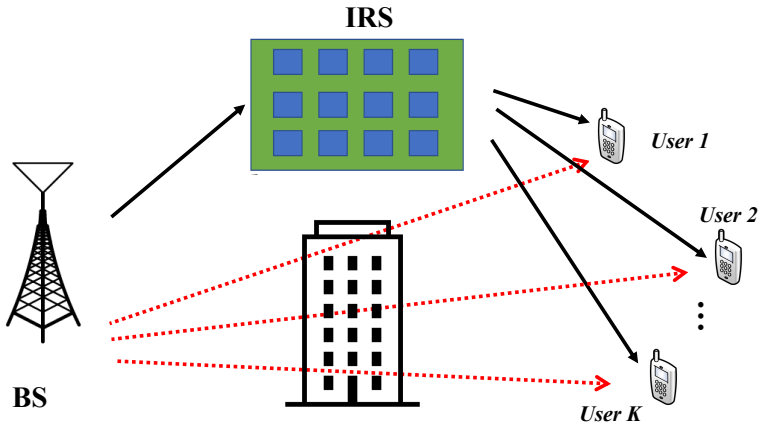

Fig. 2.4 Illustration of IRS-assisted NOMA communication system.

2.2.2 NOMA with OMA Techniques

Combining NOMA with other orthogonal multiple access techniques, such as OFDMA, TDMA, offers additional degrees of freedom as different domains can be exploited to serve a larger number of users [33, 50]. Different hybrid OMA-NOMA approaches have been considered in the literature, such as OFDMA-NOMA [41, 42, 51–54] and TDMA-NOMA approaches [55, 56]. In addition, the practical challenges of employing NOMA in dense networks can be addressed through such combined (i.e., hybrid) OMA techniques. For example, in the hybrid TDMA-NOMA system, the power-domain based NOMA and the time-domain based TDMA are jointly utilized to serve the users in the DL transmission [55, 57]. In particular, the hybrid TDMA-NOMA system has been recently identified as a promising multiple access technique for various applications in 5G and beyond [55]. They include IoT applications [58], multi-satellite relay transmission systems based on TDMA and NOMA [59], and TDMA-based unified resource management scheme for MEC over Ethernet-based fiber-wireless network [60].

2.2.3 NOMA with IRS technology

IRS is a planar surface comprising a large number of passive reflecting elements, each of which is able to dynamically alter wireless channels to enhance the communication performance [61]. This technology can be beneficial in many scenarios to improve the

performance of the system. For example, it can be deployed on the outside walls of buildings or the ceiling of factories to bypass obstacles, and can be utilized to suppress the co-channel or inter-cell interference between users. In contrast to the conventional relaying systems, the IRS-assisted wireless communications reflect the signals in a passive way, which can support the link capacity and provide a new solution to solve the channel fading and interference issues [61].

Moreover, IRS can also be incorporated into NOMA communication systems. As shown in Fig. 2.4, a DL IRS-assisted NOMA communication system is illustrated, where there is one BS, one IRS and K users. It is obvious to see that the direct links between the users and the BS are blocked by obstacles. By employing an IRS between the users and the BS, the new set of transmission links can be successfully established by adjusting the reflection coefficients, including phase shift and amplitude of the IRS [62].

2.3 Energy-Efficient Strategies

With the exponential growth of connected devices in future wireless networks, the corresponding power consumption becomes a significant issue, requiring careful consideration in the relevant designs [63]. In particular, the excessive power consumption not only introduces an uncontrolled increase of the CO_2 emission levels, but also imposes additional financial pressures on the users [64, 65]. Therefore, different solution approaches have been proposed to address these challenges of the unprecedented growth in the power consumption. These approaches include the deployment of renewable energy resources [66], recent developments of energy harvesting techniques using simultaneous wireless information and power transfer (SWIPT) [67], and integration of energy efficient resource allocation techniques into wireless networks [68, 69].

EE is defined as the ratio of the achieved sum-rate to the corresponding total power consumption [63]. The EE based resource allocation techniques strike a good balance between these two conflicting goals of maximizing the achieved sum-rate and minimizing the corresponding power consumption [68, 70]. In addition, several EE designs have been investigated for NOMA transmissions system in the literature, such as the works in [39, 71–73]. The power and bandwidth allocation have been jointly considered for a SISO-NOMA system in [71] to maximize the system EE with the transmit power constraint and minimum rate constraints. In [72], an EE maximization (EE-Max) problem for the DL SISO-NOMA system has been solved through the EE optimal power allocation strategy. An EE fairness for MISO-NOMA systems has been

studied in [73], in which max-min EE and proportional fairness designs have been addressed relying on the SCA technique. The transmission scheme for EE has been proposed for the multi-user MIMO-NOMA system under a QoS constraint for each user in [39]. Numerical results show that the proposed NOMA design improves both EE and the number of accessible users compared with the conventional OMA.

2.4 Literature Review

NOMA has been recently combined with a wide range of techniques. These include multiple-antenna [39, 74–78] and conventional OMA techniques [41, 42, 55, 56, 79]. In these combined systems, power domain multiplexing is utilized along with the other existing spatial and orthogonal domains multiplexing to meet the demanding massive connectivity requirements. In particular, these hybrid systems not only exploit different multiplexing domains to enhance the performance but also facilitate the practical implementation of NOMA in dense networks [68, 80]. For example, the SE-EE trade-off based design has been proposed for a MISO-NOMA system in [81]. In this design, a weighted sum approach based on the priori articulation is utilized to convert the MOO problem into a SOO problem. An EE maximization problem has been studied for the massive MIMO-NOMA uplink system in [68]. Then, the proposed schemes achieve superior EE when compared with the conventional OMA scheme. Another example SE-EE trade-off based design for massive MIMO systems has been considered in [82] through the particle swarm optimization algorithm.

In addition, some research works for hybrid OMA-NOMA systems investigate different resource allocation techniques [40–42]. In hybrid OMA-NOMA systems, the available OMA resources (i.e., time or frequency) are divided into several sub-resource blocks, where each sub-resource block corresponds to a set of multiple users via NOMA technique [40, 80]. For example, in a hybrid OFDMA-NOMA system, an EE maximization problem has been considered in [41], in which the sub-channel assignment and the power allocation algorithms were proposed for the system. In [42], the max-min sum of DL and UL transmit rates among all users joint resource allocation problem of OFDMA-NOMA system has been investigated. Specifically, two scenarios have been considered, perfect CSI estimation and imperfect CSI estimation. Then, an asymptotically optimal algorithm and a sub-optimal algorithm have been proposed. The energy harvesting capabilities of a hybrid TDMA-NOMA system have been explored in [40], in which several users are divided into different groups (i.e., clusters) and each cluster is assigned to the equal time slot for transmission, aiming to

minimize the transmit power under minimum rate and minimum energy harvesting requirements at each user.

Several IRS-assisted NOMA communication systems have been considered in the literature, such as the works in [62, 70, 83–85]. In [83], a DL IRS-assisted NOMA transmission scheme has been studied, where IRSs have been deployed at cell-edge regions to maximize the total number of served users. In [62], a DL multi-channel IRS-NOMA system has been considered, specifically, the channel assignment, decoding order, power allocation and reflection coefficients have been jointly optimized to maximize the achievable system throughput. Further, the design of the UL IRS-NOMA transmission scheme under individual power constraint has been investigated in [84]. By considering the secrecy of the IRS-aided NOMA network, a robust beamforming scheme with imperfect CSI of the eavesdropper has been addressed in [85]. Furthermore, to maximize the EE, a power allocation and phase shift design strategy has been considered in [70] by the AO algorithm. The joint power allocation and the reflection matrix of IRS in this context has been further explored in [61, 86, 87] under the assumption of perfect CSI at the BS. In [86], a two-user IRS-assisted DL communication with discrete phase shifts has been studied in both NOMA and OMA systems, with FDMA and TDMA. A comparison with FDMA/TDMA showed that TDMA can yield superior performance compared with FDMA when the latter case lacked frequency-selective IRS reflection coefficients. In addition, it was demonstrated that TDMA requires lower transmit power than NOMA under symmetric user channels, which is due to the fact that the IRS reflection has an effect on users' individual channels in the former case. Furthermore, this two-user case has been extended to a general case in [61, 87], which the number of users is more than two.

2.5 Summary

In this chapter, the fundamental principles of NOMA are introduced, including SC, and SIC, which lays a comprehensive foundation for the technical work in the thesis. Then, the main advantages of NOMA are discussed, including bandwidth efficiency, fairness, compatibility, and connectivity, which provides the background knowledge for the NOMA. Furthermore, the combinations between NOMA and other techniques are presented, including multiple-antenna techniques, conventional OMA techniques, and IRS technology, which offers new insight into these emerging technologies. Finally, recent research works related to NOMA in the literature are reviewed.

Chapter 3

Mathematical Background

Optimization techniques play an important role in the design of many practical communication systems, especially in research areas, such as signal processing and resource allocations [88–90]. In this chapter, the fundamental concepts of resource allocation techniques, convex optimization techniques, and the MOO framework are provided.

3.1 Resource Allocation Techniques in Wireless Networks

Over the past decade, the resource allocation has drawn a significant attention in wireless networks due to its various benefits [91, 92]. For example, different users may have different requirements and capabilities in terms of their locations and channel conditions. The system performances may worsen without considering user diversity and the assignment of the limited resources. In addition, considering the communication resources, such as space, time slots, frequency bandwidths, or power levels, the system performances can be improved by taking the advantages of those diverse resources. Therefore, a suitable performance metric needs to be selected based on several factors when designing any future wireless communication systems. In the following, the resource allocation techniques that have been considered in this thesis are introduced briefly.

3.1.1 Power Control Techniques

In wireless communications, transmission power is one of the most important radio resources [75, 93]. Power control, also known as transmit power control, is a crucial design problem in wireless networks, as it mitigates the consequences of two fundamental limitations of wireless networks, including radio spectrum and the life of the battery [92]. The objective of power control in wireless networks is to control the transmit power to guarantee a certain link quality at each user. For instant, a power minimization problem, which is referred as P-Min problem, is to minimize the transmit power consumption under different QoS constraints at each user:

$$\begin{aligned} \underset{P_k}{\text{minimize}} \quad & \sum_k P_k \\ s.t. \quad & SINR_k \geq \widetilde{SINR}_k, \quad \forall k, \end{aligned} \tag{3.1}$$

where P_k is the transmit power allocated for the kth user, and $\widetilde{SINR}_k$ is the predefined QoS requirements for the kth user.

3.1.2 Rate-Aware Techniques

Compared to the power control, rate-aware based designs provide additional degrees of freedom which can be utilized to improve the SE in 6G and beyond wireless networks [91]. Several rate-aware resource allocation techniques have been widely proposed for different wireless networks in the literature [94, 95]. In some application scenarios, the SE is maximized with the available power constraint at the transmitter. In particular, this utility function is defined as the ratio between the total achievable sum rate and the total bandwidth and expressed in bits per second per Hz (bps/Hz). This can be mathematically expressed as

$$\begin{aligned} \text{minimize} \quad & \frac{\sum_k R_k}{B} \\ s.t. \quad & \sum_k P_k \leq \widetilde{P}, \quad \forall k, \end{aligned} \tag{3.2}$$

where R_k is the achievable data rate of the kth user, B is the bandwidth, and $\widetilde{P}$ is the available total transmit power.

However, maximizing overall throughput of the system degrades the performance of individual users while compromising user-fairness in terms of achievable rates. Hence, this fairness problems have been developed, which can be expressed as the minimum

achievable rate between all users:

$$\begin{aligned} \text{maximum} \quad & \min \; R_k \\ s.t. \quad & \sum_k P_k \leq \widetilde{P}, \quad \forall k. \end{aligned} \tag{3.3}$$

In such a max-min problem, equal rates will be obtained for all users while satisfying the power constraints [96].

The power consumption of wireless networks is one of the dominant factors that has a considerable impact in the design of future wireless communications systems. On one hand, the excessive power consumption will cause an uncontrolled increase of CO_2 emission levels [64, 65, 97], which raises different environmental issues including global warming and natural disasters [98]. On the other hand, the accelerated growth of the power consumption will be inherently reflected on the overall costs of the wireless communication systems. This, as a result, will impose additional financial pressures on the service providers and consumers. Therefore, EE is defined as the ratio between the achieved sum rate and the corresponding total power consumption [63]. The EE of a communication system with unit of bit-per joule can be defined as the number of bits transmitted per joule of energy consumption in the system [99], which can be written as

$$\begin{aligned} \text{minimize} \quad & \frac{\sum_k R_k}{\sum_k P_k} \\ s.t. \quad & \sum_k P_k \leq \widetilde{P}, \quad \forall k, \\ & SINR_k \geq \widetilde{SINR}_k, \quad \forall k. \end{aligned} \tag{3.4}$$

This section provides different resource allocation techniques to optimize crucial performance metrics in future wireless networks including the motivations behind each performance metric are introduced. Specifically, the EE and power minimization performance metric is mainly focused on in Chapter 4 and Chapter 6. In addition, the SE and max-min fairness performance metric has been investigated in Chapter 5. In the following sections, an overview of the fundamentals of convex optimization is provided to mathematically express these resource allocation techniques.

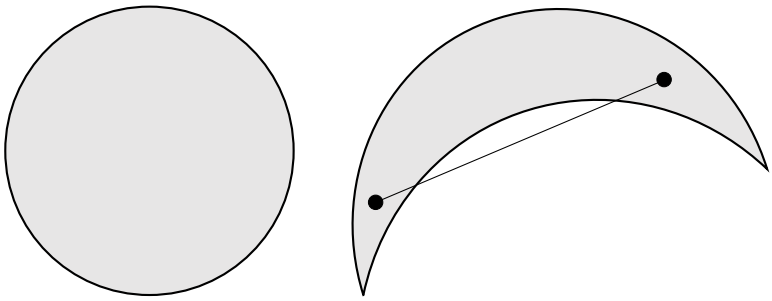

Fig. 3.1 Example of a convex set (left) and a non-convex set (right).

3.2 Convex Optimization

Generally, optimization problems can be categorized as convex optimization problems or non-convex optimization problems. The former problems can be solved efficiently to determine the optimal solution through interior point methods, while the latter problems cannot be solved directly by convex optimization tools and software due to their non-convex nature [88, 100].

3.2.1 Fundamentals of Convex Optimization

Convex Sets

A set $\mathcal{C} \in \mathbb{R}^n$ is called convex if the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$, i.e., for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$, and any $\theta \in [0, 1]$ [90]. This can be mathematically expressed as

$$\theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in \mathcal{C}. \tag{3.5}$$

For example, the unit circle $\{x \in \mathbb{R}^2 | x_1^2 + x_2^2 = 1\}$ is convex, where (x_1, x_2) is a point on the unit circle's circumference. However, a crescent shape, is not convex since the line segment joining two distinct points is no longer on the crescent shape, as shown in Fig. 3.1. In general, a convex set must be a solid body, containing no holes, and always curve outward. It is worthy to note that the convex set remains as convex in the following operations:

- The convex set is convex under intersection, even if the number of sets is infinite.

- If $\mathcal{C} \in \mathbb{R}^n$ is a convex set, $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, then its affine transform set $\mathbf{A}\mathcal{C} + \mathbf{b}$ also preserves convexity, where

$$\mathbf{A}\mathcal{C} + \mathbf{b} = \{\mathbf{A}\mathbf{x} + \mathbf{b} \mid \mathbf{x} \in \mathcal{C}\}.$$

- If $\mathcal{C}$ is a convex set, $\mathcal{C} \subseteq \mathbf{dom}\ P = \mathbb{R}^n \times \mathbb{R}_{++}$, then its perspective transform

$$P(\mathcal{C}) = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \mathbf{x}/t \in \mathcal{C}, t > 0\},$$

  is convex. Note that $\mathbb{R}_{++}$ denotes the set of positive numbers. The perspective function scales or normalizes vectors so the last component is one, and then drops the last component [90]. The perspective function can be interpreted as the action of a pinhole camera. For example, a pinhole camera (in $\mathbb{R}^3$) consists of an opaque horizontal plane $x_3 = 0$ and a horizontal image plane $x_3 = -1$. Then, an object at x, such as (x_1, x_2, x_3) through a pinhole at $(0, 0, 0)$ on the opaque horizontal plane $x_3 = 0$ forms an image at the point $-(x_1/x_3, x_2/x_3, 1)$ on the image plane $x_3 = -1$. The last component of the image point can be dropped since it is always fixed.

Convex Cones

A convex cone $\mathcal{C}$ is a special type of convex set, that is, for each $\mathbf{x} \in \mathcal{C}$ and each $\theta \geq 0, \theta\mathbf{x} \in \mathcal{C}$. This can be mathematically expressed as

$$\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 \in \mathcal{C},\ \forall\theta_1 \geq 0,\ \forall\theta_2 \geq 0,\ \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}. \tag{3.6}$$

In communications, engineering application and signal processing, different forms of convex cones are utilized, and the most common convex cones are as follows [90]:

- Nonnegative orthant $\mathbb{R}^n_+$

  Any subspace is affine, and a convex cone.

- Second-order cone (Quadratic cone)

  $\mathcal{C} = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid ||\mathbf{x}||_2 \leq t\}$, where $||\cdot||_2$ denotes the Euclidean norm, i.e., $||\mathbf{x}||_2 = (\mathbf{x}^T\mathbf{x})^{1/2}$.

- Positive semidefinite cone

  $\mathcal{C} = \{\mathbf{X} \in \mathbb{S}^n_+ | \mathbf{X} \succeq 0\}$, where $\mathbb{S}^n_+$ denotes the set of symmetric positive $n \times n$ matrices.

Convex Functions

A function $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is convex if **dom** f is a convex set and if $\forall \mathbf{x}, \mathbf{y} \in$ **dom** $f(\mathbf{x})$, and $\forall \theta \in [0, 1]$ the following inequality holds [90]

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}). \tag{3.7}$$

The above inequality means that the line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ lies above the graph of f. A function $f(\mathbf{x})$ is called strictly convex if strict inequality holds in (3.7) whenever $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$. The function $f(\mathbf{x})$ is concave if $-f(\mathbf{x})$ is convex.

Suppose $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, for all $\mathbf{x}, \mathbf{y} \in$ **dom** $f(\mathbf{x})$, then $f(\mathbf{x})$ is convex if and only if **dom** $f(\mathbf{x})$ is convex and the following inequality holds

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}). \tag{3.8}$$

It is clear to see that the inequality in (3.8) is the first-order Taylor series approximation of f near $\mathbf{x}$, which is a global underestimator of the function. In other words, if the first-order Taylor series expansion of a function is always a global underestimator of the function, then the function is convex.

Furthermore, if $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, then $f(\mathbf{x})$ is convex if the following inequality holds,

$$\nabla^2 f(\mathbf{x}) \succeq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \tag{3.9}$$

where $\nabla^2 f$ is its Hessian or second derivative. The condition in (3.9) can be interpreted geometrically as the requirement that the graph of the function has positive (upward) curvature at $\mathbf{x}$ [90]. Thus, all linear and affine functions are convex, and for quadratic functions $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{P} \in \mathbb{S}^n, \mathbf{q} \in \mathbb{R}^n, a \in \mathbf{R}$, with $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + a$ is convex if and only if $\mathbf{P} \succeq 0$ since $\nabla^2 f(\mathbf{x}) = \mathbf{P}$.

3.2.2 Convex Optimization Problems

A convex optimization problem can be mathematically expressed as follows [90]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_0(\mathbf{x}) \\ s.t. \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\ & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, p, \end{aligned} \tag{3.10}$$

where the vector $\mathbf{x} \in \mathbb{R}^n$ is the *optimization variables* of the problem, f_0 is called the objective function (or cost function), the functions $f_0, f_1 \ldots, f_m$ are convex functions and the functions $h_1, h_2, \ldots, h_p$ are linear functions, $f_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m$, and $h_i(x) = 0, i = 1, 2, \ldots, p$, are called the inequality and equality constraint functions, respectively. Generally, the convex optimization problem in (3.10) aims to find an $\mathbf{x}$ that minimizes $f_0(\mathbf{x})$ among all $\mathbf{x}$ while satisfying all inequalities and equality constraints. The problem is called an *unconstrained problem* if there is no constraints, i.e., $m = p = 0$. The domain of the problem (3.10) is the set of points, for which the objective and the constraints can be defined as

$$D = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\ h_i. \tag{3.11}$$

If a point $\mathbf{x} \in D$ satisfies all inequalities and equality constraints, $f_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m$, and $h_i(\mathbf{x}) = 0, i = 1, 2, \ldots, p$, then it is *feasible*. The problem (3.10) is said to be *feasible* if there exists at least one feasible point, and otherwise it is *infeasible*. A feasible solution $\mathbf{x}^*$ is said to be *optimal point* if

$$f_0(\mathbf{x}^*) \leq f_0(\mathbf{x}), \quad \forall \mathbf{x} \in D. \tag{3.12}$$

A feasible point $\bar{\mathbf{x}}$ is called locally optimal if there exists an $R > 0$, such that $f_0(\bar{\mathbf{x}}) \leq f_0(\mathbf{x})$ for all feasible $\mathbf{x}$ satisfying $||\mathbf{x} - \bar{\mathbf{x}}||_2 \leq R$. Based on the above definition, a problem can be classified as a convex optimization problem, which should satisfy the following requirements:

- the objective function $f_0(\mathbf{x})$ must be convex,
- the inequality constraint functions $f_i(\mathbf{x}), i = 1, 2, \ldots, m$ must be convex,
- the equality constraint functions $h_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} - b_i$ must be affine.

Linear Programming

When the objective and constraint functions are all affine, then the convex optimization problem is known as a LP [90]. A standard form of LP can be defined as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^T\mathbf{x} + d \\ s.t. \quad & \mathbf{G}\mathbf{x} \preceq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.13}$$

where $\mathbf{G} \in \mathbb{R}^{m \times n}$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$. Since the maximization problem $\mathbf{c}^T\mathbf{x} + d$ can be solved by minimizing the function $-\mathbf{c}^T\mathbf{x} - d$ subject to the constraints, a maximization problem with affine objective and constraint functions can also be referred as an LP.

Quadratic Programming

The convex optimization problem is called a QP, when the objective function is quadratic and the constraint functions are affine. A QP takes the following form [90]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r \\ s.t. \quad & \mathbf{G}\mathbf{x} \preceq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.14}$$

where $\mathbf{P} \in \mathbb{S}_+^n$, $\mathbf{G} \in \mathbb{R}^{m \times n}$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$. LP is a special case of QP by setting $\mathbf{P} = 0$ in the objective function of (3.14).

Quadratic Constrained Quadratic Programming

If the objective and constraint functions are quadratic, then this convex optimization problem is called QCQP. A QCQP can be expressed as follows [90]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}_0^T\mathbf{x} + r_0 \\ s.t. \quad & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} + \mathbf{q}_i^T\mathbf{x} + r_i \leq 0, \quad i = 1, 2, \ldots, m, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.15}$$

where $\mathbf{P}_i \in \mathbb{S}_+^n, i = 0, 1, \ldots, m$. QCQP includes QP (and therefore also LP) as a special case. This may be obtained by setting $\mathbf{P}_i = 0, i = 1, 2, \ldots, m$ in the constraints of (3.15).

Second Order Cone Programming

A SOCP can be defined in the following form [90]:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{f}^T\mathbf{x} \\ s.t. \quad & ||\mathbf{A}_i\mathbf{x} + \mathbf{b}_i||_2 \leq \mathbf{c}_i^T\mathbf{x} + d_i, \quad i = 1, 2, \ldots, m, \\ & \mathbf{F}\mathbf{x} = \mathbf{g}, \end{aligned} \tag{3.16}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}_i \in \mathbb{R}^{n_i \times n}$, and $\mathbf{F} \in \mathbb{R}^{p \times n}$. In particular, the constraint $||\mathbf{A}\mathbf{x} + \mathbf{b}||_2 \leq \mathbf{c}^T\mathbf{x} + d$ is called a second order cone constraint, where $\mathbf{A} \in \mathbb{R}^{k \times n}$. A QCQP can be obtained by setting $\mathbf{c}_i = 0, i = 1, 2, \ldots, m$ and squaring both sides of the constraints of (3.16). When $\mathbf{A}_i = 0, i = 1, 2, \ldots, m$, then this problem in (3.16) will be reduced to a LP.

Semidefinite Programming

The general form of SDP can be written as [90]

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^T\mathbf{x} \\ s.t. \quad & \mathbf{x}_1\mathbf{F}_1 + \mathbf{x}_2\mathbf{F}_2 + \cdots + \mathbf{x}_n\mathbf{F}_n + \mathbf{G} \preceq 0, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.17}$$

where $\mathbf{G}, \mathbf{F}_1, \ldots, \mathbf{F}_n \in \mathbb{S}^{k \times k}$, and $\mathbf{A} \in \mathbb{R}^{p \times n}$. A LP can be obtained if the matrices $\mathbf{G}, \mathbf{F}_1, \ldots, \mathbf{F}_n$ are all diagonal.

A standard form of SDP can be expressed as [90]

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \mathbf{tr}(\mathbf{C}\mathbf{X}) \\ s.t. \quad & \mathbf{tr}(\mathbf{A}_i\mathbf{X}) = b_i, \quad i = 1, 2, \ldots, p, \\ & \mathbf{X} \succeq 0, \end{aligned} \tag{3.18}$$

where $\mathbf{C}, \mathbf{A}_1, \ldots, \mathbf{A}_p \in \mathbb{S}^{k \times k}$. This form of SDP in (3.18) has linear equality constraints, and a (matrix) non-negativity constraint on the variable $\mathbf{X} \in \mathbb{S}^{n \times n}$.

The SDP can also be represented as the form of multiple LMIs and linear inequalities, such as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^T\mathbf{x} \\ s.t. \quad & \mathbf{F}^{(i)}(\mathbf{x}) = \mathbf{x}_1\mathbf{F}_1^{(i)} + \cdots + \mathbf{x}_n\mathbf{F}_n^{(i)} + \mathbf{G}^{(i)} \preceq 0, \quad i = 1, 2, \ldots, K, \\ & \mathbf{G}\mathbf{x} \preceq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{3.19}$$

This form of SDP in (3.19) has linear objective, linear equality and inequality constraints, and several LMI constraints.

3.3 Cholesky Decomposition

The Cholesky decomposition is useful if the matrix $\mathbf{A}$ is symmetric and positive definite [101], which can be defined as follows:

If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric and positive definite, then there exists a unique lower triangular matrix $\mathbf{G} \in \mathbb{R}^{n \times n}$ with positive diagonal entries such that $\mathbf{A} = \mathbf{G}\mathbf{G}^T$. Note that Cholesky decomposition can be rewritten in the form of $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, where $\mathbf{B}$ is an upper triangular matrix with positive diagonal elements [102]. Every Hermitian positive-definite matrix has a unique Cholesky decomposition.

The advantages of Cholesky decomposition is that the computation (i.e., without exploiting any structure, $(\frac{1}{3})n^3$ flops) faster than lower upper (LU) decomposition since only a lower triangular is needed in this processing. Based on the above definition, it is clear that the Cholesky decomposition is composed of a lower triangular matrix $\mathbf{G}$ and its transpose $\mathbf{G}^T$ serving as an upper triangular part. The elements of $\mathbf{G}^T$ are related to the elements of the lower triangular matrix $\mathbf{G}$ as follows:

$$\forall g_{i,j} \in \mathbf{G}, \forall b_{i,j} \in \mathbf{G}^T \Rightarrow b_{i,j} = g_{j,i}. \tag{3.20}$$

The diagonal elements $(g_{k,k})$ of $\mathbf{G}$ are determined as follows:

$$g_{k,k} = \sqrt{a_{k,k} - \sum_{j=1}^{k-1} g_{k,j}^2}, \qquad k = 1, 2, \ldots, n, \tag{3.21}$$

and the other elements $(g_{i,k}, i > k)$ are

$$g_{i,k} = \frac{1}{g_{k,k}} \left(a_{i,k} - \sum_{j=1}^{k-1} g_{i,j} g_{k,j} \right). \tag{3.22}$$

3.4 Multi-Objective Optimization

The MOO method has been recently utilized to solve different problems with conflicting objectives in wireless networks. In contrast to a unique global optimal solution in the SOO problems, the MOO problems have many Pareto-optimal solutions and each of those solutions is capable of striking a good balance between the conflicting objectives. Thus, they can simultaneously achieve a better performance in all objectives without significantly sacrificing any of those objectives [103]. A MOO problem has a number

of conflicting objectives:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_M(\mathbf{x})]^T \\ s.t. \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p, \end{aligned} \tag{3.23}$$

where M is the number of objective functions, m is the number of inequality constraints, and p is the number of equality constraints. $\mathbf{F}(\mathbf{x})$ is a vector function containing M objective functions and $\mathbf{x}_i^*$ is the point that minimizes the objective function $\mathbf{F}_i(\mathbf{x})$. $\mathbf{X}$ is the feasible design space, such that,

$$\mathbf{X} = [\mathbf{x} | g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m; h_j(\mathbf{x}) = 0, j = 1, 2, \ldots, p.] \tag{3.24}$$

The conventional SOO aims to find a single global solution that meets all requirements for an optimal solution. However, in MOO problems, the solutions obtained with no articulation of preferences are arbitrarily relative to the Pareto optimal, which is defined as follows [104]:

A feasible solution $\mathbf{x}^* \in \mathbf{X}$ *is defined as a Pareto-optimal solution if there exists no other feasible solution* $\mathbf{x} \in \mathbf{X}$ *such that* $\mathbf{F}(\mathbf{x}) \leq \mathbf{F}(\mathbf{x}^*)$*, and* $\mathbf{F}_i(\mathbf{x}) < \mathbf{F}_i(\mathbf{x}^*)$ *for at least one function.*

3.4.1 MOO problems Based on Different Utility Functions

Considering multiple objective functions, some methods are introduced to solve MOO problems based on different utility functions in this subsection.

- Weighted Sum Method

  The main idea of this method is to transform a set of objectives into a single objective by adding scaled version of each objective. The weighted sum method can be written as [104],

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{m=1}^{M} w_m f_m(\mathbf{x}) \\ s.t. \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p, \end{aligned} \tag{3.25}$$

where w_i is the weight of an objective typically set by the users, such that $\sum_{m=1}^{M} w_i = 1$. The relative importance of the objective is determined by the chosen proportion of this value. Although the weighted sum method is easy to apply in some scenarios, it also introduces some challenges. Firstly, a priori selection of weights does not necessarily guarantee that the final solution will be acceptable. In other words, it is difficult to set the weights to obtain the Pareto optimal solutions in a desired region in the objective space. The second problem with the weighted sum approach is that it is impossible to obtain points on non-convex portions of the Pareto optimal set in the objective space [104].

- ϵ-Constraint Method

  The ϵ-constraint method is to minimize one of the objective functions, $f_u(\mathbf{x})$, while the rest of objectives are transferred to constraints that bound within some limits, such that $f_m(\mathbf{x}) \leq \epsilon_m$, for all $m \neq u$. It can be expressed as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f_u(\mathbf{x}) \\ s.t. \quad & f_m(\mathbf{x}) \leq \epsilon_m, \quad m = 1, 2, \ldots, M, m \neq u, \\ & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p. \end{aligned} \tag{3.26}$$

  The ϵ-constraint method can be applied to either convex or non-convex problems. But the value of ϵ has to be chosen carefully so that it is within the minimum or maximum values of the individual objective function [105].

3.5 Summary

In this chapter, the required mathematical background and related techniques for this thesis are presented. Firstly, basic elements of convex optimization techniques, including convex set, convex cones, convex functions, as well as convex optimization problems have been discussed. In addition, a brief introduction of Cholesky decomposition for a symmetric and positive matrix has been provided. Next, a MOO framework with different utility functions has been discussed. Finally, for the addressed radio resource allocation optimization problems in this thesis, some classic objective functions used in this thesis have been provided. In general, this resource allocation problem has three main components, such as a utility function as the objective which can be a

performance metric, variables to be optimized and the constraints which are usually QoS requirements or some resource limitations in wireless networks.

Chapter 4

Energy Efficiency Maximization for Hybrid TDMA-NOMA System

In this chapter, an energy efficient resource allocation technique is considered for a hybrid TDMA-NOMA system. In such a hybrid system, the available time for transmission is divided into several sub-time slots, and a sub-time slot is allocated to serve a group of users (i.e., cluster). Furthermore, signals for the users in each cluster are transmitted based on the NOMA approach. In particular, an EE-Max problem is formulated to maximize the overall EE of the system with a per-user minimum rate and transmit power constraints. Simulation results demonstrate that the performance of the proposed hybrid TDMA-NOMA system with joint resource allocation outperforms the system with equal time allocation in terms of the overall EE.

4.1 System Model

A DL transmission of a hybrid TDMA-NOMA multi-user SISO system is considered. In this hybrid system, a single-antenna BS communicates with K single-antenna users, as shown in Fig. 4.1. As such, the total number of users is K, which are grouped into C clusters with a time-slot $t_i, \forall i = 1, 2, ..., C$, per cluster. Furthermore, $u_{j,i}$ represents the jth user in the ith cluster. As shown in Fig. 4.2, the available time for transmission is denoted by T. The number of users in the ith cluster, L_i, is denoted by K_i, $\forall i \in \mathcal{C} \triangleq \{1, 2, ..., C\}$, satisfying $K = \sum_{i=1}^{C} K_i$. It is assumed that signals are transmitted over a quasi-static flat Rayleigh fading channel, where the channel coefficients remain constant over each transmission block but vary independently between different blocks.

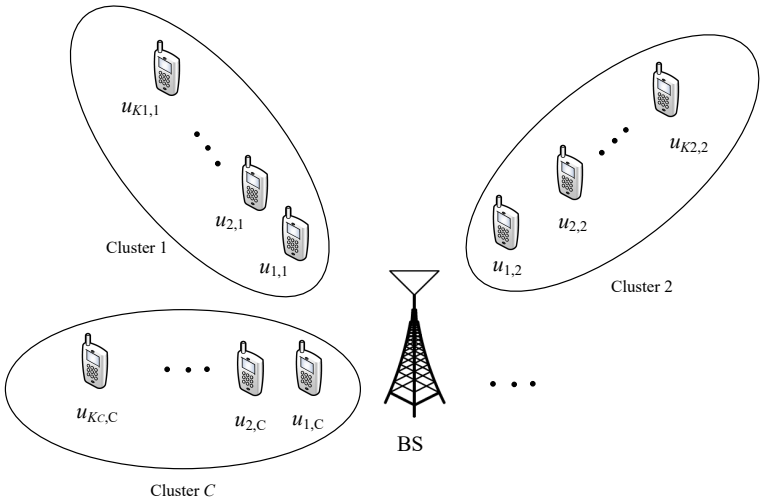

Fig. 4.1 A hybrid TDMA-NOMA multi-user SISO system.

The power assigned for $u_{j,i}$ is denoted as $p_{j,i}^2$, and thus, define the total transmit power at the BS by P_t, such that $P_t = \sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2$. The maximum transmit power available at the BS is P^{max}; then, the total transmit power constraint can be mathematically expressed as

$$P_t = \sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}. \tag{4.1}$$

As the power-domain NOMA technique is applied among the users in each cluster, the symbol transmitted from the BS during t_i can be written as

$$x_i = \sum_{j=1}^{K_i} p_{j,i} x_{j,i}, \tag{4.2}$$

where $x_{j,i}$ is the message intended to $u_{j,i}$. Accordingly, at user $u_{j,i}$, the received signal is given by

$$r_{j,i} = h_{j,i} x_i + n_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i \triangleq \{1, 2, ..., K_i\}, \tag{4.3}$$

where $h_{j,i}$ denotes the Rayleigh fading channel coefficient between the BS and the $u_{j,i}$, and $n_{j,i} \sim \mathcal{CN}(0, \sigma_{j,i}^2)$ represents the AWGN at receiver. Note that it is assumed that perfect CSI is available at the BS. The corresponding channel gain is defined as

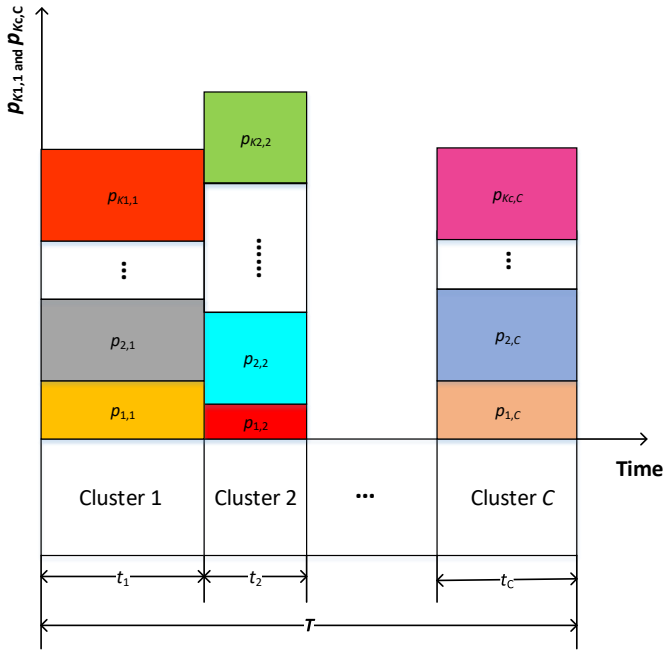

Fig. 4.2 A time-slot is assigned to serve each cluster, while the users in each cluster communicate with the BS based on the power-domain NOMA.

$|h_{j,i}|^2 = \frac{\beta}{(d_{j,i}/d_0)^\kappa}$ [106], where $d_{j,i}$ and d_0 are the distances between $u_{j,i}$ and the BS, and a reference distance, respectively. Furthermore, denote the signal attenuation at the reference distance, d_0, by β and the path loss exponent by κ. Without loss of generality, the channel gains for the users at each cluster are assumed to be ordered as

$$|h_{1,i}|^2 \geq |h_{2,i}|^2 \geq ...|h_{K_i,i}|^2, \forall i \in \mathcal{C}. \tag{4.4}$$

Accordingly, the SIC process is implemented at stronger users, i.e., users with higher channel strengths. In particular, the user $u_{j,i}$ aims to cancel the interference from any other weaker users from $u_{j+1,i}$ to $u_{K_i,i}$ using SIC. It is assumed that SIC is implemented perfectly without any errors. Therefore, the SINR at $u_{j,i}$ to decode the message of weaker users $u_{d,i}, \forall d \in \{j+1, j+2, ..., K_i\}$, can be defined as

$$SINR^d_{j,i} = \frac{|h_{j,i}|^2 p^2_{d,i}}{|h_{j,i}|^2 \sum_{s=1}^{d-1} p^2_{s,i} + \sigma^2_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{4.5}$$

To successful perform the SIC process, the received SINR levels of the users with weaker channel strengths should be higher than the users with stronger channel strengths [47]. This requirement can be only satisfied by imposing the following constraint on

the power allocation [107]:

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}. \tag{4.6}$$

The necessary power constraints for efficient SIC can be given by (4.6), which is referred to as the SIC constraint throughout this chapter; this has been widely adopted in several NOMA downlink transmissions [38, 80]. Thus, the received signal at $u_{j,i}$ after employing SIC can be expressed as

$$r_{j,i}^{SIC} = h_{j,i}p_{j,i}x_{j,i} + h_{j,i}\sum_{s=1}^{j-1} p_{s,i}x_{s,i} + n_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.7}$$

Therefore, the SINR of $u_{j,i}$ can be given by [40]

$$SINR_{j,i} = \min\{SINR_{j,i}^1, SINR_{j,i}^2, ..., SINR_{j,i}^j\}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.8}$$

Then, the achieved rate $R_{j,i}$ at $u_{j,i}$ can be given by

$$R_{j,i} = Bt_i \log_2\left(1 + SINR_{j,i}\right), \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.9}$$

where B is the available bandwidth of the channel. It is worthy to note that grouping (i.e., clustering) strategy plays a crucial role in the performance of the TDMA-NOMA system. Therefore, the proposed user grouping strategy will be discussed in the following.

4.2 Power Consumption Model

The total power consumption at the BS can be defined as [41]

$$P_{total} = \frac{1}{\epsilon}P_t + P_{loss}, \tag{4.10}$$

where $\epsilon \in [0, 1]$ denotes the efficiency of the power amplifier. Furthermore, P_{loss} represents the total power losses, and it can be expressed as [41]

$$P_{loss} = P_{dyn} + P_{sta}, \tag{4.11}$$

where P_{dyn} is the dynamic power consumption and P_{sta} is the static power consumption required to maintain the system [63].

The GEE can be defined as the ratio between the achieved sum-rate and the corresponding power consumption [39, 41]. Considering the hybrid TDMA-NOMA system, the overall EE, i.e., GEE, can be given by

$$GEE = \frac{\sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}}{P_{total}}. \tag{4.12}$$

4.3 Problem Formulation

By taking into account the importance of energy efficient resource allocation, an GEE-Max optimization design is developed for the hybrid TDMA-NOMA system. In particular, the aim is to maximize the overall system EE, i.e., GEE, while satisfying a set of relevant constraints.

Note that the GEE function presented in (4.12) is fractional in nature; as such, maximizing GEE can be viewed as a joint optimization of the achieved sum-rate (maximization) and the corresponding power consumption (minimization). In this chapter, an GEE-Max design is proposed to maximize EE under a set of constraints, including QoS requirements, as well as total time and power constraints at the BS. With these constraints, the GEE-Max optimization problem for the TDMA-NOMA system can be formulated as follows:

$$\textbf{(P1)}: \quad \max_{\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}} \quad \frac{\sum_{i=1}^{C} \sum_{j=1}^{K_i} t_i \log_2\left(1 + SINR_{j,i}\right)}{\frac{1}{\epsilon}\sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}} \tag{4.13}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \leq T, \tag{4.14}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}, \tag{4.15}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{4.16}$$

$$R_{j,i} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.17}$$

where $\bar{R}_{j,i}$ is the minimum rate requirement associated with the QoS constraint for user $u_{j,i}$, and the constraint in (4.14) ensures the maximum time constraint. Furthermore, the constraint in (4.16) is necessary for the successful application of the SIC technique [38].

In particular, there are several challenges associated with solving **P1**, which are summarized in the following discussion. Firstly, unlike equal time allocation considered

in previous works in the literature including [40], the joint allocations of the time and power resources introduce additional complexity to solve the problem and evaluate the corresponding design parameters. Secondly, as it can be seen in **P1**, the objective function is not only non-convex, but also fractional. This fractional non-convex nature of the objective function introduces additional complexity in solving the optimization problem. Thirdly, due to the constraints in (4.16) and (4.17), the optimization problem **P1** might turn out to be infeasible when the minimum rate requirements cannot be achieved with the available power budget at the BS. Hence, determining the solution for **P1** should take all these issues into account. Note that this optimization problem is significantly different from those solved in other related work, such as in [80], in terms of both objective function and constraints. Therefore, a comprehensive algorithm to solve this problem is provided in the following section. Before presenting the detailed steps of the proposed algorithm, a method is first provided to validate the feasibility of the problem in the following section.

4.4 Feasibility Analysis of the GEE-Max Problem

Now, the feasibility issues of the GEE-Max problem **P1** is discussed. Towards this end, some light is shed on the total time constraint in (4.14) with the following Lemma:

Lemma 1 *The condition* $\sum_{i=1}^{C} t_i = T$ *is necessary for achieving the maximum EE for the optimization problem* **P1**.

Proof: This lemma is proven by contradiction. First, it is assumed that $\sum_{i=1}^{C} t_i = T$ does not hold for an optimal solution $\mathbf{T}^* = [t_1^*, t_2^*, ..., t_C^*]$, that is, $\sum_{i=1}^{C} t_i < T$. Now, a new solution $\mathbf{T}_{new} = [t_1^*, t_2^* + \left(T^* - \sum_{i=1}^{C} t_i^*\right), t_3^*, ..., t_C^*]$ [95] is constructed. Obviously, this new solution still satisfies the time allocation constraint given in (4.14). Additionally, cluster L_2 has a larger throughput than that of solution $\mathbf{T}^*$ since $R_{j,i}(\mathbf{T}, p_{j,i})$ is a strictly monotonically increasing function with respect to t_i, and clusters $L_1, L_3, ..., L_C$ achieve the same throughputs as that obtained with $\mathbf{T}^*$. Therefore, $\mathbf{T}_{new}$ yields better throughputs than $\mathbf{T}^*$. This contradicts the initial assumption that $\mathbf{T}^*$ is optimal and therefore $\sum_{i=1}^{C} t_i = T$ should be satisfied. This completes the proof of Lemma 1.

Based on Lemma 1, the time allocation constraint is transformed to the following equality constraint to reduce the feasible region of the original problem **P1**:

$$\sum_{i=1}^{C} t_i = T. \tag{4.18}$$

Provided a solution set $\{p_{j,i}{}^*, t_i{}^*\}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$ is feasible, then, the minimum rate constraint in (4.17) is automatically fulfilled. Obviously, **P1** turns out to be infeasible if the corresponding minimum rate constraints cannot be met with the total power constraint in (4.15). Therefore, this infeasibility issue can be examined through evaluating the required minimum power to satisfy the corresponding QoS constraints, as follows:

$$\textbf{(P-Min)}: \quad P^{min} = \min_{\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \tag{4.19}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i = T, \tag{4.20}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq \cdots \geq p_{1,i}^2, \quad \forall i \in \mathcal{C}, \tag{4.21}$$

$$SINR_{j,i} \geq 2^{\frac{\bar{R}_{j,i}}{t_i}} - 1, \quad \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.22}$$

where P^{min} is the minimum total transmit power that is required to meet the user data rate requirements. It indicates that the BS has insufficient power budget to achieve the user data rate requirements when $P^{min} > P^{max}$. Under this condition, the optimization problem **P1** is classified as an infeasible problem. To handle this infeasibility issue, an alternative resource allocation problem, namely sum-rate maximization (**SR-Max**) problem, is considered. With this technique, the maximum achievable sum-rate is investigated under available power and SIC constraints. This problem can be formulated as

$$\textbf{(SR-Max)}: \quad \max_{\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \sum_{i=1}^{C}\sum_{j=1}^{K_i} R_{j,i} \tag{4.23}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i = T, \tag{4.24}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq \cdots \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{4.25}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.26}$$

The solution of the optimization problem **SR-Max** can be accessed using the SCA technique. In the following section, two algorithms are developed to solve the original GEE-Max problem **P1**.

4.5 Proposed Methodology

In this section, two algorithms are provided to solve the non-convex optimization problem **P1**, which are based on the SCA technique and Dinkelbach's algorithm (DA). Note that the solution of the original GEE-Max problem **P1** depends on the users that are selected for each cluster. Hence, it is necessary to determine the grouping strategy in the considered hybrid TDMA-NOMA system.

4.5.1 Grouping Strategy

Obviously, the optimal user grouping sets can be determined through an exhaustive search or combinatorial search algorithms among all possible user sets. However, this exhaustive search is impractical due to its computational complexity. Furthermore, there are several factors that should be considered when choosing a clustering algorithm, which are summarized in the following discussion. Firstly, grouping users should consider the objective function of the original design problem. In particular, for a given hybrid TDMA-NOMA system, the user clusters for the SR-Max design should be different from those of the GEE-Max design. Secondly, it has been proven in the literature that SIC is successfully implemented with relatively small error when the gap between channel strengths of the users is as high as possible [108]. This imposes that users with diverse channel strengths should be grouped into the same cluster. Considering the above key facts, and similar to the grouping strategy proposed in [80] and [109], a clustering algorithm based on the difference between the channel strengths of the users is employed. In particular, users with higher channel strengths' gaps are grouped into the same cluster. Clusters with only two users have been considered due to practical implementation challenges, including high computational complexity and potential propagation in SIC. With this restriction, the grouping sets of the hybrid TDMA-NOMA can be presented as

$$(\{u_{1,1}, u_{2,1}\}, \{u_{1,2}, u_{2,2}\}, ...\{u_{1,C}, u_{2,C}\}) \equiv \left(\{u_1, u_K\}, \{u_2, u_{K-1}\}, ...\{u_{\frac{K}{2}}, u_{\frac{K}{2}+1}\}\right), \tag{4.27}$$

where u_1 is the strongest user, while u_K is the weakest user from all users in the considered system. With this grouping strategy, two algorithms are developed to solve the non-convex optimization problem **P1** in the following subsections.

4.5.2 Sequential Convex Approximation (SCA) - based Approach

The SCA technique is a local optimization method for evaluating the solutions of non-convex problems [110]. The key idea behind this iterative approach is to approximate the non-convex functions into convex ones, and then solve iteratively the approximated convex optimization problems. It is worthy to mention that SCA is heuristic; therefore, the solutions generally depend on the initializations [110].

It is obvious that the optimization problem **P1** is non-convex due to both non-convex objective functions and constraints. Thus, this non-convexity issue can be solved through employing the SCA technique. Starting with an approximation of the non-convex objective function. A slack variable γ is introduced to approximate the objective function with a convex one. With this slack variable, the optimization problem **P1** can be rewritten as,

$$\textbf{(P2)}: \quad \max_{\gamma,\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \gamma \tag{4.28}$$

$$\text{s.t.} \quad \frac{\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i \log_2(1+SINR_{j,i})}{\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}} \geq \sqrt{\gamma}, \tag{4.29}$$

$$\sum_{i=1}^{C} t_i = T, \tag{4.30}$$

$$(4.15), (4.16), (4.17). \tag{4.31}$$

Note that the objective function of the original optimization problem **P1** is replaced with the slack variable $\sqrt{\gamma}$ by using epigraph. However, the non-convex constraint in (4.29) should be approximated by a convex constraints such that problem **P2** turns out to be a convex problem. To handle these non-convexity issues, the SCA technique is exploited through introducing additional slack variables. The details of these approximations are provided in the following.

Firstly, by incorporating a positive slack variable z, the constraint in (4.29) can equivalently be decomposed into the following two constraints:

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i \log_2(1+SINR_{j,i}) \geq \sqrt{\gamma z}, \tag{4.32}$$

$$\sqrt{z} \geq \frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}. \tag{4.33}$$

Now, the non-convexity of (4.32) is dealt by introducing new slack variables $\alpha_{j,i}$ and $\vartheta_{j,i}$ as follows:

$$(1 + SINR_{j,i}) \geq \alpha_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, ..., K_i\}, \tag{4.34a}$$

$$\log_2(1 + SINR_{j,i}) \geq \vartheta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.34b}$$

$$\alpha_{j,i} \geq 2^{\vartheta_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.34c}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} t_i \vartheta_{j,i} \geq \sqrt{\gamma z}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.34d}$$

Note that the constraint in (4.34c) is convex while the rest of the constraints in (4.34a), (4.34b), (4.34d) are not. To overcome these non-convexity issues, another slack variable $\eta_{j,i}$ is introduced, such that the constraint in (4.34a) can be rewritten as

$$\frac{|h_{j,i}|^2 p_{d,i}^2}{|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2} \geq \frac{(\alpha_{j,i} - 1)\eta_{j,i}^2}{\eta_{j,i}^2}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{4.35}$$

Accordingly, the constraint in (4.35) can now be decomposed into the following two constraints:

$$|h_{j,i}|^2 p_{d,i}^2 \geq (\alpha_{j,i} - 1)\eta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{4.36a}$$

$$|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2 \leq \eta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{4.36b}$$

Then, based on the approximation of first-order Taylor series expansion method, the constraint in (4.36a) can be represented as

$$\begin{aligned} |h_{j,i}|^2 \left({p_{d,i}^2}^{(t)} + 2p_{d,i}^{(t)}(p_{d,i} - p_{d,i}^{(t)}) \right) &\geq {\eta_{j,i}^2}^{(t)} \left(\alpha_{j,i}^{(t)} - 1 \right) \\ +2 \left(\alpha_{j,i}^{(t)} - 1 \right) \eta_{j,i}^{(t)} \left(\eta_{j,i} - {\eta_{j,i}}^{(t)} \right) + {\eta_{j,i}^2}^{(t)} \left(\alpha_{j,i} - {\alpha_{j,i}}^{(t)} \right), \\ \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \end{aligned} \tag{4.37}$$

where $p_{d,i}^{(t)}$, ${\eta_{j,i}}^{(t)}$ and $\alpha_{j,i}^{(t)}$ represent the approximations of $p_{d,i}$, $\eta_{j,i}$ and $\alpha_{j,i}$ at the tth iteration, respectively. Note that both sides of (4.37) are linear in terms of the optimization variables, i.e., $p_{d,i}$, $\eta_{j,i}$, and $\alpha_{j,i}$. Furthermore, the constraint in (4.36b)

can be rewritten as the following SOC constraint:

$$\|\|h_{j,i}|p_{1,i}, |h_{j,i}|p_{2,i}, ..., |h_{j,i}|p_{d-1,i}, \sigma_{j,i}\| \leq \eta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{4.38}$$

where $||\cdot||$ denotes the Euclidean norm of a vector. With these approximations, the constraint in (4.34a) can be rewritten as the convex constraints in (4.37) and (4.38).

Now, the non-convexity issue of the constraint in (4.34d) is considered. Similar to the previous approximations, the non-convex constraint in (4.34d) with a new slack variable $\nu_{j,i}$ can be rewritten as

$$t_i \vartheta_{j,i} \geq \nu_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.39a}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} \nu_{j,i} \geq \sqrt{\gamma z}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.39b}$$

To deal with the non-convexity issue of (4.39a), a non-negative term $t_i^2 + \vartheta_{j,i}^2$ is added to both sides of inequality (4.39a) without loss of generality. By taking the square root of both sides of the inequality, the following SOC constraint can be defined:

$$t_i + \vartheta_{j,i} \geq \left\| \begin{matrix} 2\sqrt{\nu_{j,i}} \\ t_i - \vartheta_{j,i} \end{matrix} \right\|, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.40a}$$

Furthermore, the left-side of (4.39b) is approximated with a lower-bounded convex approximation using the first-order Taylor series expansion. As such, (4.39b) can be reformulated as

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} \nu_{j,i} \geq \sqrt{\gamma^{(t)} z^{(t)}} + \frac{1}{2} \sqrt{\left(\frac{z^{(t)}}{\gamma^{(t)}}\right)} (\gamma - \gamma^{(t)}) + \frac{1}{2} \sqrt{\left(\frac{\gamma^{(t)}}{z^{(t)}}\right)} (z - z^{(t)}). \tag{4.40b}$$

Similar to the previous approximations, the non-convexity of the constraint in (4.33) can be dealt with by introducing a new slack variable $\widetilde{z}$,

$$\sqrt{z} \geq \widetilde{z}, \tag{4.41a}$$

$$\widetilde{z} \geq \frac{1}{\epsilon} \sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}. \tag{4.41b}$$

Now, a SOC relaxation is exploited to cast the non-convex constraints in (4.41) with convex ones [111] as follows:

$$\frac{z+1}{2} \geq \left|\left| \begin{matrix} \frac{z-1}{2} \\ \widetilde{z} \end{matrix} \right|\right|, \tag{4.42a}$$

$$\frac{\epsilon(\widetilde{z}-P_{loss})+1}{2} \geq \left|\left| \begin{matrix} \frac{\epsilon(\widetilde{z}-P_{loss})-1}{2} \\ p_{1,i} \\ ... \\ p_{K,i} \end{matrix} \right|\right|, \forall i \in \mathcal{C}. \tag{4.42b}$$

Considering the aforementioned approximations, the constraint in (4.29) can be equivalently rewritten as a set of the following convex constraints: (4.34c), (4.37), (4.38), (4.40a), (4.40b), (4.42a) and (4.42b). Accordingly, the minimum rate constraints can be redefined as the following convex constraints:

$$\nu_{j,i} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.43}$$

Next, the non-convexity issue of (4.16) can be handled by approximating each non-convex term in the inequality by a lower-bounded convex term using the first-order Taylor series expansion. With this approximation, each term in (4.16) can be equivalently written as

$$p_{K,i}^2 \geq {p_{K,i}^2}^{(t)} + 2{p_{K,i}}^{(t)}(p_{K,i} - {p_{K,i}}^{(t)}), \forall i \in \mathcal{C}. \tag{4.44}$$

With the above relaxations, the original non-convex optimization problem **P1** can be equivalently written as the following approximated convex one:

$$\textbf{(P3)}: \quad \max_{\Gamma} \quad \gamma \tag{4.45}$$

$$\text{s.t.} \quad (4.34c), (4.37), (4.38), (4.40a), (4.40b), (4.42a), (4.42b), \tag{4.46}$$

$$(4.30), (4.15), (4.44), (4.43), \tag{4.47}$$

where Γ comprises all the optimization variables, such that $\Gamma = \{p_{j,i}, t_i, \gamma, \alpha_{j,i}, \vartheta_{j,i}, \eta_{j,i}, z, \widetilde{z}, \nu_{j,i}\}$, $\forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$. More specifically, the solution of **P1** is obtained through iteratively solving the approximated convex optimization problem **P3** and updating initialized variables. In particular, the solution of each iteration is fed into the optimization problem **P3** to update the corresponding initial parameter for the convergence iteration. However, the initial parameters of the first iteration have to be carefully selected to guarantee the success of the iterative algorithm. In particular, for this

Table 4.1 GEE-Max Joint Resource Allocation Algorithm.

Algorithm 1: SCA method to solve GEE-Max Problem.

1: Group the users into clusters based on the grouping strategy defined in (4.27)
2: Initialize: Set the parameters $\Gamma^{(0)}$
3: Repeat
4: Solve the problem **P3** in (4.45) - (4.47)
5: Update all parameters $\Gamma^{(t)}$
6: Until $|\gamma^{*(t+1)} - \gamma^{*(t)}| \leq \tau$.

iterative SCA algorithm, selecting initial optimization parameters, $\Gamma^{(0)}$, has a considerable impact on both efficiency of the solution and convergence of the algorithm itself. Hence, the initialization of variables should be carefully defined. These initial values can be chosen by defining random power allocations $p_{j,i}^{(0)}$ that fulfill the constraints in the problem **P-min**. Then, the corresponding slack variables are determined by substituting these power allocations in (4.34), (4.37), (4.38), (4.40) and (4.42). Consequently, the solutions obtained in each iteration are used as initialized variables to solve the optimization problem **P3** in the subsequent iteration. Note that the objective function of the optimization problem **P3**, i.e., γ, is a non-decreasing function [4]. Therefore, a solution of the SCA algorithm can be selected as a reasonable solution for the original optimization problem if the difference between two consecutive solutions is less than a pre-defined threshold, τ. This stopping criteria of the proposed iterative algorithm can be mathematically given by $|\gamma^{*(t+1)} - \gamma^{*(t)}| \leq \tau$. The proposed iterative SCA technique based algorithm to solve **P1** is summarized in Table 4.1.

It is worth pointing out that the convergence of the proposed SCA-based approach to solve the GEE-Max problem has been carefully investigated in [112], where it has been stated that the SCA guarantees at least a local optimal solution, and in most cases, a global optimal solution. On the other hand, the required aspects to guarantee the convergence of the SCA technique in Algorithm 1 are discussed in the following. Firstly, the iterative algorithm should be initialized with appropriate initial parameters $\Gamma^{(0)}$, which ensures the feasibility of the problem at each iteration. It can be realized that the solutions returned at the iteration t are also feasible solutions for the problem at the successive iteration $t+1$. This implies that Algorithm 1 yields a non-decreasing sequence of the objective values, i.e., $\gamma^{(t+1)} > \gamma^{(t)}$. In addition, the total transmit power at BS is limited by an upper bound of P^{max}, which confirms that γ will converge to the solution with a finite number of iterations.

4.5.3 Dinkelbach's algorithm (DA) - based Approach

In this subsection, an alternative approach based on DA is developed to solve the original GEE-Max problem. This approach not only validates the solution obtained through the SCA algorithm, but also provides an alternative technique to deal with the fractional nature of the objective function in the original optimization problem **P1**. With DA, a new non-negative variable λ is introduced to parametrize the fractional objective function into a non-fractional one [113]. Based on the variable λ, the problem **P1** can be defined as follows:

$$\textbf{(P4)}: \max_{\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}} \sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i \log_2\left(1+SINR_{j,i}\right) - \lambda\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}\right) \tag{4.48}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i^{(t)} = T, \tag{4.49}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} {p_{j,i}^2}^{(t)} \leq P^{max}, \tag{4.50}$$

$${p_{K,i}^2}^{(t)} \geq {p_{K-1,i}^2}^{(t)} \geq ... \geq {p_{1,i}^2}^{(t)}, \forall i \in \mathcal{C}, \tag{4.51}$$

$$R_{j,i}^{(t)} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.52}$$

Obviously, the objective function is convex with respect to λ. Then, the following theorem presents the solution to the problem **P4**.

Theorem 1 *[113] The optimal objective value of **P4** equals to zero, i.e.,*

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i^* \log_2(1+SINR_{j,i}^*) - \lambda^*\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} {p_{j,i}^*}^2 + P_{loss}\right) = 0, \tag{4.53}$$

*where $\{t_i^*, p_{j,i}^*, \lambda^*\}, \forall i, \forall j$ denote the corresponding optimal solutions for **P4**.*

With Theorem 1, the solution of the original fractional problem **P1** (i.e., $\{t_i^*, p_{j,i}^*\}, \forall i, \forall j$) can be determined by solving the non-fractional optimization problem **P4**, where the optimal objective value of **P4** is zero [113]. The proof of Theorem 1 can be found in [113].

According to Theorem 1, the fractional objective function can now be transformed into a subtractive form, and thus, obtaining the variables $p_{j,i}, t_i$ that maximize the GEE in the original problem **P1** is equivalent to solving the parameterized optimization

problem **P4**. Therefore, starting to initialize the parameter λ with zero, then use the convex approximation techniques to solve the parameterized optimization problem **P4** [113]. Then, update λ in the tth iteration can be written as follows:

$$\lambda^{(t)} = \frac{\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i^{(t-1)} \log_2(1+SINR_{j,i}^{(t-1)})}{\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^{2\ (t-1)} + P_{loss}}. \tag{4.54}$$

In particular, the variables $t^{(t)}, p_{j,i}^{(t)}$ in the tth iteration can be found by solving the following optimization problem:

$$\textbf{(P5)}: \max_{\{p_{j,i}^{(t)}, t_i^{(t)}\}_{i=1\ j=1}^{C\ K_i}} \sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i^{(t)} \log_2(1+SINR_{j,i}^{(t)}) - \lambda^{(t-1)} \left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^{2\ (t)} + P_{loss}\right) \tag{4.55}$$

$$\text{s.t.} \qquad (4.49), (4.50), (4.51), (4.52). \tag{4.56}$$

Now, the following observations are highlighted by comparing the optimization problems **P1** and **P5**. Firstly, note that the parametrization carried out using DA deals with the fractional nature of the objective function, as seen in **P5**. However, it can be shown that the first part of the objective function in **P5** still remains non-convex because the optimization variables are coupled. Secondly, the convex approximations implemented in **P3** can be utilized to deal with the non-convex constraints in **P5**.

Considering the above, to deal with the non-convexity issue of the objective function of **P5**, the same approach that has been developed to approximate the constraints in problem **P3** is exploited, such as the SCA technique, through introducing a set of slack variables as follows:

$$t_i \log_2(1+SINR_{j,i}) \geq y_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.57a}$$

$$(1+SINR_{j,i}^{d}) \geq \beta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, ..., K_i\}, \tag{4.57b}$$

$$\log_2(1+SINR_{j,i}) \geq \chi_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.57c}$$

$$\beta_{j,i} \geq 2^{\chi_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.57d}$$

$$t_i \chi_{j,i} \geq y_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.57e}$$

The constraint in (4.57b) can be equivalently rewritten as the following set of convex constraints:

$$\begin{aligned}|h_{j,i}|^2 \left(p_{d,i}^{2\ (t)} + 2p_{d,i}^{(t)}(p_{d,i} - p_{d,i}^{(t)})\right) \geq \theta_{j,i}^{2\ (t)} \left(\beta_{j,i}^{(t)} - 1\right) \\ +2\left(\beta_{j,i}^{(t)} - 1\right)\theta_{j,i}^{(t)}\left(\theta_{j,i} - \theta_{j,i}^{(t)}\right) + \theta_{j,i}^{2\ (t)}\left(\beta_{j,i} - \beta_{j,i}^{(t)}\right), \\ \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\},\end{aligned} \tag{4.58a}$$

$$\||h_{j,i}|p_{1,i}, |h_{j,i}|p_{2,i}, ..., |h_{j,i}|p_{d-1,i}, \sigma_{j,i}\| \leq \theta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{4.58b}$$

where $\theta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$, are newly introduced variables. As can be seen, the constraints in (4.57e) are jointly convex with respect to the involved variables where the right side is an affine function and the left side is a quadratic-over-affine function [4]. The inequality (4.57e) can be formulated into a SOC constraint as follows:

$$t_i + \chi_{j,i} \geq \left\| \begin{matrix} 2\sqrt{y_{j,i}} \\ t_i - \chi_{j,i} \end{matrix} \right\|, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{4.59}$$

It is obvious that $\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}\right)$ is a convex function in terms of $p_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$. Thus, the function $\lambda\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}\right)$ is also a convex function of $p_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$, as λ is a constant and consequently $\sum_{i=1}^{C}\sum_{j=1}^{K_i} y_{j,i} - \lambda\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}\right)$ is a convex function. From these observations, the original non-convex optimization problem **P1** can be approximated using the DA as the following optimization problem:

$$\textbf{(P6)}: \quad \max_{\Phi^{(t)}} \quad \sum_{i=1}^{C}\sum_{j=1}^{K_i} y_{j,i}^{(t)} - \lambda^{(t-1)}\left(\frac{1}{\epsilon}\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^{2\ (t)} + P_{loss}\right) \tag{4.60}$$

$$\text{s.t.} \quad (4.49), (4.50), (4.51), (4.58a), (4.58b), (4.57d), (4.59), \tag{4.61}$$

$$y_{j,i}^{(t)} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{4.62}$$

where $\Phi^{(t)} = \{p_{j,i}^{(t)}, t_i^{(t)}, \theta_{j,i}^{(t)}, \chi_{j,i}^{(t)}, y_{j,i}^{(t)}\}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$. With the solution of **P6**, the involved variables are updated at the successive iteration. This iterative process is carried out until the algorithm converges. Table 4.2 summarizes the proposed iterative algorithm for solving the original problem **P1**. In this DA-based iterative algorithm, the variables are first initialized with $\lambda^{(0)}$ and $\Phi^{(0)}$. Then, for current λ, the optimal

Table 4.2 Dinkelbach's Method to Solve GEE-Max Problem.

Algorithm 2: Dinkelbach's method to solve GEE-Max Problem.

1: Initialize: $\lambda^{(0)}$ to satisfy $G(\lambda^{(0)}) \geq 0$, iteration number $t = 0$ and set the parameters $\Phi^{(0)}$
2: Repeat
3: Solve the problem **P6** with $\Phi^{(t-1)}$, then obtain the optimal Φ^*
4: Update $\Phi^{(t)} = \Phi^*$
5: Until required accuracy is achieved
6: Update $\lambda^{(t+1)}$ according to (4.54), and set $t \leftarrow t + 1$
7: Until convergence.

variables Φ^* are determined by solving the problem **P6** until the required accuracy is achieved. The proof for the convergence of the proposed algorithm is provided in Appendix.

4.5.4 Complexity Analysis of the Proposed Schemes

In this subsection, the analysis for the computational complexity of solving the original GEE-Max optimization problem **P1** is provided.

The complexity of the SCA based approach

With the SCA-based approach, the solution of the original GEE-Max optimization problem **P1** is obtained through solving the approximated optimization problem **P3**, iteratively. Therefore, the complexity of solving **P1** can be defined by quantifying the complexity of solving the approximated **P3** and the average number of required iterations, where the interior-point method is utilized to solve the SOCP with SOC and linear constraints [100, 114]. In particular, the complexity of solving the SOCP constraints at each iteration is given by $\mathcal{O} = (\mathcal{A}^2\mathcal{B})$ [114], where $\mathcal{A}$ and $\mathcal{B}$ represent the number of optimization variables and the dimensions of the SOC constraints, respectively. In fact, as **P3** is solved in a number of iterations, the overall complexity of solving the original problem is upper bounded by $\mathcal{O}\big(\mathcal{A}^2\mathcal{B}\log(\frac{1}{\tau})\big)$, where τ is the required solution accuracy.

The complexity of the Dinkelbach's algorithm (DA) based approach

In the developed DA-based algorithm, the original optimization problem is transformed into a standard SOCP problem for the given non-negative variable λ. However,

two iterative algorithms are deployed to determine the solution in the DA based approach. Thus, the upper-bound complexity of DA based approach can be defined as $\mathcal{O}\big(\mathcal{A}^2\mathcal{B}\log(\frac{1}{\tau})\log(\frac{1}{\varpi})\big)$, where ϖ represents the required solution accuracy.

4.6 Simulation Results

In this section, simulation results are provided to demonstrate the effectiveness of the proposed joint GEE-Max design for the considered hybrid TDMA-NOMA system. Additionally, the performance of the proposed schemes have been compared with that of the other baseline designs, namely, **SR-Max** and **P-Min** designs. In particular, the performance of the proposed GEE-Max design is evaluated with opportunistic time allocations against schemes with equal time allocations.

In the simulations, a hybrid TDMA-NOMA system is considered with 10 users divided into 5 clusters. The users are assumed to be uniformly distributed over a circle area with a radius of 10 meters around the BS, where the minimum distance d_0 is selected 1 meter ($d_0 = 1$ m), where d_0 is the reference distance. The corresponding channel gain is $|h_{j,i}|^2 = \frac{\beta}{(d_{j,i}/d_0)^\kappa}$, where $\beta = -30$ dB and $\kappa = 2$. The noise variance at each user $\sigma^2_{j,i}$ depends on the noise power spectral density N_0 and the channel bandwidth B, which is expressed as $\sigma^2_{j,i} = N_0 B$. In these simulations, N_0 is assumed to be -70 dBm/Hz and the bandwidth B is set to 1 MHz. T is chosen to be 10 seconds. The power amplifier efficiency ϵ for both algorithms is 0.35, the static power consumption is P_{sta} =15 dBm, and P_{dyn} =10 dBm [99]. In addition, the stopping-criteria threshold for both algorithms is set to 0.01 [38]. Furthermore, the CVX software is used to solve the convex problems in these simulations [4].

Fig. 4.3 compares the EE of our proposed design with the existing conventional designs in the literature, namely the resource allocation techniques with **P-Min** and **SR-Max** in hybrid TDMA-NOMA system. As seen in Fig. 4.3, the proposed GEE-Max based design outperforms the conventional design criteria of **P-Min** and **SR-Max** in terms of achieved EE. In addition, the EE of the **SR-Max** based design is not monotonically increasing with the available power and decreases when the transmit power exceeds a certain available power budget. This is due to the fact that this design fully uses all the available power for maximizing the achievable sum rate instead of maximizing the EE. In other words, maximizing the achievable sum-rate does not always maximize the EE. It can also be observed that the **P-Min** based design achieves lower EE than the proposed scheme. This is due to the fact that the **P-Min** design seeks for the minimum power that is required to achieve the minimum rate requirements.

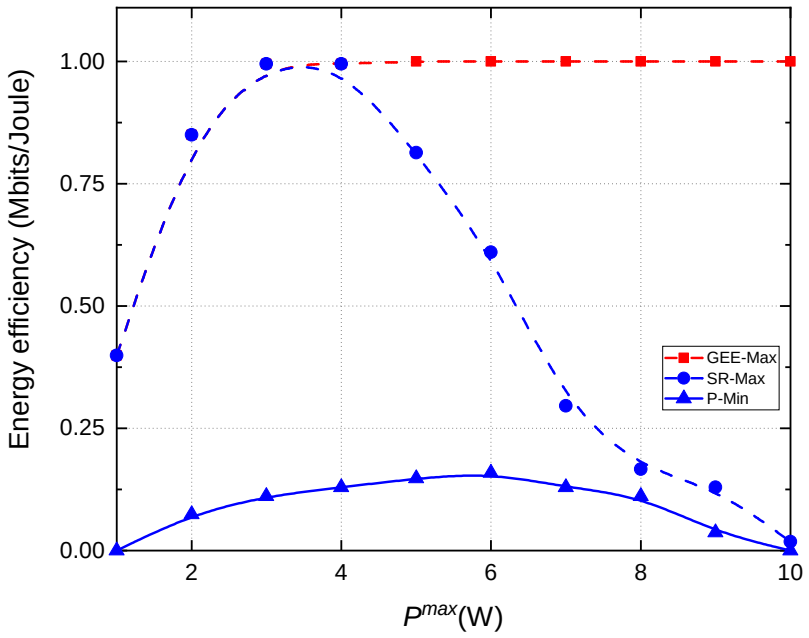

Fig. 4.3 Energy efficiency of the hybrid TDMA-NOMA system with different design criteria, $\bar{R}_{j,i} = 5bits/s/Hz$.

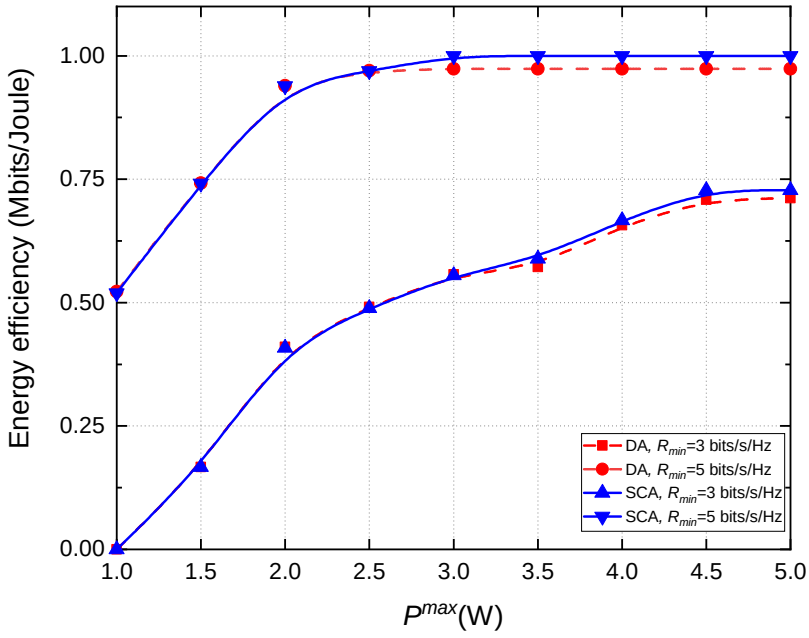

Fig. 4.4 Energy efficiency of the proposed algorithm with different QoS requirements.

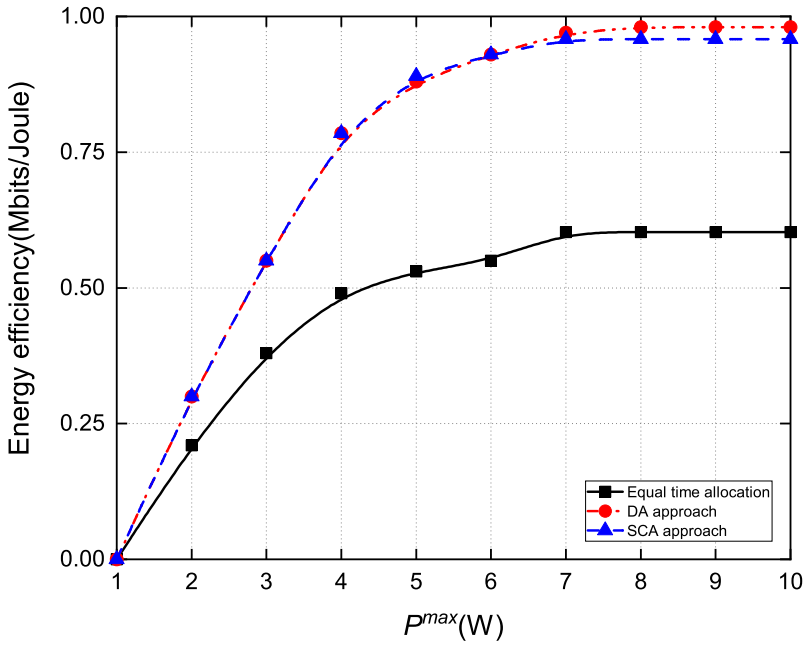

Fig. 4.5 Energy efficiency of the proposed algorithm and equal time allocation scheme with different transmit power, $\bar{R}_{j,i} = 3bits/s/Hz$.

Fig. 4.4 depicts the average EE versus transmit power for different QoS requirements using the algorithms developed through the SCA and DA techniques. It can be observed that the EE of both algorithms first increases until reaches a certain value, and then it remains constant after a certain maximum power P^{max}. The EE performance with $R_{min} = 5$ bits/s/Hz [1] is better than that with $R_{min} = 3$ bits/s/Hz. This performance difference can be justified through the following argument. With higher R_{min}, the increasing rate in the sum rate is higher than the increasing rate in the total power consumption, which results in an improvement of EE. Note that the performance gap between these two algorithms is not significant in terms of the achieved EE. By setting the stopping-criteria to zero (i.e., $\lambda^{(t+1)} - \lambda^{(t)} = 0$), then both approaches should achieve the same performance.

Next, Fig. 4.5 illustrates the achieved EE against different transmit power levels for the proposed scheme with opportunistic time allocation and the conventional schemes with equal time allocation. As seen in Fig. 4.5, the achieved EE of the GEE-Max design with opportunistic time allocation outperforms that of the conventional equal time allocation. This can be achieved by solving the GEE-Max problem using either

[1] Note that R_{min} and $\bar{R}_{j,i}$ carry the same meaning

Table 4.3 Power Allocations For Each User In The Hybrid TDMA-NOMA Through The Proposed Opportunistic Time Allocations And The Conventional Equal Time One.

	Channels	$p_{1,1}$	$p_{2,1}$	$p_{1,2}$	$p_{2,2}$	$p_{1,3}$	$p_{2,3}$	$p_{1,4}$	$p_{2,4}$	$p_{1,5}$	$p_{2,5}$
Scheme with opportunistic time allocations	Channel 1	9.407	2.000	8.678	2.000	7.443	2.000	2.461	2.000	2.000	2.000
	Channel 2	1.2509	2.6832	1.2509	2.6832	1.2509	2.5327	1.2898	2.2687	1.3310	2.1908
	Channel 3	0.8979	2.1111	0.8979	2.1111	0.8979	1.9323	0.8979	1.6852	0.8979	1.5376
	Channel 4	1.1093	2.6235	1.1093	2.6235	1.1093	2.5285	1.1093	2.4644	1.1093	2.4644
	Channel 5	0.8393	1.8862	0.8393	1.8433	0.8393	1.7023	0.8393	1.6952	1.0169	1.5410
Scheme with equal time allocations	Channel 1	7.517	2.000	7.517	2.000	7.091	2.000	3.293	2.000	2.687	2.000
	Channel 2	1.1546	3.2741	1.1546	2.8266	1.1546	2.1453	1.1546	1.9478	1.1855	1.8550
	Channel 3	0.9145	2.6055	0.9145	2.3931	0.9145	1.8665	0.9145	1.3254	0.9145	1.2887
	Channel 4	1.1637	3.1452	1.1637	2.9782	1.1637	2.3582	1.1637	2.1550	1.1637	2.1295
	Channel 5	0.8568	2.2729	0.8568	1.9871	0.8568	1.5232	0.8568	1.4743	0.9514	1.3963

Table 4.4 Time Allocation And Achieved Minimum Throughout In The Hybrid TDMA-NOMA And The Conventional Schemes.

	Scheme with opportunistic time allocations						Scheme with equal time allocations					
Channels	$t_1(s)$	$t_2(s)$	$t_3(s)$	$t_4(s)$	$t_5(s)$	EE (Mbits/Joule)	$t_1(s)$	$t_2(s)$	$t_3(s)$	$t_4(s)$	$t_5(s)$	EE (Mbits/Joule)
Channel 1	2.254	2.329	2.118	1.980	1.516	0.356	2	2	2	2	2	0.327
Channel 2	1.751	1.819	2.097	2.114	2.219	0.275	2	2	2	2	2	0.251
Channel 3	2.737	2.399	1.892	1.425	1.547	0.490	2	2	2	2	2	0.438
Channel 4	2.599	2.388	1.783	1.628	1.601	0.304	2	2	2	2	2	0.284
Channel 5	2.634	2.220	1.695	1.630	1.819	0.563	2	2	2	2	2	0.533

the SCA or the DA algorithm, as shown in Fig. 4.5. In these algorithms, both time and power resources are utilized efficiently to achieve the best EE for a given system.

Next, the achieved per-user power allocations and per-cluster time allocations in the hybrid TDMA-NOMA schemes with the opportunistic time allocations versus that of the conventional schemes with equal time allocation are presented in Table 4.3 and 4.4, respectively. For the sake of comparison, it is assumed that both schemes use the same minimum data rate requirements ($\bar{R}_{j,i} = 2$ bits/s/Hz). The achieved rate of each user and the time allocations using the proposed opportunistic time allocation are given for five different sets of random channels. As seen in Table 4.3, most users achieve better rates in our proposed opportunistic time allocations based hybrid TDMA-NOMA schemes when compared with the conventional scheme with equal time allocations.

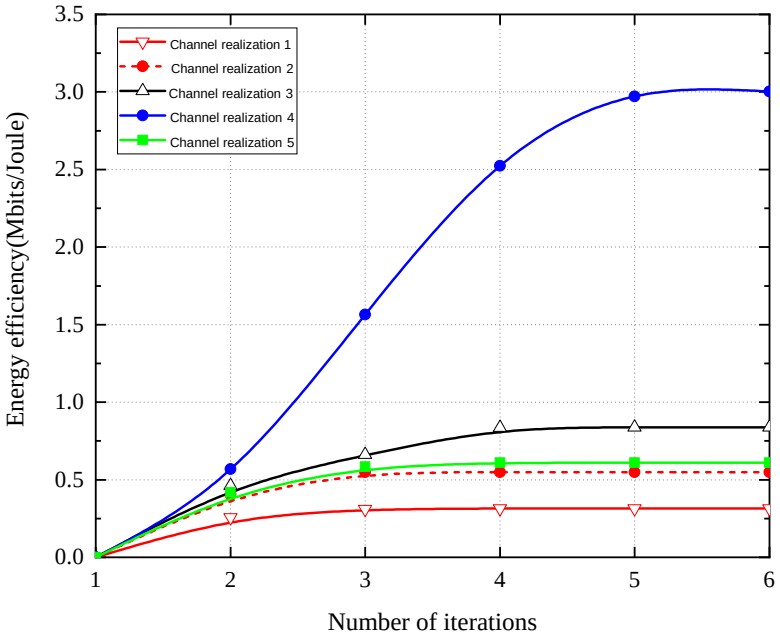

Fig. 4.6 The convergence of the SCA algorithm for five different sets of channels.

The convergence of SCA and DA-based EE maximization algorithms is studied in Fig. 4.6 and Fig. 4.7, respectively. In particular, five different sets of channels are considered to evaluate the convergence. In these simulations, the maximum transmit power P^{max} is set to 10 W. As seen in Fig. 4.6, the SCA-based algorithm converges to the optimal EE faster by using the relaxation of constraints. The convergence of DA follows the same procedure as for the convergence of the SCA technique for each λ. In addition, the simulation results confirm that both algorithms converge within a few number of iterations.

4.7 Summary

In this chapter, the GEE-Max problem with joint power-time resource allocation has been studied for a hybrid TDMA-NOMA system. In particular, the users are grouped into a number of clusters, the available transmission time is divided into several time-slots, and the power-domain NOMA is exploited to serve multiple users within each cluster. However, due to the non-convexity of the formulated GEE-Max optimization problem, two different algorithms based on the SCA and DA techniques have been proposed, respectively. Simulation results have demonstrated the effectiveness of the

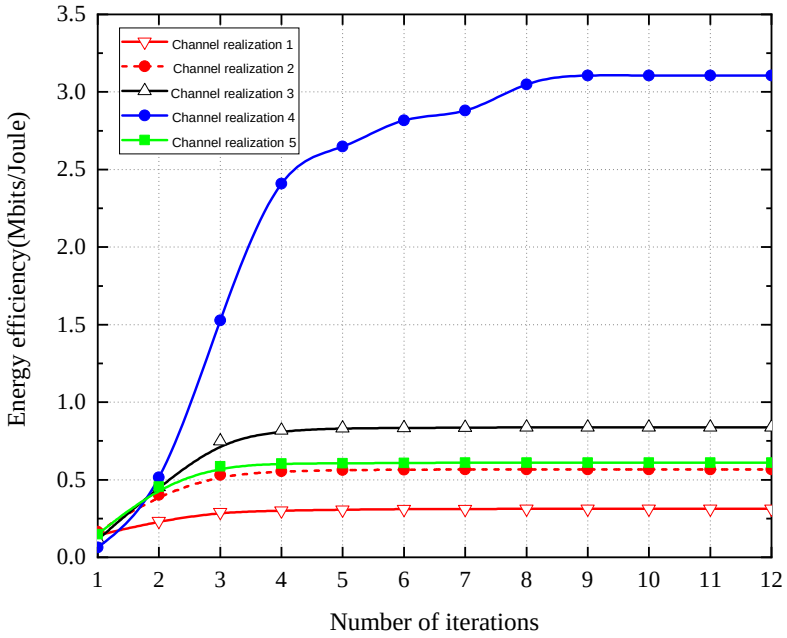

Fig. 4.7 The convergence of the DA for five different sets of channels.

proposed schemes. In particular, the proposed hybrid TDMA-NOMA system with opportunistic time allocation outperforms the conventional resource allocations with equal time assignment in terms of the required minimum transmit power and achieved sum rate. In other words, the proposed schemes achieve better EE than the conventional schemes with equal time allocations.

4.8 Appendix

In order to prove the convergence of the DA-based iterative approach to the optimal solution, the following conditions can be equivalently proven [113]:

- Firstly, it is proven that $\lambda^{(t+1)} > \lambda^{(t)}$ for all t.

  Lemma 2 *Let* $G(\lambda^{(t)}) = \max_{\{t_i^*, p_{j,i}^*\}} \left\{ \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i, p_{j,i}) - \lambda^{(t)} P_{total}(p_{j,i}) \right\}$, *then* $G(\lambda)$ *is a strictly monotonic decreasing function of* λ, *i.e., if* $\lambda^{(t)} < \lambda^{(t+1)}$, *then* $G(\lambda^{(t)}) > G(\lambda^{(t+1)})$.

Proof: Let $\{t_i^*, p_{j,i}^*\}$ be the optimal power allocation and time slot assignment for the proposed schemes for $G(\lambda^{(t+1)})$. Then

$$\begin{aligned} G(\lambda^{(t+1)}) &= \max_{\{t_i^*, p_{j,i}^*\}} \left\{ \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i, p_{j,i}) - \lambda^{(t+1)} P_{total}(p_{j,i}) \right\} \\ &= \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i^*, p_{j,i}^*) - \lambda^{(t+1)} P_{total}(p_{j,i}^*) \\ &< \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i^*, p_{j,i}^*) - \lambda^{(t)} P_{total}(p_{j,i}^*) \\ &\leq \max_{t_i, p_{j,i}} \left\{ \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i, p_{j,i}) - \lambda^{(t)} P_{total}(p_{j,i}) \right\} \\ &= G(\lambda^{(t)}). \end{aligned} \tag{4.63}$$

This completes the proof of Lemma 2.

Lemma 3 *Let $\{t_i, p_{j,i}\}$ be an arbitrary power allocation and time slot assignment and $\lambda^{(t+1)} = \frac{\sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(p_{j,i}^{(t+1)}, t_i^{(t+1)})}{P_{total}(p_{j,i}^{(t+1)})}$, then $G(\lambda^{(t+1)}) \geq 0$.*

Proof: $G(\lambda^{(t+1)}) = \max_{\{t_i, p_{j,i}\}} \left\{ \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i, p_{j,i}) - \lambda^{(t+1)} P_{total}(p_{j,i}) \right\}$ $\geq \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i^{(t+1)}, p_{j,i}^{(t+1)}) - \lambda^{(t+1)} P_{total}(p_{j,i}^{(t+1)}) = 0$. Hence, $G(\lambda^{(t+1)}) \geq 0$. This completes the proof of Lemma 3 and this lemma implies that $G(\lambda^{(t)}) \geq 0$. By definition,

$$\begin{aligned} G(\lambda^{(t)}) &= \sum_{i=1}^{C} \sum_{j=1}^{K_i} R_{j,i}(t_i{}^{(t)}, p_{j,i}^{(t)}) - \lambda^{(t)} P_{total}(p_{j,i}^{(t)}) \\ &= \lambda^{(t+1)} P_{total}(p_{j,i}^{(t)}) - \lambda^{(t)} P_{total}(p_{j,i}^{(t)}) > 0. \end{aligned} \tag{4.64}$$

Since $P_{total}(p_{j,i}^{(t)}) > 0$, then $\lambda^{(t+1)} > \lambda^{(t)}$.

- Secondly, it is proven that $\lim_{t\to\infty} \lambda^{(t)} = \lambda^*$, where λ^* is the maximum EE and the $\lambda^{(t)}$ equals to λ^* when iteration number t approaches to infinity. It is proven by contradiction. Assume that $\lim_{t\to\infty} \lambda^{(t)} = \lambda^*$ does not hold, that is, $\lim_{t\to\infty} \lambda^{(t)} = \widetilde{\lambda} < \lambda^*$. Based on this argument, $G(\widetilde{\lambda}) = 0$. However, $G(\lambda)$ is a strictly monotonic decreasing function based on Lemma 2, and therefore it is obtained

$$0 = G(\widetilde{\lambda}) > G(\lambda^*) = 0, \tag{4.65}$$

which contradicts the initial assumption. Therefore, this confirms that the Dinkelbach's algorithm based method converges to the optimal solution. ■

Chapter 5

Resource Allocation Techniques for Hybrid TDMA-NOMA System

In this chapter, two resource allocation techniques for hybrid TDMA-NOMA system are proposed. While recent resource allocation techniques aimed to individually maximize either SE or EE, this chapter firstly considers an SE-EE trade-off based technique for a hybrid TDMA-NOMA system. The original problem has been formulated as an MOO problem with the conflicting objective functions SE and EE. Then, a max-min problem is investigated to maintain a user fairness in terms of the achieved rates in the hybrid TDMA-NOMA system. The performance of the proposed designs are evaluated and compared with that of conventional designs in the literature.

5.1 Spectral-Energy Efficiency Trade-Off Based Design

5.1.1 Introduction

To meet the requirements of future wireless networks, several resource allocation techniques have already been proposed for hybrid TDMA-NOMA systems. SE-Max has been investigated in [94] and the work in [63] has considered the EE-Max based resource allocation technique. The EE and SE are conflicting performance metrics. Optimizing SE degrades the overall EE, provided the available transmit power is more than the green power [69]. The maximum GEE in **P1** is achieved within certain available power budget, which is referred to as the green power. Similarly, EE maximization does not offer maximum SE.

An SE-EE trade-off based design has been proposed for a MISO-NOMA system in [81]. In [82], the SE-EE trade-off for massive MIMO systems has been investigated through the particle swarm optimization algorithm. The EE-SE trade-off has been studied for the IRS-aided multi-user MIMO uplink system with the partial CSI in [115]. The solution to MOO problem can be achieved by converting it into a SOO problem [116, 117].

Motivated by the importance of both SE and EE in future wireless networks, an SE-EE trade-off based resource allocation technique is considered for a hybrid TDMA-NOMA system in this chapter. The hybrid TDMA-NOMA system with single-antenna has potential capabilities to achieve better performance and meet different requirements in specific scenarios compared to the conventional stand-alone NOMA or TDMA designs, including some practical applications, such as M2M communications [55], UAV communications [57], and IoT networks [118]. Unlike the existing stand-alone SE or EE resource allocation techniques in hybrid TDMA-NOMA system, the proposed design aims to strike a good balance between those performance metrics while fulfilling the requirements of future wireless networks. The SE-EE trade-off based design is formulated as a MOO problem, and the weighted sum utility function is utilized to reformulate the MOO framework as a SOO problem. Then, an iterative approach is proposed to solve the SOO problem.

5.1.2 Problem Formulation

The DL hybrid TDMA-NOMA system has been introduced in the previous chapter. Based on the definition of the achieved rate $R_{j,i}$ at $u_{j,i}$, SE can be defined as $SE = \frac{\sum_{i=1}^{C}\sum_{j=1}^{K_i} R_{j,i}}{B}$. Since it is assumed that $B = 1$ throughout this thesis, both SE and sum-rate carry the same meaning throughout this thesis.

Note that the objective functions of EE-Max and SE-Max are conflicting in nature. In particular, maximizing the sum rate in EE-Max might degrade the EE of the system. Similarly, maximizing EE through solving SE-Max has a negative impact on the achieved sum-rate. To overcome this conflicting issue and to align with different requirements of both users and service providers, an SE-EE trade-off based design is proposed for the hybrid TDMA-NOMA system.

The objective function of this MOO design consists of the conflicting performance metrics, i.e., SE and EE. For simplicity, SE and EE are represented by the functions $f_1(\{p_{j,i}, t_i\}_{i=1\,j=1}^{C\;\;K_i})$ and $f_2(\{p_{j,i}, t_i\}_{i=1\,j=1}^{C\;\;K_i})$, respectively. In fact, the objective function of this trade-off based design can be defined as a vector function $\mathbf{f}$, with the elements of

both performance metrics f_1 and f_2. Accordingly, the proposed trade-off based design can be formulated as

$$\textbf{(P7)}: \quad \max_{\{p_{j,i},t_i\}_{i=1\,j=1}^{C\;\;K_i}} \quad \mathbf{f}\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\;\;K_i}\right) \tag{5.1}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \leq T, \tag{5.2}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}, \tag{5.3}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.4}$$

$$R_{j,i} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.5}$$

where

$$\mathbf{f}(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\;\;K_i}) = [f_1(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\;\;K_i}), f_2(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\;\;K_i})]. \tag{5.6}$$

There are several challenges associated with solving **P7**. Firstly, it is essential to identify the users that have to be grouped and served at each sub-time slot as different solutions can be obtained with different grouping strategies. Secondly, once the grouping strategy is determined, a feasibility check has to be carried out prior to solving the problem. This is due to the fact that the minimum rate constraints in **P7** cannot be satisfied for some power budget at BS. Finally, given a multi-objective function in **P7**, the conventional approaches cannot be directly employed to determine its feasible solution. Thus, a solution approach is proposed to address all of these issues in the following subsection.

5.1.3 Proposed Methodology

The solution of the original problem **P7** depends on the selected users for each cluster. Hence, it is important to determine an appropriate grouping strategy in the considered hybrid TDMA-NOMA system. The optimal grouping strategy can be only determined through exhaustive search [119], which has a high computational complexity. To reduce the computational complexity, different sub-optimal grouping strategies have been considered in the literature [97, 120].

Feasibility Analysis

The feasibility of **P7** can be investigated with the grouping strategy. Note that **P7** might turn out to be infeasible due to the limited total power constraint in (5.3). Therefore, it is important to firstly examine the required minimum transmit power to fulfill these minimum rate constraints. It can be evaluated by solving the following problem:

$$\textbf{(P-Min)}: \qquad P^{min} = \min_{\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \tag{5.7}$$

$$\text{s.t.} \qquad \sum_{i=1}^{C} t_i \leq T, \tag{5.8}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.9}$$

$$SINR_{j,i} \geq 2^{\frac{R_{j,i}^{min}}{t_i}} - 1, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.10}$$

where P^{min} is the minimum total transmit power required to meet the minimum rate requirements. The **P7** is feasible, and thus worthy to solve if $P^{min} \leq P^{max}$. When $P^{min} > P^{max}$, **P7** turns out to be infeasible. In this chapter, it is assumed that an alternative SE-Max design is considered if **P7** is infeasible.

The Proposed Algorithm

Given that **P7** is feasible, an approach is proposed to solve this MOO problem. Note that no single unique optimal solution exists to simultaneously optimize both $f_1(\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i})$ and $f_2(\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i})$. Therefore, the set of the best trade-off solutions should be determined, referred to as the Pareto-optimal solutions [121]. A feasible solution $\{p_{j,i}^*,t_i^*\}$ is considered to be a Perto-optimal solution if no other solution exists such that $f(\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}) \preceq f(\{p_{j,i}^*,t_i^*\}_{i=1}^{C}{}_{j=1}^{K_i})$ [104]. To determine the Pareto-optimal solution, the multi-objective function should be firstly replaced with a single-objective function [104, 121]. The weighted-sum utility function is selected to determine the Pareto-optimal solution [104]. A weight factor $\alpha_i \in [0,1]$ is assigned to the ith objective function to reflect its relative importance on the overall design, and the sum of both weighted functions is considered. The SOO framework to represent **P7** can be formulated as

$$\textbf{(P8)}: \quad \max_{\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}} \quad \sum_{l=1}^{2} \alpha_l f_l^{Norm}\left(\{p_{j,i}\},\{t_i\}_{i=1\,j=1}^{C\quad K_i}\right) \tag{5.11}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \leq T, \tag{5.12}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}, \tag{5.13}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.14}$$

$$R_{j,i} \geq \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.15}$$

where $f_1^{Norm}(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i})$ and $f_2^{Norm}(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i})$ are the normalized versions of $f_1(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i})$ and $f_2(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i})$, respectively. These can be expressed as

$$f_1^{Norm}\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right) = \frac{f_1\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right)}{f_1^*}, \tag{5.16a}$$

$$f_2^{Norm}\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right) = \frac{f_2\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right)}{f_2^*}, \tag{5.16b}$$

where f_1^* and f_2^* are the maximum values of SE and EE, respectively. With such normalization, a non-dimensional objective function with a unity upper bound is obtained. For simplicity, let $\alpha_2 = \alpha$ and $\alpha_1 = 1-\alpha$. Note that **P8** is non-convex problem. Therefore, the SCA technique is exploited to deal with its non-convexity issue.

Sequential Convex Approximation (SCA)

The SCA technique is a local optimization method for evaluating the solutions of non-convex problems, and it has been utilized to obtain the solutions for several optimization frameworks in wireless communications [40, 81]. Firstly, the objective function can be solved by introducing two slack variables γ_1 and γ_2 such that

$$(1-\alpha) f_1^{Norm}\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right) \geq \gamma_1, \tag{5.17a}$$

$$\alpha f_2^{Norm}\left(\{p_{j,i},t_i\}_{i=1\,j=1}^{C\quad K_i}\right) \geq \gamma_2. \tag{5.17b}$$

With γ_1 and γ_2, **P8** can be equivalently written as

$$\textbf{(P9)}: \quad \max_{\gamma_1,\gamma_2,\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \gamma_1+\gamma_2 \tag{5.18}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \le T, \tag{5.19}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \le P^{max}, \tag{5.20}$$

$$p_{K,i}^2 \ge p_{K-1,i}^2 \ge ... \ge p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.21}$$

$$R_{j,i} \ge \bar{R}_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.22}$$

$$(1-\alpha) f_1^{Norm}\left(\{p_{j,i}\}, \{t_i\}_{i=1}^{C}{}_{j=1}^{K_i}\right) \ge \gamma_1, \tag{5.23}$$

$$\alpha f_2^{Norm}\left(\{p_{j,i}\}, \{t_i\}_{i=1}^{C}{}_{j=1}^{K_i}\right) \ge \gamma_2. \tag{5.24}$$

Note that the objective function in **P9** is linear in terms of γ_1 and γ_2. However, the non-convex constraints in (5.23) and (5.24) are introduced to **P9**. The SCA is exploited to approximate the non-convex terms.

Firstly, the constraint in (5.23) can be rewritten as

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i \log_2\left(1+SINR_{j,i}\right) \ge \frac{f_1^*}{1-\alpha}\gamma_1. \tag{5.25}$$

To deal with the non-convexity of (5.25), slack variables $z_{j,i}$ and $\chi_{j,i}$ are introduced, such that

$$(1+SINR_{j,i}) \ge z_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, ..., K_i\}, \tag{5.26a}$$

$$\log_2(1+SINR_{j,i}) \ge \chi_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.26b}$$

$$z_{j,i} \ge 2^{\chi_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.26c}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} t_i\chi_{j,i} \ge \frac{f_1^*}{1-\alpha}\gamma_1, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.26d}$$

Note that the constraint in (5.26c) is convex while the others are not. To overcome the non-convexity issues of (5.26a), another slack variable $\xi_{j,i}$ is introduced, such that

(5.26a) can be rewritten as

$$\frac{|h_{j,i}|^2 p_{d,i}^2}{|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2} \geq \frac{(z_{j,i}-1)\xi_{j,i}^2}{\xi_{j,i}^2}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.27}$$

Furthermore, the constraint in (5.27) can now be decomposed into two constraints as follows:

$$|h_{j,i}|^2 p_{d,i}^2 \geq (z_{j,i}-1)\xi_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{5.28a}$$

$$|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2 \leq \xi_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.28b}$$

Then, the first-order Taylor series expansion is exploited to approximate both sides of (5.28a) with their corresponding linear approximations, such that

$$\begin{aligned} &|h_{j,i}|^2 \left({p_{d,i}^2}^{(n)} + 2p_{d,i}^{(n)}(p_{d,i} - p_{d,i}^{(n)}) \right) \geq {\xi_{j,i}^2}^{(n)} \left(z_{j,i}^{(n)} - 1 \right) \\ &+ 2\left(z_{j,i}^{(n)} - 1 \right) \xi_{j,i}^{(n)} \left(\xi_{j,i} - {\xi_{j,i}}^{(n)} \right) + {\xi_{j,i}^2}^{(n)} \left(z_{j,i} - {z_{j,i}}^{(n)} \right), \\ &\forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \end{aligned} \tag{5.29}$$

where $p_{d,i}^{(n)}$, ${\xi_{j,i}}^{(n)}$ and $z_{j,i}^{(n)}$ represent the approximations of $p_{d,i}$, $\xi_{j,i}$ and $z_{j,i}$ at the nth iteration, respectively. Note that both sides of (5.29) are now linear in terms of $p_{d,i}$, $\xi_{j,i}$, and $z_{j,i}$. Furthermore, the constraint in (5.28b) can be rewritten as the following SOC constraint:

$$\| |h_{j,i}|p_{1,i}, |h_{j,i}|p_{2,i}, ..., |h_{j,i}|p_{d-1,i}, \sigma_{j,i} \| \leq \xi_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{5.30}$$

where $||\cdot||$ denotes the Euclidean norm of a vector. With these approximations, (5.26a) can now be approximated as the convex constraints in (5.29) and (5.30).

Next, the non-convexity issue of the constraint in (5.26d) is considered. This can be dealt by incorporating a new slack variable $\nu_{j,i}$, such that

$$t_i \chi_{j,i} \geq \nu_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.31a}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} \nu_{j,i} \geq \frac{f_1^*}{1-\alpha} \gamma_1, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.31b}$$

To tackle the non-convexity issue of (5.31a), the non negative $t_i^2 + \chi_{j,i}^2$ is incorporated to the both sides of inequality (5.31a) without affecting the original inequality. This constraint can be now formulated as the following SOC constraint:

$$t_i + \chi_{j,i} \geq \left\| \begin{matrix} 2\sqrt{\nu_{j,i}} \\ t_i - \chi_{j,i} \end{matrix} \right\|_2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.32}$$

To this end, the non-convex constraint in (5.23) is replaced with the following convex constraints:

$$(5.23) \Leftrightarrow (5.29), (5.30), (5.26c), (5.32), (5.31b).$$

Similarly, the non-convexity of the constraint in (5.24) is tackled by introducing a new slack variable b such that

$$\frac{\sum_{i=1}^{C} \sum_{j=1}^{K_i} t_i \log_2 (1 + SINR_{j,i})}{\frac{1}{\epsilon} \sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}} \geq \frac{\gamma_2 f_2^* b^2}{\alpha b^2}. \tag{5.33}$$

The constraint in (5.33) can be split into the following two constraints:

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} t_i \log_2 (1 + SINR_{j,i}) \geq \frac{f_2^*}{\alpha} \gamma_2 b^2, \tag{5.34}$$

$$b^2 \geq \frac{1}{\epsilon} \sum_{i=1}^{C} \sum_{j=1}^{K_i} p_{j,i}^2 + P_{loss}. \tag{5.35}$$

To handle the non-convexity issue of (5.34), the same approaches that were used for constraint in (5.23) is exploited. A set of new slack variables $\varpi_{j,i}$, $\upsilon_{j,i}$, $\delta_{j,i}$ and $\beta_{j,i}$ are introduced, such that

$$(1 + SINR_{j,i}) \geq \upsilon_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, ..., K_i\}, \tag{5.36a}$$

$$\upsilon_{j,i} \geq 2^{\varpi_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.36b}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} t_i \varpi_{j,i} \geq \frac{f_2^*}{\alpha} \gamma_2 b^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.36c}$$

The constraint in (5.36a) can be written as

$$\frac{|h_{j,i}|^2 p_{d,i}^2}{|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2} \geq \frac{(\upsilon_{j,i} - 1)\delta_{j,i}^2}{\delta_{j,i}^2}. \tag{5.37}$$

Following a similar approach,

$$|h_{j,i}|^2 p_{d,i}^2 \geq (\upsilon_{j,i} - 1)\delta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{5.38a}$$

$$|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2 \leq \delta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.38b}$$

The inequalities in (5.38) can now be approximated with linear function using the first-order Taylor series approximation as

$$\begin{aligned} &|h_{j,i}|^2 \left({p_{d,i}^2}^{(n)} + 2p_{d,i}^{(n)}(p_{d,i} - p_{d,i}^{(n)}) \right) \geq {\delta_{j,i}^2}^{(n)} \left(\upsilon_{j,i}^{(n)} - 1 \right) \\ &+ 2\left(\upsilon_{j,i}^{(n)} - 1 \right) \delta_{j,i}^{(n)} \left(\delta_{j,i} - {\delta_{j,i}}^{(n)} \right) + {\delta_{j,i}^2}^{(n)} \left(\upsilon_{j,i} - {\upsilon_{j,i}}^{(n)} \right), \\ &\forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \end{aligned} \tag{5.39a}$$

$$\| |h_{j,i}|p_{1,i}, |h_{j,i}|p_{2,i}, ..., |h_{j,i}|p_{d-1,i}, \sigma_{j,i} \| \leq \delta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.39b}$$

The constraint in (5.36c) can be reformulated with the following convex constraints:

$$t_i \varpi_{j,i} \geq \beta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.40a}$$

$$t_i + \varpi_{j,i} \geq \left\| \begin{matrix} 2\sqrt{\beta_{j,i}} \\ t_i - \varpi_{j,i} \end{matrix} \right\|_2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.40b}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} \beta_{j,i} \geq \frac{f_2^*}{\alpha} \gamma_2 b^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.40c}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{K_i} \beta_{j,i} \geq \frac{f_2^*}{\alpha} \left({b^2}^{(n)} {\gamma_2}^{(n)} + 2{\gamma_2}^{(n)} b^{(n)} \left(b - b^{(n)} \right) + {b^2}^{(n)} \left(\gamma_2 - {\gamma_2}^{(n)} \right) \right), \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.40d}$$

Following a similar approach in (5.28b), the constraint in (5.35) can be cast as the following SOC constraint:

$$b \geq \frac{1}{\epsilon} \left\| \left[p_{1,i}, p_{2,i}, ..., p_{d-1,i}, \sqrt{P_{loss}} \right]^T \right\|_2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.41}$$

To this end, the non-convex constraint in (5.24) is replaced with the following convex constraints:

$$(5.24) \Leftrightarrow (5.39a), (5.39b), (5.36b), (5.40b), (5.40d), (5.41).$$

The non-convexity issue of (5.9) can be dealt by approximating each non-convex term of the inequality by a lower bounded convex term using the first-order Taylor series. Each term in (5.9) can be written as

$$p_{K,i}^2 \geq {p_{K,i}^2}^{(n)} + 2{p_{K,i}}^{(n)} \left(p_{K,i} - {p_{K,i}}^{(n)}\right), \forall i \in \mathcal{C}. \tag{5.42}$$

Finally, the minimum rate constraints in (5.10) can be formulated as the following convex constraints:

$$\nu_{j,i} \geq R_{j,i}^{min}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.43}$$

With the above approximations, **P7** can be equivalently written as the following approximated convex one:

$$\begin{aligned}
\textbf{(P10)}: \quad & \max_{\Gamma} && \gamma_1 + \gamma_2 && (5.44)\\
& \text{s.t.} && (5.2), (5.3), (5.42), (5.43), && (5.45)\\
& && (5.29), (5.30), (5.26c), (5.32), (5.31b), && (5.46)\\
& && (5.39a), (5.39b), (5.36b), (5.40b), (5.40d), (5.41), && (5.47)
\end{aligned}$$

where Γ consists of all the optimization parameters, such that $\Gamma = \{p_{j,i}, t_i, \gamma_1, \gamma_2, \beta_{j,i}, \chi_{j,i}, \xi_{j,i}, z_{j,i}, \nu_{j,i}, \upsilon_{j,i}, \varpi_{j,i}, \delta_{j,i}, b\}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$. In fact, the solution to **P7** can be obtained by iteratively solving **P10**. With this iterative algorithm, the initial value of $\Gamma^{(0)}$ needs to be carefully selected as it plays a crucial role in determining the solution. Therefore, a simplified approach is discussed to select these initial values. Firstly, an appropriate initial power allocation and time slots are selected to fulfill all the constraints of **P10**. Then, the corresponding slack variables can be evaluated based on the initial power allocation and time slots. Note that the iterative process is continued until the required accuracy which can be defined as the difference between two consecutive objective values (i.e., $|\Phi^*(\Gamma^{(n+1)}) - \Phi^*(\Gamma^{(n)})| \leq \varsigma$, where $\Phi(\Gamma^{(n)}) = \gamma_1(\Gamma^{(n)}) + \gamma_2(\Gamma^{(n)})$ and ς is the predefined threshold. The proposed iterative algorithm to solve the original problem **P7** is summarized in Table 5.1.

Table 5.1 SE-EE Trade-Off Resource Allocation Algorithm.

Algorithm 1: SCA method to solve SE-EE trade-off design problem.

1: Group the users into clusters based on the grouping strategy
2: Check the feasibility of the problem by solving problem **P-Min**
3: Initialize: Set the parameters $\Gamma^{(0)}$ with initial values
4: Repeat
5: Solve the problem **P10** in (5.44) - (5.47)
6: Update $\Gamma^{(n)}$
7: Until required accuracy is achieved.

5.2 Max-min Fairness Design with Opportunistic Time Assignment

To overcome the practical challenges of employing SIC in dense networks, and to meet the unprecedented requirements of future wireless networks, NOMA has been integrated with different other technologies, including multiple-antenna techniques [36, 74, 75, 122] and conventional OMA schemes [40, 41, 43, 68, 79, 93, 123]. In a hybrid OMA-NOMA systems, the available resources (i.e., time or frequency) are divided into several sub-resource blocks and each sub-resource block is assigned to serve multiple users based on NOMA [40, 41]. For example, a hybrid TDMA-NOMA system has been considered in [40], in which the available time for transmission is divided equally among several groups of users (i.e., clusters), and the energy harvesting capabilities of such system is investigated. A hybrid OFDMA-NOMA system is considered in [41], where the available bandwidth is divided into several sub-bandwidths, and the available resources are allocated to maximize the energy efficiency of the system. In [93] and [123], different resource allocation techniques for hybrid OFDMA-NOMA systems are developed. In fact, these combinations not only simplify the implementation of SIC, but also offers additional degrees of freedom by utilizing different domains to serve multiple users. Considering the hybrid TDMA-NOMA system, the work in the literature assume equal time assignments to serve the available groups of users to reduce computational burden at the receiver ends. However, this equal time assignments limit the performance enhancement of such a hybrid TDMA-NOMA system owing the fact that opportunistic time allocations provides additional benefits to the groups of users. Furthermore, serving each user in such hybrid TDMA-NOMA systems to achieve reasonable throughput is one of the key objectives of such systems. However, maximizing overall throughput of the system degrades the performance of individual

users while compromising user-fairness in terms of achievable rates. Motivated by these facts, an max-min problem is considered, which aiming to minimum per user rate while satisfying the relevant constraints on the system.

5.2.1 Problem Formulation

Based on the aforementioned facts, user-fairness is one of the crucial requirements for 5G and beyond wireless networks. Hence, the problem is studied to allocate the available resources, i.e., time and transmit power, among the users to maintain a user fairness in terms of the achieved rates in the hybrid TDMA-NOMA system. In particular, the objective is to maximize the minimum achieved rate at the individual users. This can be accomplished by solving the following max-min optimization problem:

$$\textbf{(P11)}: \quad \max_{\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \min_{1\le j\le K_i, 1\le i\le C} R_{j,i} \tag{5.48}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \le T, \tag{5.49}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} {p_{j,i}}^2 \le P^{max}, \tag{5.50}$$

$$p_{K,i}^2 \ge p_{K-1,i}^2 \ge ... \ge p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.51}$$

where the constraint in (5.49) ensures that the total allocated time does not exceed the available time T for transmission. Furthermore, the constraint in (5.51) facilitates the successful implementation of SIC. However, additional complexity is introduced due to the joint allocation of both the time and the transmit power to all served users in the system. Furthermore, the non-convex objective function makes the original problem defined in (5.49) - (5.51) more challenging to solve. Hence, an iterative algorithm is developed to realize the solution in the next subsection.

The grouping strategy is first discussed to group the users into a number of clusters in the hybrid TDMA-NOMA system. Next, approximation techniques are used to transform the non-convex optimization problem P1 into an approximated convex problem.

5.3 Proposed Methodology

The objective function of problem **P11** is non-convex; hence, a new slack variable γ is firstly introduced to approximate it into a convex one. Based on this slack variable, problem **P11** can be reformulated equivalently as,

$$\textbf{(P12)}: \quad \max_{\gamma,\{p_{j,i},t_i\}_{i=1}^{C}{}_{j=1}^{K_i}} \quad \gamma \tag{5.52}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \leq T, \tag{5.53}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} {p_{j,i}}^2 \leq P^{max}, \tag{5.54}$$

$$p_{K,i}^2 \geq p_{K-1,i}^2 \geq ... \geq p_{1,i}^2, \forall i \in \mathcal{C}, \tag{5.55}$$

$$t_i \log_2 (1 + SINR_{j,i}) \geq \gamma, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.56}$$

Note that the objective function of the original optimization problem **P11** is replaced with a new single scalar slack variable by using epigraph. However, this non-convex objective function has been formulated into a constraint in (5.56) in **P12**. In other words, maximizing $\min_{1\leq j\leq K_i, 1\leq i\leq C} R_{j,i}$ is equivalent to maximizing the slack variable γ with a new constraint in (5.56). However, the overall problem still remains intractable due to the non-convex constraints (5.55) and (5.56) in **P12**. In order to solve this non-convex problem, the SCA technique is exploited, in which a set of lower bounded convex terms are introduced to approximate the non-convex terms in the constraints (5.55) and (5.56) [40].

The non-convexity of the constraint in (5.56) is handled by introducing new slack variables $\alpha_{j,i}$ and $\vartheta_{j,i}$, such that

$$(1 + SINR_{j,i}^d) \geq \alpha_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, ..., K_i\}, \tag{5.57}$$

$$\log_2(1 + SINR_{j,i}) \geq \vartheta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.58}$$

$$\alpha_{j,i} \geq 2^{\vartheta_{j,i}}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.59}$$

$$t_i \vartheta_{j,i} \geq \gamma, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \tag{5.60}$$

where the constraint in (5.59) is convex. Next, to address the non-convexity issue of the constraint in (5.57), another slack variable $\eta_{j,i}$ is introduced, such that

$$\frac{|h_{j,i}|^2 p_{d,i}^2}{|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2} \geq \frac{(\alpha_{j,i}-1)\eta_{j,i}^2}{\eta_{j,i}^2}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.61}$$

Secondly, the above constraint in (5.61) can be decomposed into two constraints as follows:

$$|h_{j,i}|^2 p_{d,i}^2 \geq (\alpha_{j,i}-1)\,\eta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \tag{5.62}$$

$$|h_{j,i}|^2 \sum_{s=1}^{d-1} p_{s,i}^2 + \sigma_{j,i}^2 \leq \eta_{j,i}^2, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.63}$$

Then, the first-order Taylor series are exploited to approximate both sides of (5.62) with linear convex terms, such that

$$\begin{aligned} &|h_{j,i}|^2 \left({p_{d,i}^2}^{(t)} + 2p_{d,i}^{(t)}(p_{d,i} - p_{d,i}^{(t)}) \right) \geq {\eta_{j,i}^2}^{(t)} \left(\alpha_{j,i}^{(t)} - 1 \right) \\ &+ 2\left(\alpha_{j,i}^{(t)} - 1 \right) \eta_{j,i}^{(t)} \left(\eta_{j,i} - {\eta_{j,i}}^{(t)} \right) + {\eta_{j,i}}^{2(t)} \left(\alpha_{j,i} - {\alpha_{j,i}}^{(t)} \right), \\ &\forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}, \end{aligned} \tag{5.64}$$

where $p_{d,i}^{(t)}$, ${\eta_{j,i}}^{(t)}$ and $\alpha_{j,i}^{(t)}$ represent the approximations of $p_{d,i}$, $\eta_{j,i}$ and $\alpha_{j,i}$ at the tth iteration, respectively. The constraint in (5.63) can be rewritten as follows using SOC [88, 124]:

$$\left\| |h_{j,i}|p_{1,i}, |h_{j,i}|p_{2,i}, ..., |h_{j,i}|p_{d-1,i}, \sigma_{j,i} \right\| \leq \eta_{j,i}, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i, \forall d \in \{j+1, j+2, ..., K_i\}. \tag{5.65}$$

Based on the above multiple slack variables $\alpha_{j,i}$ and $\eta_{j,i}$, the constraint in (5.57) can be approximated with the convex constraints in (5.64) and (5.65).

The non-convexity issue of the constraint in (5.60) can be solved by formulating it as the following SOC constraint [88, 90]:

$$t_i + \vartheta_{j,i} \geq \left\| \begin{matrix} 2\sqrt{\gamma} \\ t_i - \vartheta_{j,i} \end{matrix} \right\|, \forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i. \tag{5.66}$$

Furthermore, each non-convex term in (5.55) can be approximated by a lower bounded convex term using the first-order Taylor series expansion,

$$p_{K,i}^2 \geq {p_{K,i}^2}^{(t)} + 2{p_{K,i}}^{(t)}(p_{K,i} - {p_{K,i}}^{(t)}), \forall i \in \mathcal{C}. \tag{5.67}$$

Therefore, the original non-convex optimization problem **P11** can be approximated by the following convex optimization problem:

$$\textbf{(P13)}: \quad \max_{\Gamma} \quad \gamma \tag{5.68}$$

$$\text{s.t.} \quad \sum_{i=1}^{C} t_i \leq T, \tag{5.69}$$

$$\sum_{i=1}^{C}\sum_{j=1}^{K_i} p_{j,i}^2 \leq P^{max}, \tag{5.70}$$

$$(5.55), (5.59), (5.64), (5.65), (5.66), \tag{5.71}$$

where Γ consists of all the optimization variables, such that $\Gamma = \{p_{j,i}, t_i, \gamma, \alpha_{j,i}, \vartheta_{j,i}, \eta_{j,i}\}$, $\forall i \in \mathcal{C}, \forall j \in \mathcal{K}_i$. It is worth pointing out that the solution of **P11** is obtained iteratively, such that the approximated convex optimization problem **P13** is solved at each iteration. In particular, this requires appropriate selection of the initial variables, i.e., $\Gamma^{(0)}$. These initial values can be chosen by defining random power allocations $p_{j,i}^{(0)}$ that fulfills the maximum power constraint in (5.50). Then, the corresponding slack variables are evaluated by substituting these power allocations in (5.64) and (5.65). The solutions obtained in each iteration are used as initial points for the Taylor series approximation to the next iteration. In fact, the iterative algorithm keeps improving the solutions at each iteration until the difference between two consecutive objective values is less than a pre-defined threshold, ς, (i.e., $|\gamma^{*(n+1)} - \gamma^{*(n)}| \leq \varsigma$). The proposed iterative algorithm is summarized in Table 5.2.

5.4 Simulation Results

5.4.1 Spectral-Energy Efficiency Trade-Off Based Design

In this sub-section, simulation results are provided to demonstrate the effectiveness of the proposed SE-EE trade-off based design for the hybrid TDMA-NOMA system. For the grouping strategy, clusters with two users in each cluster have been considered, i.e., $K_i = 2$, due to practical implementation challenges, including high computa-

Table 5.2 Max-min Joint Resource Allocation Algorithm.

Algorithm: Solving Max-min Joint Resource Allocation Problem.

1: Group the users into clusters based on the grouping strategy,
2: Initialize: Set the parameters $\Gamma^{(0)}$,
3: Repeat
4: Solve the problem **P13** in (5.68) - (5.71),
5: Update all parameters $\Gamma^{(n)}$ based on (5.55),(5.59),(5.64),(5.65),(5.66),
6: Until $|\gamma^{*(n+1)} - \gamma^{*(n)}| \leq \varsigma$, where ς is a predefined error tolerance threshold.

Table 5.3 Parameter values used in the simulations

Simulation Parameter	Value(s)
Number of users (K)	10
Number of users in each cluster (K_i)	2
Distances of users (m)	$1.0 \leq d_{j,i} \leq 30.0$
Pass loss exponent(κ)	2
Noise variance of users ($\sigma^2_{j,i}$)	0.01
Power amplifier efficiency (ϵ)	0.35
Threshold of algorithm	0.01
Bandwith B (MHz)	1

tional complexity and error propagations in SIC. However, the analysis provided in this work is applicable to clusters with any number of users. Motivated by the fact that SIC can be successfully implemented when the difference of the channel gains is high, users are grouped with higher difference in their channel gains. Based on this grouping strategy, the clusters for the considered system can be presented as $(\{u_{1,1}, u_{2,1}\}, \{u_{1,2}, u_{2,2}\}, ...\{u_{1,C}, u_{2,C}\}) \equiv \left(\{u_1, u_K\}, \{u_2, u_{K-1}\}, ...\{u_{\frac{K}{2}}, u_{\frac{K}{2}+1}\}\right)$, where u_1 and u_K are the strongest and the weakest users, respectively. Table 5.3 provides simulation parameters [99, 107]. Similar to the works in the literature [107, 125], a pico-cell is considered in this simulation.

Fig. 5.1 presents the achieved SE and EE with different weight factors α. As seen in Fig. 5.1, both SE and EE remain the same when the weight factor α is small. Then, with increasing α, the SE decreases whereas the EE increases. This is due to the fact that more resources are allocated for maximizing EE as the weight factor α increases. As the weight factor α increases, more resources are allocated for maximizing EE, which provides a better performance. With an appropriate weight factor α, the BS has the flexibility to achieve different performance trade-off according to the requirements

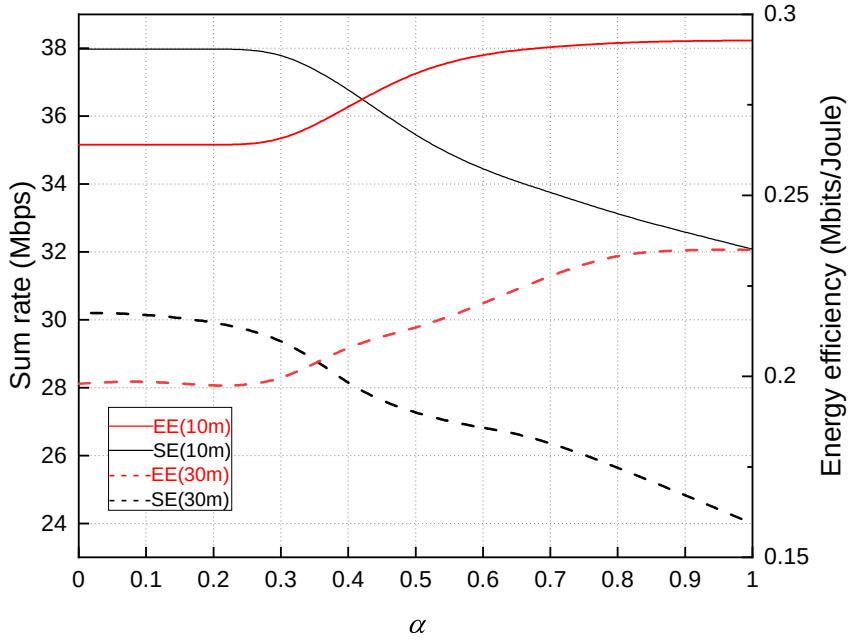

Fig. 5.1 Achieved SE and EE with different weight factors.

of the systems. Furthermore, Fig. 5.1 shows the achieved EE and SE for the proposed design against α with 10 and 30 meters radii around the BS. As expected, with the 30 meters radius, both the achieved EE and SE decrease compared to those achieved with 10 meter radius.

Fig. 5.2 and Fig. 5.3 depict the achieved SE and EE performance versus P^{max} with different α, respectively. It can be observed that the achieved SE first increases until reaching a certain value, and it then remains constant. Similarly, the performance of EE first increases until reaching the maximum value, however, it then decreases as the transmit power increases. The proposed SE-EE trade-off based design becomes the conventional SE and EE designs with $\alpha = 0$ and 1, respectively.

Finally, the SE-EE performance trade-off is illustrated in Fig. 5.4 with the set of Pareto-optimal solutions for the original optimization problem. Note that each point on this curve represents a Pareto-optimal solution for a particular α, based on sum rate and EE performance. In other words, no other solution exists to simultaneously improve both the SE and EE objective functions.

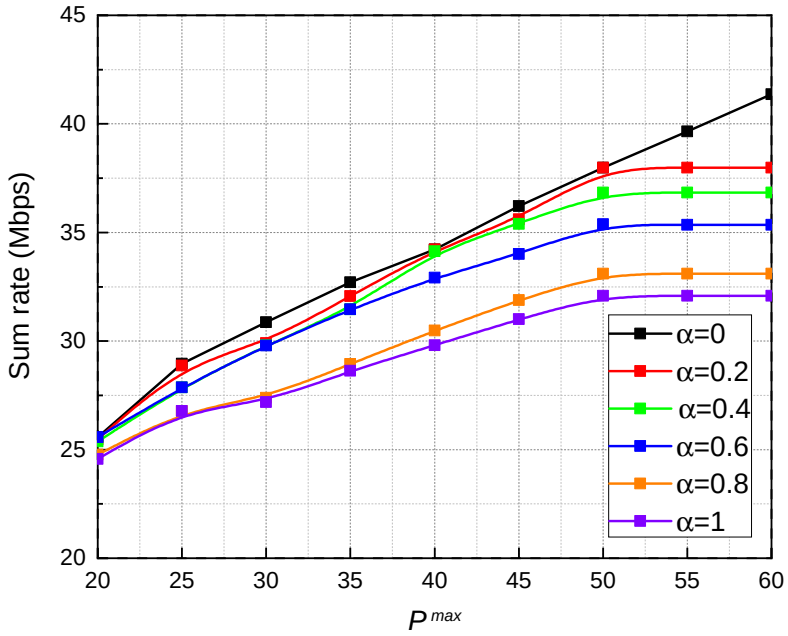

Fig. 5.2 The achieved SE performance versus P^{max} with different α.

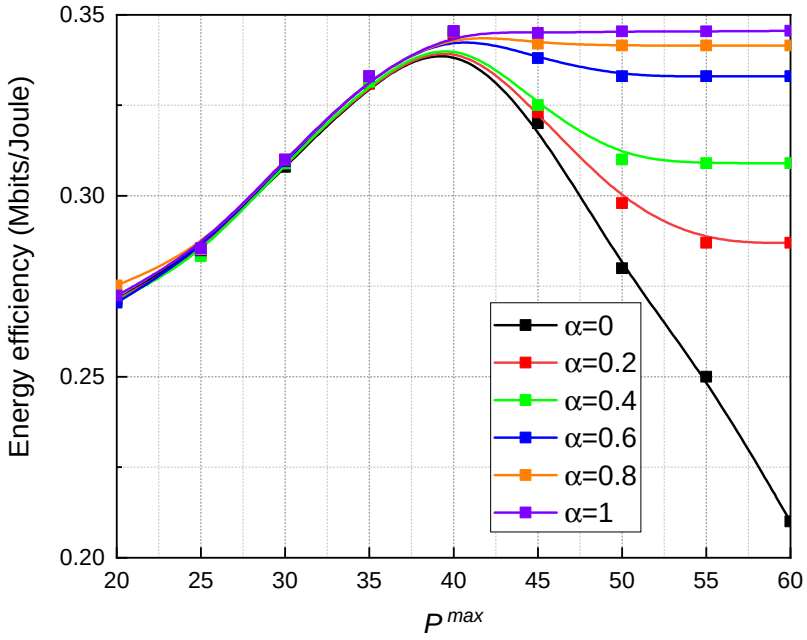

Fig. 5.3 The achieved EE performance versus P^{max} with different α.

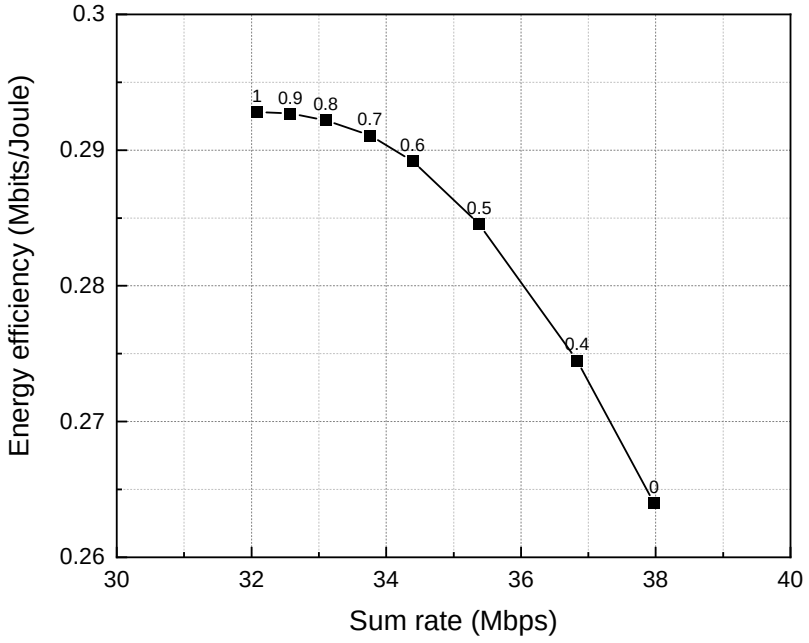

Fig. 5.4 A set of Pareto-optimal solutions of the proposed design.

5.4.2 Max-min Fairness Design with Opportunistic Time Assignment

To evaluate the performance of the proposed hybrid TDMA-NOMA scheme with opportunistic time allocations, the conventional scheme with equal time allocations is selected as the benchmark. In simulations, the users are uniformly distributed over a circle with a radius of 50 meters around the BS, and the minimum distance is selected such that $d_0 = 1$ meters. The corresponding channel gain is $|h_{j,i}|^2 = \frac{\beta}{(d_{j,i}/d_0)^\kappa}$, where $d_{j,i}$ is the distance between $u_{j,i}$ and the BS, measured in meters and $\beta = -30$ dB is the signal attenuation at d_0, and $\kappa = 2$ is the path-loss exponent. Ten users ($K = 10$) are considered, which are divided into five clusters ($C = 5$); accordingly, each cluster consists of two users. The noise variance at each user is assumed to be -100 dBm/Hz ($\sigma^2_{j,i} = -100$). Furthermore, the CVX software is utilized to solve the convex problems in this work [124].

Table 5.4 and 5.5 illustrate the performance of the hybrid TDMA-NOMA with the opportunistic time allocations versus that of the conventional schemes with equal time allocation. As seen in Table 5.4, the proposed opportunistic time allocations based hybrid TDMA-NOMA outperforms the conventional scheme with equal time

Table 5.4 Power Allocations For Each User In The Hybrid TDMA-NOMA Through The Proposed Opportunistic Time Allocations And The Conventional Equal Time One.

Scheme with opportunistic time allocations										Scheme with equal time allocations									
$p_{1,1}$	$p_{2,1}$	$p_{1,2}$	$p_{2,2}$	$p_{1,3}$	$p_{2,3}$	$p_{1,4}$	$p_{2,4}$	$p_{1,5}$	$p_{2,5}$	$p_{1,1}$	$p_{2,1}$	$p_{1,2}$	$p_{2,2}$	$p_{1,3}$	$p_{2,3}$	$p_{1,4}$	$p_{2,4}$	$p_{1,5}$	$p_{2,5}$
0.081	1.483	0.098	1.462	0.122	1.414	0.179	1.330	0.256	1.330	0.057	1.407	0.071	1.407	0.092	1.407	0.141	1.407	0.264	1.407
0.120	1.409	0.178	1.391	0.266	1.391	0.273	1.391	0.301	1.391	0.079	1.397	0.126	1.397	0.243	1.397	0.253	1.397	0.302	1.397
0.144	1.386	0.188	1.386	0.243	1.386	0.370	1.386	0.383	1.386	0.089	1.384	0.125	1.384	0.177	1.384	0.420	1.384	0.439	1.384
0.034	1.572	0.036	1.523	0.091	1.306	0.121	1.306	0.253	1.306	0.023	1.409	0.023	1.409	0.053	1.409	0.079	1.409	0.263	1.409
0.074	1.410	0.105	1.410	0.123	1.410	0.127	1.410	0.160	1.410	0.040	1.410	0.072	1.410	0.097	1.410	0.102	1.410	0.160	1.410

Table 5.5 Time Allocation And Achieved Minimum Throughout In The Hybrid TDMA-NOMA Through The Proposed Opportunistic Time Allocations And The Conventional Equal Time One.

	Scheme with opportunistic time allocations						Scheme with equal time allocations					
Channels	$t_1(s)$	$t_2(s)$	$t_3(s)$	$t_4(s)$	$t_5(s)$	Minimum throughput (bit/second)	$t_1(s)$	$t_2(s)$	$t_3(s)$	$t_4(s)$	$t_5(s)$	Minimum throughput (bit/second)
Channel 1	1.878	1.904	1.934	1.980	2.304	10.664	2	2	2	2	2	9.423
Channel 2	1.751	1.819	2.097	2.114	2.219	9.706	2	2	2	2	2	8.765
Channel 3	1.714	1.789	1.886	2.284	2.327	8.433	2	2	2	2	2	7.211
Channel 4	1.948	1.896	1.817	1.908	2.432	11.494	2	2	2	2	2	9.667
Channel 5	1.744	1.928	2.031	2.044	2.253	14.124	2	2	2	2	2	12.540

allocations in terms of minimum achieved rate. In particular, the opportunistic time allocations provides improvement to the overall system performance. Furthermore, Table 5.5 presents the power allocations of all users in the system for both schemes: with opportunistic and equal time allocations.

Next, Fig. 5.5 depicts the performance of these schemes in terms of the minimum achieved rate for different transmission power P^{max}. Simulation results confirm that the proposed scheme with opportunistic time allocation outperforms the scheme with equal time allocation in terms of the minimum achieved rate.

Finally, simulation results are presented to validate the convergence of the proposed algorithm in Fig. 5.6. Five different channels are considered, and as seen in Fig. 5.6, the proposed algorithm converges within a few iterations.

5.5 Summary

In this chapter, two different resource allocation designs are proposed for hybrid TDMA-NOMA system. Firstly, the SE-EE trade-off based resource allocation technique is

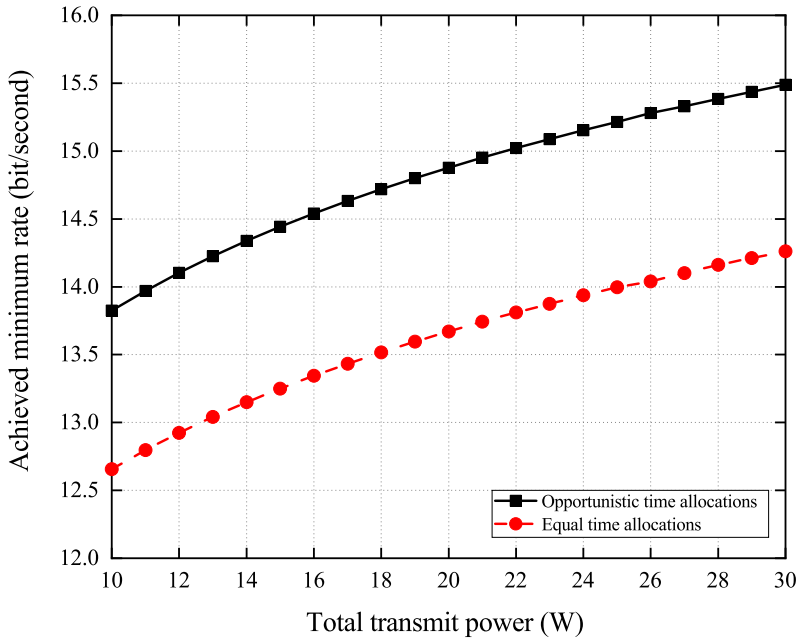

Fig. 5.5 The achieved minimum rate versus different total transmit power.

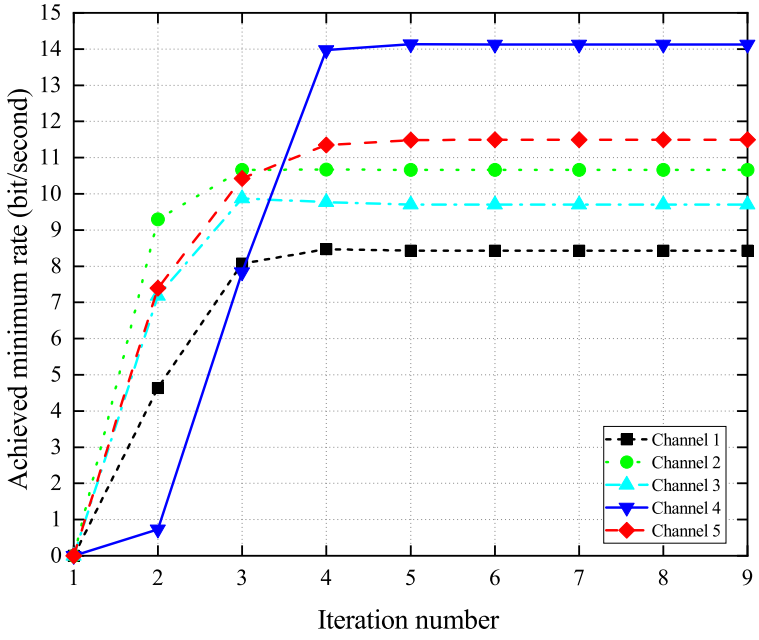

Fig. 5.6 The convergence of the SCA algorithm for five different channels.

proposed for a hybrid TDMA-NOMA system. Simulation results have demonstrated that the proposed SE-EE trade-off based design has the flexibility to strike a good balance between the conflicting metrics SE and EE compared to the conventional SE and EE designs. Secondly, the max-min joint resource allocation problem is studied of a SISO hybrid TDMA-NOMA system. Specifically, the available transmission time is divided into several time-slots and power-domain NOMA is exploited to serve multiple users within a cluster. However, the formulated max-min optimization problem is non-convex in nature. To cope with this challenge, an iterative algorithm is developed by exploiting SCA and a novel form of a SOC formulation to realize a solution to the original problem. Simulation results demonstrate that the proposed hybrid TDMA-NOMA system outperforms the conventional resource allocations with equal time assignment in terms of the minimum achieved rate and overall system throughput.

Chapter 6

Energy Efficiency Maximization for IRS-Assisted NOMA System with Imperfect CSI

In this chapter, a DL transmission of an IRS-assisted NOMA system is considered, where an IRS is deployed to support the communication between a multi-antenna BS and multiple single-antenna users. Specifically, the EE is maximized by jointly optimizing the transmit beamforming at BS and the phase shift of IRS. However, due to the non-convex nature of this joint optimization problem, an iterative algorithm based on the alternating optimization is proposed. The effectiveness of the proposed scheme is verified through simulations with different estimation error values and with OMA scheme.

6.1 Introduction

IRS is considered to be a viable and promising technology to improve the SE and EE, and the signal coverage in wireless networks [61, 126, 127]. At the same time, NOMA has been recently combined with the conventional OMA techniques, which offers additional degrees of freedom [128, 129]. Most existing works on NOMA have assumed the perfect CSI at the BS, which might be unrealistic in terms of practical implementation. It is impossible to have perfect CSI at the BS in terms of inevitable errors from the channel estimation and quantization in practice. Furthermore, the overall system might suffer a performance degradation due to these channel uncertainties [130]. Therefore, these channel uncertainties have been considered and incorporated

in the NOMA design to introduce the robustness and enjoy the full potential benefits offered by NOMA. In particular, the robust design can be categorized into two groups, the worst-case design with the bounded channel uncertainties [131, 132], and the outage probability based design with the statistics of channels and uncertainties [133–135]. In [131], the impact of the SINR balancing problem on the performance of a two-user MIMO-NOMA system with channel errors has been investigated. NOMA has been applied in a heterogeneous vehicular network to enhance the spectral efficiency, where the uplink was adopted with the imperfect channel estimation [132]. Considering the outage probability based robust design, [133] has investigated the power allocation and beamforming design problem to maximize the system utility under outage probability constraints. In [134], a robust power minimization optimization problem has been formulated under the constraints of the outage probability constraints of users' required rate. The expressions for the outage probability have been derived for a robust NOMA-aided design in vehicular communications and networking where two main causes of imperfect CSI have been considered, imposed by the channel estimation process and the Doppler spread, respectively [135].

Motivated by the above-mentioned discussion, this chapter investigates a worst-case robust design for IRS-assisted NOMA MISO system. A bounded channel uncertainty model is considered to define the CSI errors, where the transmit beamforming and reflecting matrix are designed to maximize the EE with a set of QoS constraints. In particular, the DL transmission of an IRS-assisted NOMA-based MISO system is considered and a joint EE-Max design is formulated, which maximizes the EE under a set of relevant QoS constraints, including the worst-case rate constraints under the bounded CSI error model, the minimum rate constraint of each user, the power budget constraint at the BS, and phase shift unit-modulus at IRS. However, the formulated EE-Max problem is non-convex in nature due to the fractional objective function and coupled variables in constraints. The following summarize the major contributions of this design.

- Assuming no direct links between BS and users and the ellipsoid-based channel uncertainties, the robust EE-Max problem is solved by jointly designing both the transmit beamforming for BS and reflection matrix for IRS based on the definition of the system EE, which represents the ratio between the total sum rate and total power consumption. To guarantee the QoS requirement of each user, the individual minimum data rate is chosen as a performance metric. Moreover, the total transmit power and the phase shift unit-modulus constraints are also incorporated in our formulated problem to guarantee the maximum transmit

power at BS and the unit-modulus requirements of the reflection elements at the IRS, respectively.

- An AO algorithm is proposed by applying the S-procedure and SDP to address the original non-convex problem. In particular, the reflection matrix is firstly initialized to solve the beamforming optimization problem. The beamforming optimization problem is still non-convex as the objective function is not only non-convex, but also fractional. In the first stage, given phase shift matrix, the optimal solutions of the beamforming optimization sub-problem are obtained by developing an iterative algorithm. In the second stage, the reflection matrix from the phase shift optimization sub-problem based on the beamforming in the first stage is obtained through an iterative algorithm. The process described above is repeated until convergence with the required accuracy. Furthermore, due to the imperfect CSI, the constraints are reformulated in terms of convex LMIs that can be easily solved, exploiting the S-procedure.

- The performance of the proposed robust IRS-assisted NOMA EE-Max design is demonstrated by comparing that of the non-robust design. Numerical results further verify that the proposed AO algorithm can provide the solution to the original EE-Max optimization problem as well as convergence within a few iterations.

6.2 System Model and Problem Formulations

6.2.1 System Model

In this subsection, a DL transmission of a multi-user MISO system is considered, where a BS equipped with M antennas serves K single-antenna users, and an IRS with N passive reflecting elements, as shown in Fig. 6.1. The IRS is deployed to establish virtual line-of-sight links to assist the communications between BS and users.

At the transmission stage, the desired signals from the BS are superimposed by employing SC, and then are encoded with different power levels. Hence, the combined transmitted complex base band signal, $\mathbf{x}$, at the BS can be represented as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k x_k, \tag{6.1}$$

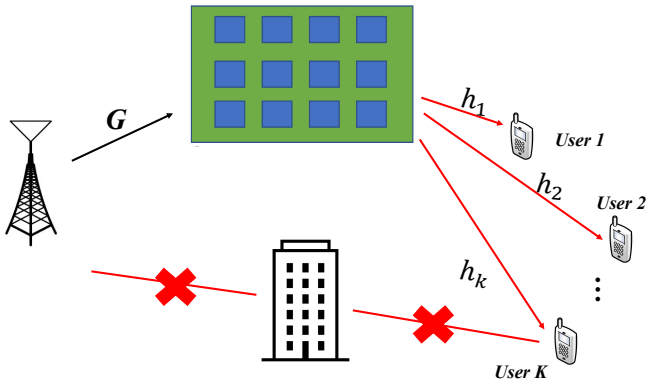

Fig. 6.1 IRS-assisted MISO NOMA system.

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$, and x_k denote the corresponding beamforming vector intended for each user and the transmitted symbol for user k, respectively. It is assumed that $x_k, k = 1, 2, \ldots, K$, are independent and have unit variance, then the transmission power of each symbol is 1, i.e., $\mathbb{E}[|x_k|^2] = 1$. The received signal at user k can be defined as follows:

$$y_k = \left(\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}\right) \sum_{j=1}^{K} \mathbf{w}_j x_j + n_k, \forall k \in \mathcal{K}, \tag{6.2}$$

where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$, and $\mathbf{G} \in \mathbb{C}^{N\times M}$ represent the channel coefficients from the IRS to kth user and from the BS to the IRS, respectively, while n_k denotes the AWGN with zero-mean and variance σ_k^2 at receiver. In addition, the matrix $\boldsymbol{\Theta} = \text{diag}(\beta e^{j\theta_1}, ..., \beta e^{j\theta_N})$ is the phase shift matrix at the IRS, where $\beta \in [0, 1]$ and $\theta_n \in [0, 2\pi), n \in \{1, 2, ..., N\}$ denote the amplitude reflection coefficient of the incident signal and the phase shift of nth reflecting element, respectively. Without loss of generality, and similar to the recent works in IRS-assisted systems, it is assumed that the amplitude reflection coefficient is one, i.e., $\beta = 1$. In this work, it is assumed that the direct links between the BS and the users are unavailable and obstructed by obstacles, such as building clusters, and industrial factories that produce heavy metallic objects and smog can block this communication path. With the help of the IRS, it can be deployed on the outside walls of buildings or the ceiling of the factories to bypass the obstacles and improve the performance of the system. This means the signals are transmitted through the IRS.

Furthermore, it is assumed that it is difficult to have perfect CSI in practical schemes due to errors from the channel estimation and quantization. To overcome this

problem, a worst-case approach is considered to deal with such channel uncertainties. In particular, the channel uncertainties can be defined by incorporating norm-bounded channel uncertainties model [136], i.e., the actual $\mathbf{h}_k$ takes the form as

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k, ||\triangle\mathbf{h}_k||_2 = ||\mathbf{h}_k - \widehat{\mathbf{h}}_k||_2 \leq \varepsilon, \tag{6.3}$$

where $\widehat{\mathbf{h}}_k$ denotes the estimate of $\mathbf{h}_k$ and $\triangle\mathbf{h}_k$ is the norm bounded channel estimation error. In addition, ε is the error bound of channel estimation. The concatenated channel $\mathbf{h}_k^H\mathbf{\Theta}\mathbf{G}$ depends on $\mathbf{h}_k^H$, $\mathbf{\Theta}$ and $\mathbf{G}$, which significantly complicates the user ordering. In the NOMA scheme, users can be multiplexed in the power domain and the user ordering can be determined according to their channel strengths, which is given by

$$||\widehat{\mathbf{h}}_1||_2 \leq ||\widehat{\mathbf{h}}_2||_2 \leq ... \leq ||\widehat{\mathbf{h}}_K||_2. \tag{6.4}$$

For example, user 2 has better channel conditions than user 1 based on the constraint in (6.4), then user 2 would decode the signal x_1 and remove it from the combined signals to decode x_2 by employing SIC. User 1 decodes its signal x_1 directly with the interference, $x_j, j > 1$. Considering this, the lth user can decode and remove the kth user's signals for $1 \leq k \leq l-1$ before decoding its individual desired signal. Consequently, the received signal at the lth user after SIC processing can be expressed as follows [136]:

$$\tilde{y}_l^k = \left(\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G}\right)\mathbf{w}_k x_k + \sum_{j=1}^{k-1}\left(\triangle\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G}\right)\mathbf{w}_j x_j + \sum_{j=k+1}^{K}\left(\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G}\right)\mathbf{w}_j x_j + n_k, \quad \forall k \in \mathcal{C}, l \in \{k, k+1, ..., K\}, \tag{6.5}$$

where the first term in (6.5) indicates the intended signal for the kth user. Since the considered framework has imperfect CSI, the second term represents that the l th user cannot completely subtract the signals intended for the users from 1 to $k-1$ when employing the SIC. The third term denotes the interference caused by the signals intended from user $k+1$ to user K. The last term is the AWGN.

Considering the above, the lth user decodes the message intended for the weaker user, i.e., $k < l$, with the following SINR:

$$SINR_l^k = \frac{|(\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G})\mathbf{w}_k|^2}{\sum_{j=1}^{k-1}|(\triangle\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G})\mathbf{w}_j|^2 + \sum_{j=k+1}^{K}|(\mathbf{h}_l^H\mathbf{\Theta}\mathbf{G})\mathbf{w}_j|^2 + \sigma_l^2}. \tag{6.6}$$

Therefore, the SINR of user k can be defined as

$$\gamma_k = \min_{l\in\{k,k+1,...,K\}} SINR_l^k = \min\{SINR_k^k, SINR_{k+1}^k, ..., SINR_K^k\}. \tag{6.7}$$

With the above definition, the achieved rate at user k can be expressed as

$$R_k = B\log_2(1+\gamma_k), \tag{6.8}$$

where B represents the available bandwidth in Hz. The total required transmit power P_t at the BS is $\sum_{k=1}^{K}||\mathbf{w}_k||^2$. Consequently, the total transmit power should satisfy the following constraint:

$$P_t \leq P^{max}, \tag{6.9}$$

where P^{max} is the maximum available transmit power at the BS.

6.2.2 Power Consumption Model

P_{total} denotes the total power consumption at the BS, which is thus formulated as [70, 73]

$$P_{total} = P_0(\mathbf{w}_k) + MP_{dyn} + P_{sta} = \frac{1}{\epsilon}||\mathbf{w}_k||^2 + P_{loss}, \tag{6.10}$$

where $P_0(\mathbf{w}_k)$ is a function of the corresponding beamforming vector, $\mathbf{w}_k$, and $\epsilon \in [0,1]$ is the efficiency of the power amplifier. In addition, P_{dyn} and P_{sta} denote the dynamic power consumption and the static power consumption, respectively. These two components can be modeled as P_{loss} since P_{dyn} and P_{sta} are not dependent on the corresponding beamforming vector.

Therefore, the EE in bits/Joule can be defined as [63]

$$EE = \frac{\sum_{k=1}^{K} R_k}{P_{total}}. \tag{6.11}$$

6.2.3 Problem Formulation

In this subsection, the robust EE-Max optimization problem is formulated by jointly designing of the beamforming vectors (i.e., $\mathbf{w}_k$) at the BS and the reflection matrix (i.e., $\mathbf{\Theta}$) at the IRS. In particular, a set of relevant constraints is also considered, including the data rate requirements of all users, the unit modulus constraints of all reflecting elements, and total power budget at the BS. Therefore, this EE-Max problem can be

formulated as

$$\textbf{(P14)}: \quad \max_{\mathbf{w}_k, \boldsymbol{\Theta}} \quad \frac{\sum_{k=1}^{K} R_k}{\frac{1}{\epsilon}\sum_{k=1}^{K} ||\mathbf{w}_k||^2 + MP_{dyn} + P_{sta}} \tag{6.12}$$

$$\text{s.t.} \quad R_k \geq R_k^{min}, \forall k \in \mathcal{K}, \tag{6.13}$$

$$\sum_{k=1}^{K} ||\mathbf{w}_k||^2 \leq P^{max}, \tag{6.14}$$

$$0 \leq \theta_n \leq 2\pi, n = 1, 2, \ldots, N, \tag{6.15}$$

where R_k^{min} is the threshold for the required QoS of user k. The constraint (6.13) refers to the QoS requirement for each user. The constraint (6.14) denotes the limitation of total transmit power, P^{max}. The constraint (6.15) ensures that the phase shift is between 0 and 2π. However, problem **P14** is a highly-coupled non-convex problem as the objective function and the constraints in (6.13) are non-convex. Therefore, the existing convex optimization techniques cannot be directly applied to determine the solution of the problem **P14**.

6.3 Alternating Optimization Framework

An AO framework is proposed to transform the original problem **P14** into two sub-problems, namely, the beamforming design and phase shift design of IRS, respectively. In other words, $\{\mathbf{w_k}\}$ and $\boldsymbol{\Theta}$ can be obtained by optimizing alternatively until required convergence accuracy is achieved.

6.3.1 Beamforming Design

For a given reflection matrix $\boldsymbol{\Theta}$, problem **P14** is still non-convex. The original problem **P14** can be reformulated into a epigraph form by introducing a slack variable t, as

follows:

$$\textbf{(P15)}: \quad \max_{\{\mathbf{w}_k\}_{k=1}^K, t} \quad t \tag{6.16}$$

$$\text{s.t.} \quad \frac{\sum_{k=1}^K \log_2(1+\gamma_k)}{\frac{1}{\epsilon}\sum_{k=1}^K ||\mathbf{w}_k||^2 + MP_{dyn} + P_{sta}} \geq t, \tag{6.17}$$

$$R_k \geq R_k^{min}, \forall k \in \mathcal{K}, \tag{6.18}$$

$$\sum_{k=1}^K ||\mathbf{w}_k||^2 \leq P^{max}. \tag{6.19}$$

To deal with the non-convex constraint in (6.17), another slack variable ξ is introduced, and thus, the following two constraints can be used to represent the constraint in (6.17) [137]:

$$\sum_{k=1}^K \log_2(1+\gamma_k) \geq t\xi, \tag{6.20a}$$

$$\frac{1}{\epsilon}\sum_{k=1}^K ||\mathbf{w}_k||^2 + MP_{dyn} + P_{sta} \leq \xi. \tag{6.20b}$$

A set of new slack variables α_k is introduced to deal with the non-convexity of (6.20a). Hence, the constraint (6.20a) can be expressed as

$$\sum_{k=1}^K \log_2(\alpha_k) \geq t\xi, \tag{6.21a}$$

$$1+\gamma_k \geq \alpha_k, \forall k \in \mathcal{K}. \tag{6.21b}$$

By introducing another variable β_k, the constraint in (6.21a) can be equivalently represented by

$$\sum_{k=1}^K \beta_k \geq t\xi, \tag{6.22a}$$

$$\alpha_k \geq 2^{\beta_k}, \forall k \in \mathcal{K}. \tag{6.22b}$$

According to the above new slack variables, the constraint in (6.20a) can equivalently be expressed into a set of constraints, as

$$(6.20a) \Leftrightarrow \begin{cases} (6.21b): 1+\gamma_k \geq \alpha_k, \forall k \in \mathcal{K}, \\ (6.22a): \sum_{k=1}^{K} \beta_k \geq t\xi, \\ (6.22b): \alpha_k \geq 2^{\beta_k}, \forall k \in \mathcal{K}. \end{cases}$$

Although the constraints in (6.22b) are not non-convex anymore, the constraints in (6.21b) and (6.22a) are still non-convex considering some variables are tightly coupled together. To deal with the non-convexity of the constraint in (6.22a), the upper bound of $t\xi$ can be obtained as [137, 138]

$$\frac{t^{(n)}}{2\xi^{(n)}}\xi^2 + \frac{\xi^{(n)}}{2t^{(n)}}t^2 \geq t\xi, \tag{6.23}$$

where $t^{(n)}$ and $\xi^{(n)}$ are the values of t and ξ at the nth iteration, respectively. Therefore, the constraint in (6.22a) can be transformed into the following convex constraints

$$\sum_{k=1}^{K} \beta_k \geq \frac{t^{(n)}}{2\xi^{(n)}}\xi^2 + \frac{\xi^{(n)}}{2t^{(n)}}t^2. \tag{6.24}$$

To deal with non-convexity of constraints in (6.21b),

$$1 + \min_{||\triangle \mathbf{h}_l||_2 \leq \varepsilon} \{SINR_k^k, SINR_{k+1}^k, \ldots, SINR_K^k\} \geq \alpha_k. \tag{6.25}$$

A new matrix variable $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ is introduced and define $\mathbf{F} = \mathbf{\Theta}\mathbf{G}$, then the constraint in (6.21b) can be rewritten as

$$\min_{||\triangle \mathbf{h}_l||_2 \leq \varepsilon} \frac{\left(\left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{F} \right) \mathbf{w}_k \right|^2 \right)}{\sum_{j=1}^{k-1} \left| \left(\triangle \mathbf{h}_l^H \mathbf{F} \right) \mathbf{w}_j \right|^2 + \sum_{j=k+1}^{K} \left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{F} \right) \mathbf{w}_j \right|^2 + \sigma_l^2} \geq \alpha_k - 1, \quad \forall l = k, k+1, \ldots, K. \tag{6.26}$$

Due to the uncertain term $\triangle\mathbf{h}_l$, the following Lemma 4 is presented, which relates to reformulate the non-convex constraints as LMI constraints. This is referred to as the *S*-procedure [139, 140] Lemma.

Lemma 4 *(S-procedure) Let* $\mathbf{C}_i = \mathbf{C}_i^H \in \mathbb{C}^{N\times N}$, $\mathbf{b}_i \in \mathbb{C}^{N\times 1}$, *and* $d_i \in \mathbb{R}$. *The function* $f_1(\mathbf{x})$ *and* $f_2(\mathbf{x})$, *can be defined in the following form:*

$$\begin{aligned} f_1(\mathbf{x}) &= \mathbf{x}^H\mathbf{C}_1\mathbf{x} + 2\Re\{\mathbf{b}_1^H\mathbf{x}\} + d_1, \\ f_2(\mathbf{x}) &= \mathbf{x}^H\mathbf{C}_2\mathbf{x} + 2\Re\{\mathbf{b}_2^H\mathbf{x}\} + d_2, \end{aligned} \tag{6.27}$$

Then, the implication

$$f_1(\mathbf{x}) \leq 0 \Rightarrow f_2(\mathbf{x}) \leq 0 \tag{6.28}$$

holds if and only if there exists a non-negative number $\lambda \geq 0$, *such that*

$$\lambda \begin{bmatrix} \mathbf{C}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^H & d_1 \end{bmatrix} - \begin{bmatrix} \mathbf{C}_2 & \mathbf{b}_2 \\ \mathbf{b}_2^H & d_2 \end{bmatrix} \succeq 0,$$

is positive semi-definite.

By applying *S*-procedure, the numerator and the denominator of (6.26) can be express as (6.29) and (6.30), respectively.

$$\begin{aligned} &\left(\left|\left(\left(\widehat{\mathbf{h}}_l + \triangle\mathbf{h}_l\right)^H\mathbf{F}\right)\mathbf{w}_k\right|^2\right) = \left(\left(\widehat{\mathbf{h}}_l + \triangle\mathbf{h}_l\right)^H\mathbf{F}\mathbf{w}_k\right)\left(\left(\widehat{\mathbf{h}}_l + \triangle\mathbf{h}_l\right)^H\mathbf{F}\mathbf{w}_k\right)^H \\ &= \triangle\mathbf{h}_l^H\mathbf{F}\mathbf{W}_k\mathbf{F}^H\triangle\mathbf{h}_l + 2\Re(\widehat{\mathbf{h}}_l^H\mathbf{F}\mathbf{W}_k\mathbf{F}^H\triangle\mathbf{h}_l) + \widehat{\mathbf{h}}_l^H\mathbf{F}\mathbf{W}_k\mathbf{F}^H\widehat{\mathbf{h}}_l. \end{aligned} \tag{6.29}$$

$$\begin{aligned} &\sum_{j=1}^{k-1}\left|\left(\triangle\mathbf{h}_l^H\mathbf{F}\right)\mathbf{w}_j\right|^2 + \sum_{j=k+1}^{K}\left|\left(\left(\widehat{\mathbf{h}}_l + \triangle\mathbf{h}_l\right)^H\mathbf{F}\right)\mathbf{w}_j\right|^2 + \sigma_l^2 \\ &= \sum_{j=1}^{k-1}\triangle\mathbf{h}_l^H\mathbf{F}\mathbf{W}_j\mathbf{F}^H\triangle\mathbf{h}_l + \sum_{j=k+1}^{K}\{\triangle\mathbf{h}_l^H\mathbf{F}\mathbf{W}_j\mathbf{F}^H\triangle\mathbf{h}_l + 2\Re(\widehat{\mathbf{h}}_l^H\mathbf{F}\mathbf{W}_k\mathbf{F}^H\triangle\mathbf{h}_l) \\ &+ \widehat{\mathbf{h}}_l^H\mathbf{F}\mathbf{W}_k\mathbf{F}^H\widehat{\mathbf{h}}_l\} + \sigma_l^2 \\ &= \triangle\mathbf{h}_l^H\mathbf{F}\sum_{j\neq k}\mathbf{W}_j\mathbf{F}^H\triangle\mathbf{h}_l + 2\Re\left(\widehat{\mathbf{h}}_l^H\mathbf{F}\sum_{j=k+1}^{K}\mathbf{W}_k\mathbf{F}^H\triangle\mathbf{h}_l\right) + \widehat{\mathbf{h}}_l^H\mathbf{F}\sum_{j=k+1}^{K}\mathbf{W}_j\mathbf{F}^H\widehat{\mathbf{h}}_l \\ &+ \sigma_l^2, \forall l = k, k+1, \ldots, K. \end{aligned} \tag{6.30}$$

Next, the constraint in (6.26) can be transformed into a tractable form as follows:

$$\left(\left|\left(\left(\widehat{\mathbf{h}}_l+\triangle\mathbf{h}_l\right)^H\mathbf{F}\right)\mathbf{w}_k\right|^2\right)\geq\tau_k, \tag{6.31a}$$

$$\tau_k\geq(\alpha_k-1)v_k, \tag{6.31b}$$

$$\sum_{j=1}^{k-1}\left|\left(\triangle\mathbf{h}_l^H\mathbf{F}\right)\mathbf{w}_j\right|^2+\sum_{j=k+1}^{K}\left|\left(\left(\widehat{\mathbf{h}}_l+\triangle\mathbf{h}_l\right)^H\mathbf{F}\right)\mathbf{w}_j\right|^2+\sigma_l^2\leq v_k. \tag{6.31c}$$

Note that

$$\triangle\mathbf{h}_l^H\mathbf{I}\triangle\mathbf{h}_l-\varepsilon^2\leq 0. \tag{6.32}$$

Combining (6.29), (6.32) and Lemma 4, the constraint in (6.31a) can be transformed into the following convex LMI:

$$\begin{bmatrix}\lambda_l^k\mathbf{I}+\mathbf{d}_k & (\widehat{\mathbf{h}}_l^H\mathbf{d}_k)^H\\ (\widehat{\mathbf{h}}_l^H\mathbf{d}_k) & \widehat{\mathbf{h}}_l^H\mathbf{d}_k\widehat{\mathbf{h}}_l-\tau_k-\lambda_l^k\varepsilon^2\end{bmatrix}\succeq\mathbf{0}, \tag{6.33}$$

where $\mathbf{d}_k=\mathbf{F}\mathbf{W}_k\mathbf{F}^H$. It is obvious that constraint in (6.31b) is non-convex in nature due to the coupling of slack variables α_k and v_k. To overcome this issue, a set of upper bound of the coupling terms is utilized to replace them, then constraint in (6.31b) can be represented as follows [137, 138]:

$$\tau_k\geq\frac{(\alpha_k^{(n)}-1)}{2v_k^{(n)}}v_k^2+\frac{v_k^{(n)}}{2(\alpha_k^{(n)}-1)}(\alpha_k-1)^2. \tag{6.34}$$

where $\alpha_k^{(n)}$ and $v_k^{(n)}$ are the values of α_k and v_k at the nth iteration, respectively. Based on the constraints in (6.30), (6.32) and Lemma 4, the constraint in (6.31c) can be transformed into the following convex LMI:

$$\begin{bmatrix}\lambda_l^k\mathbf{I}-\mathbf{d}_k' & -(\widehat{\mathbf{h}}_l^H\mathbf{f}_k)^H\\ -(\widehat{\mathbf{h}}_l^H\mathbf{f}_k) & v_k-\widehat{\mathbf{h}}_l^H\mathbf{f}_k\widehat{\mathbf{h}}_l-\lambda_l^k\varepsilon^2-\sigma_l^2\end{bmatrix}\succeq\mathbf{0}, \tag{6.35}$$

where

$$\mathbf{d}_k'=\mathbf{F}\left(\sum_{j\neq k}\mathbf{W}_k\right)\mathbf{F}^H,$$

$$\mathbf{f}_k=\mathbf{F}\left(\sum_{j=k+1}^{K}\mathbf{W}_k\right)\mathbf{F}^H.$$

Next, the minimum rate constraint in (6.18) can be expressed as:

$$\gamma_k \geq \gamma_k^{min}, \tag{6.36}$$

where $\gamma_k^{min} = 2^{R_k^{min}} - 1$. Similarly, the above constraint in (6.36) can be converted into LMI form as follows:

$$\begin{aligned}
&\triangle \mathbf{h}_l^H \mathbf{I} \triangle \mathbf{h}_l - \varepsilon^2 \leq 0 \Longrightarrow \\
&\triangle \mathbf{h}_l^H \mathbf{F} \left(\gamma_k^{min} \sum_{j \neq k} \mathbf{W}_j - \mathbf{W}_k \right) \mathbf{F}^H \triangle \mathbf{h}_l + \widehat{\mathbf{h}}_l^H \mathbf{F} \left(\gamma_k^{min} \sum_{j=k+1}^{K} \mathbf{W}_j - \mathbf{W}_k \right) \mathbf{F}^H \widehat{\mathbf{h}}_l \\
&+ 2\Re \left\{ \widehat{\mathbf{h}}_l^H \mathbf{F} \left(\gamma_k^{min} \sum_{j=k+1}^{K} \mathbf{W}_j - \mathbf{W}_k \right) \mathbf{F}^H \triangle \mathbf{h}_l \right\} + \gamma_k^{min} \sigma_l^2 \leq 0, \forall l = k, k+1, \ldots, K.
\end{aligned} \tag{6.37}$$

Then,

$$\begin{bmatrix} \lambda_l^k \mathbf{I} + \mathbf{c}_k' & (\widehat{\mathbf{h}}_l^H \mathbf{c}_k)^H \\ (\widehat{\mathbf{h}}_l^H \mathbf{c}_k) & \widehat{\mathbf{h}}_l^H \mathbf{c}_k \widehat{\mathbf{h}}_l - \lambda_l^k \varepsilon^2 - \gamma_k^{min} \sigma_l^2 \end{bmatrix} \succeq \mathbf{0}, \tag{6.38}$$

where

$$\begin{aligned}
\mathbf{c}_k' &= \mathbf{F} \left(\mathbf{W}_k - \gamma_k^{min} \sum_{j \neq k} \mathbf{W}_j \right) \mathbf{F}^H, \\
\mathbf{c}_k &= \mathbf{F} \left(\mathbf{W}_k - \gamma_k^{min} \sum_{j=k+1}^{K} \mathbf{W}_j \right) \mathbf{F}^H.
\end{aligned}$$

Finally, the original problem **P15** can be transformed into the following SDP problem,

$$\begin{aligned}
\textbf{(P16)}: \quad & \max_{\{\mathbf{W}_k\}_{k=1}^K, t, \xi, \alpha_k, \beta_k, \tau_k, v_k} && t && (6.39) \\
& \text{s.t.} && (6.22b), (6.24), (6.33), (6.34), (6.35), && (6.40) \\
& && \sum_{k=1}^{K} \mathrm{tr}\left(\mathbf{W}_k\right) \leq \epsilon(\xi - MP_{dyn} - P_{sta}), && (6.41) \\
& && (6.38), && (6.42) \\
& && \sum_{k=1}^{K} \mathrm{tr}\left(\mathbf{W}_k\right) \leq P^{max} && , \quad (6.43) \\
& && \mathbf{W}_k \succeq 0, \mathrm{rank}\left(\mathbf{W}_k\right) = 1. && (6.44)
\end{aligned}$$

Table 6.1 The Gaussian Randomization Procedure

Algorithm 1: The Gaussian Randomization Procedure for Problem **P16**.

1: Solve the problem **P16** without the rank-one constraints in (6.44) and find the solution matrices $\mathbf{W}_k^*$;
2: If $\text{rank}(\mathbf{W}_k^*) = 1$, then use the Cholesky decomposition to obtain the optimal solution $\mathbf{w}_k^*$ for the original problem;
3: Else
4: Generate K random vectors $\mathbf{a}_k, k = 1, 2, \ldots, K$ of $\mathbf{W}_k^*$ by introducing the Gaussian randomization technique;
5: Calculate k^* based on the equation in (6.45);
6: Choose the $\mathbf{a}_k^*$ as an approximate solution of $\mathbf{w}_k^*$.

Due to the non-convex nature of the rank-one constraint in (6.44), this problem **P16** cannot be solved directly by using the existing convex optimization techniques. Hence, the rank-one constraint in (6.44) is firstly relaxed, and then prove that the optimal solution of the relaxed problem also satisfy the rank-one constraint. It can be observed that the relaxed problem is a standard SDP problem, which can be efficiently solved by the convex optimization techniques. Cholesky decomposition is adopted in case the solution matrices $\mathbf{W}_k^*$ of the relaxed problem meet the rank-one constraints. By using Cholesky decomposition, the optimal solution for the original problem can be constructed. For the case of the solution matrices $\mathbf{W}_k^*$ do not meet the rank-one constraints, the Gaussian randomization technique is used to generate a set of rank-one solutions [141, 142]. Specifically, the approximate solution of $\mathbf{w}_k^*$ is given by

$$k^* = \arg \min_{k \in 1,2,\ldots K} (\mathbf{a}_k)^H \mathbf{a}_k - \text{tr}(\mathbf{W}_k^*). \tag{6.45}$$

The algorithm based on Gaussian randomization for the problem **P16** is summarized in Table 6.1.

Theorem 2 *[137, 143] The optimal solutions of the relaxed problem, defined as* $\mathbf{W}_k^*$*, should satisfy the rank-one constraint* $\text{rank}(\mathbf{W}_k^*) = 1$.

With Theorem 2, the problem **P16** can be solved directly without considering the rank-one constraint in (6.44). The proof of Theorem 2 can be found in [137, 143].

6.3.2 Phase Shift Design of IRS

For a given set of beamforming vectors, $\mathbf{w}_k$, the EE-Max problem **P14** can be reduced to a sum-rate maximization problem:

$$\textbf{(P17)}: \quad \max_{\boldsymbol{\Theta}} \quad \sum_{k=1}^{K} R_k \tag{6.46}$$

$$\text{s.t.} \quad R_k \geq R_k^{min}, \forall k \in \mathcal{K}, \tag{6.47}$$

$$0 \leq \theta_n \leq 2\pi, n = 1, 2, \ldots, N. \tag{6.48}$$

To deal with the non-convexity of the sum-rate maximization problem **P17**, a slack variable a is introduced to approximate the objective function. Consequently, the problem **P17** can be rewritten as follows:

$$\textbf{(P18)}: \quad \max_{\boldsymbol{\Theta}} \quad a \tag{6.49}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} R_k \geq a, \tag{6.50}$$

$$R_k \geq R_k^{min}, \forall k \in \mathcal{K}, \tag{6.51}$$

$$0 \leq \theta_n \leq 2\pi, n = 1, 2, \ldots, N. \tag{6.52}$$

Then, a set of new slack variables η_k is introduced and then constraint (6.50) can be expressed as

$$\sum_{k=1}^{K} \log_2 (\eta_k) \geq a, \tag{6.53a}$$

$$1 + \gamma_k \geq \eta_k, \forall k \in \mathcal{K}. \tag{6.53b}$$

By introducing another slack variables ι_k, the constraint in (6.53a) can be equivalently reformulated as

$$\sum_{k=1}^{K} \iota_k \geq a, \tag{6.54a}$$

$$\eta_k \geq 2^{\iota_k}, \forall k \in \mathcal{K}. \tag{6.54b}$$

From the above discussion, the constraint in (6.50) can equivalently be represented as

$$(6.50) \Leftrightarrow \begin{cases} (6.53b): 1+\gamma_k \geq \eta_k, \forall k \in \mathcal{K}, \\ (6.54a): \sum_{k=1}^{K} \iota_k \geq a, \\ (6.54b): \eta_k \geq 2^{\iota_k}, \forall k \in \mathcal{K}, \end{cases}$$

while the (6.53b) is still non-convex, such as

$$1+\min_{||\triangle \mathbf{h}_l||_2 \leq \varepsilon} \{SINR_k^k, SINR_{k+1}^k, \ldots, SINR_K^k\} \geq \eta_k. \tag{6.55}$$

$$\min_{||\triangle \mathbf{h}_l||_2 \leq \varepsilon} \frac{\left(\left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{\Theta G} \right) \mathbf{w}_k \right|^2 \right)}{\sum_{j=1}^{k-1} \left| \left(\triangle \mathbf{h}_l^H \mathbf{\Theta G} \right) \mathbf{w}_j \right|^2 + \sum_{j=k+1}^{K} \left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{\Theta G} \right) \mathbf{w}_j \right|^2 + \sigma_l^2}$$
$$\geq \eta_k - 1, \forall l = k, k+1, \ldots, K. \tag{6.56}$$

The numerator and denominator of (6.56) can be expressed in (6.57) and (6.58),

$$\left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{\Theta G} \right) \mathbf{w}_k \right|^2 \Rightarrow \left| \left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathrm{diag}\left(\mathbf{Z}_k \right) \phi \right|^2. \tag{6.57}$$

$$\sum_{j=1}^{k-1} \left| \left(\triangle \mathbf{h}_l^H \mathbf{\Theta G} \right) \mathbf{w}_j \right|^2 + \sum_{j=k+1}^{K} \left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathbf{\Theta G} \right) \mathbf{w}_j \right|^2 + \sigma_l^2 \Rightarrow$$
$$\sum_{j=1}^{k-1} \left| \left(\triangle \mathbf{h}_l^H \mathrm{diag}\left(\mathbf{Z}_j \right) \phi \right) \right|^2 + \sum_{j=k+1}^{K} \left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathrm{diag}\left(\mathbf{Z}_j \right) \phi \right) \right|^2 + \sigma_l^2, \tag{6.58}$$

where $\mathbf{Z}_k = \mathbf{G}\mathbf{w}_k$, $\phi = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}]^T$. Considering that (6.55) has a similar expression with (6.25), the same scheme can solve this constraint. Consequently, the constraint in (6.55) can be expressed as

$$\left| \left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathrm{diag}\left(\mathbf{Z}_k \right) \phi \right|^2 \geq \varphi_k, \tag{6.59a}$$

$$\varphi_k \geq (\eta_k - 1) g_k, \tag{6.59b}$$

$$\sum_{j=1}^{k-1} \left| \left(\triangle \mathbf{h}_l^H \mathrm{diag}\left(\mathbf{Z}_j \right) \phi \right) \right|^2 + \sum_{j=k+1}^{K} \left| \left(\left(\widehat{\mathbf{h}}_l + \triangle \mathbf{h}_l \right)^H \mathrm{diag}\left(\mathbf{Z}_j \right) \phi \right) \right|^2 + \sigma_l^2 \leq g_k. \tag{6.59c}$$

Let $\mathbf{\Phi} = \phi\phi^H$, then, the equivalent transformation of the constraint in (6.59a) can be expressed into the convex LMI:

$$\begin{bmatrix} \lambda_l^k \mathbf{I} + \mathbf{m}_k & (\widehat{\mathbf{h}}_l \mathbf{m}_k)^H \\ (\widehat{\mathbf{h}}_l \mathbf{m}_k) & \widehat{\mathbf{h}}_l^H \mathbf{m}_k \widehat{\mathbf{h}}_l - \varphi_k - \lambda_l^k \varepsilon^2 \end{bmatrix} \succeq \mathbf{0}, \tag{6.60}$$

where $\mathbf{m}_k = \mathrm{diag}(\mathbf{Z}_k)\mathbf{\Phi}\mathrm{diag}(\mathbf{Z}_k)^H$. A similar approach is used to address this non-convex constraint in (6.59b), which can be written as follows:

$$\varphi_k \geq \frac{\left(\eta_k^{(n)} - 1\right)}{2g_k^{(n)}} g_k^2 + \frac{g_k^{(n)}}{2\left(\eta_k^{(n)} - 1\right)} \left(\eta_k - 1\right)^2. \tag{6.61}$$

where $\varphi_k^{(n)}$ and $g_k^{(n)}$ are the values of φ_k and g_k at the nth iteration, respectively. Similar to the aforementioned approach of the constraint in (6.59a), the constraint in (6.59c) can also be transformed into the convex LMI:

$$\begin{bmatrix} \lambda_l^k \mathbf{I} - \mathbf{m}_k' & -(\widehat{\mathbf{h}}_l \mathbf{f}_k')^H \\ -(\widehat{\mathbf{h}}_l \mathbf{f}_k') & g_k - \widehat{\mathbf{h}}_l^H \mathbf{f}_k' \widehat{\mathbf{h}}_l - \lambda_l^k \varepsilon^2 - \sigma_l^2 \end{bmatrix} \succeq \mathbf{0}, \tag{6.62}$$

where

$$\mathbf{m}_k' = \sum_{j \neq k} \mathrm{diag}\left(\mathbf{Z}_j\right) \mathbf{\Phi} \mathrm{diag}\left(\mathbf{Z}_j\right)^H,$$
$$\mathbf{f}_k' = \sum_{j=k+1}^{K} \mathrm{diag}\left(\mathbf{Z}_j\right) \mathbf{\Phi} \mathrm{diag}\left(\mathbf{Z}_j\right)^H.$$

Therefore, the constraint in (6.50) can equivalently be represented by (6.54a), (6.54b), (6.60), (6.61) and (6.62).

Next, the minimum rate constraint in (6.51) can be represented as:

$$\begin{bmatrix} \lambda_l^k \mathbf{I} + \mathbf{q}_k' & (\widehat{\mathbf{h}}_l \mathbf{q}_k)^H \\ (\widehat{\mathbf{h}}_l \mathbf{q}_k) & \widehat{\mathbf{h}}_l^H \mathbf{q}_k \widehat{\mathbf{h}}_l - \lambda_l^k \varepsilon^2 - \gamma_k^{min} \sigma_l^2 \end{bmatrix} \succeq \mathbf{0}, \tag{6.63}$$

where the $\gamma_k^{min}, \mathbf{q}_k', \mathbf{q}_k$ are represented in (6.64), (6.65) and (6.66), which are provided as follows:

$$\gamma_k^{min} = 2^{R_k^{min}} - 1, \tag{6.64}$$

$$\mathbf{q}_k' = \left(\text{diag}\left(\mathbf{Z}_k\right) - \gamma_k^{min} \sum_{j \neq k} \text{diag}\left(\mathbf{Z}_j\right)\right) \mathbf{\Phi} \left(\text{diag}\left(\mathbf{Z}_k\right) - \gamma_k^{min} \sum_{j \neq k} \text{diag}\left(\mathbf{Z}_j\right)\right)^H, \tag{6.65}$$

$$\mathbf{q}_k = \left(\text{diag}\left(\mathbf{Z}_k\right) - \gamma_k^{min} \sum_{j=k+1}^{K} \text{diag}\left(\mathbf{Z}_j\right)\right) \mathbf{\Phi} \left(\text{diag}\left(\mathbf{Z}_k\right) - \gamma_k^{min} \sum_{j=k+1}^{K} \text{diag}\left(\mathbf{Z}_j\right)\right)^H. \tag{6.66}$$

Finally, the problem **P18** can be equivalently transformed to

$$\begin{aligned}
\textbf{(P19)}: \quad & \max_{\mathbf{\Phi}, a, \eta_k, \iota_k, \varphi_k, g_k} \quad a \\
& \text{s.t.} \quad (6.54a), (6.54b), (6.60), (6.61), (6.62), && (6.67) \\
& \quad (6.63), && (6.68) \\
& \quad 0 \leq \theta_n \leq 2\pi, n = 1, 2, \ldots, N, && (6.69) \\
& \quad \mathbf{\Phi} \succeq 0, \text{rank}(\mathbf{\Phi}) = 1. && (6.70)
\end{aligned}$$

Similar to the previous approach, the existing convex optimization software can be utilized to solve the formulated SDP problem **P16** removing the rank-one constraint in (6.70). If the obtained $\mathbf{\Phi}$ satisfies the rank-one constraint $\text{rank}(\mathbf{\Phi}) = 1$, the optimal ϕ can be reconstructed by using eigenvalue decomposition of $\mathbf{\Phi} = \phi\phi^H$ and the reflection coeffcient matrix $\mathbf{\Theta}$ can be expressed as $\mathbf{\Theta} = \text{diag}\{\phi_1, \phi_2, \ldots, \phi_N\}$. Otherwise, the Gaussian randomization technique can be used to construct a rank-one solution [142].

Finally, the proposed iterative AO algorithm is summarized in Table 6.2. Specifically, the transmit beamforming at the BS for given reflection matrix $\mathbf{\Theta}$ is considered. Next, the reflection matrix $\mathbf{\Theta}$ is optimized based on the obtained beamforming vectors $\mathbf{w}_k$. The above procedure is repeated until required convergence accuracy.

6.3.3 Complexity Analysis of the Proposed Schemes

This section discusses the computational complexity of the proposed algorithm. For the proposed AO algorithm, the two sub-problems (convex) in **P16** and **P19** are alternatively solved until convergence with the required accuracy. Specifically, the two sub-problems contain with a series of linear, SOC, and LMI constraints, which can

Table 6.2 The AO Algorithm

Algorithm 2: The Proposed Iterative AO Algorithm.

1: Initialize the outer iteration index $n = 1$ and the reflection matrix $\mathbf{\Theta} = \mathbf{\Theta}^{(0)}$;
2: **Repeat**
3: **Beamforming Optimization**: Initialize variables $\{t^{(0)}, \xi^{(0)}, \alpha_k^{(0)}, v_k^{(0)}\}$ and the inner iteration index $n' = 1$;
4: **While** $t^{(n')} - t^{(n'-1)} > \omega$ **do**
5: Obtain $\{\mathbf{W}_k\}_{k=1}^K, t, \xi, \alpha_k, \beta_k, \tau_k, v_k$ by solving problem (P16).
6: **end while**
7: Update variables $\{\mathbf{W}_k^{(n)}\}_{k=1}^K = \{\mathbf{W}_k^{(n')}\}_{k=1}^K$
8: **Phase Shift Optimization**: Initialize variables $\{\varphi_k^{(0)}, g_k^{(0)}\}$ and the inner iteration index $n^\dagger = 1$;
9: Given by $\{\mathbf{W}_k^{(n)}\}_{k=1}^K$, obtain $\mathbf{\Theta}^\dagger$ by solving problem (P19).
10: Update $\mathbf{\Theta}^{(n)} = \mathbf{\Theta}^{(n^\dagger)}$
11: **Calculate** $EE^{(n)}$ by $\mathbf{\Theta}^{(n)}$ and $\{\mathbf{W}_k^{(n)}\}_{k=1}^K$ and let $n = n + 1$.
12: **Until** $EE^{(n)} - EE^{(n-1)} < \omega$
13: Obtain the beamforming $\{\mathbf{W}_k^*\}_{k=1}^K$, the reflection matrix $\mathbf{\Theta}^*$ and EE^*.

be efficiently solved by the interior-point-based method [100]. Therefore, following the complexity analysis in [144, 145], and the complexity of each sub-problem can be defined in the following.

With the two sub-problems formulated in the previous subsections, an iterative algorithm is proposed to solve problem **P14** by utilizing the AO method. Hence, the beamforming vector $\mathbf{w}_k$ and the phase shift matrix $\mathbf{\Theta}$ are alternately optimized by solving problem **P16** and **P19**, where the local input points of each iteration are obtained from the solutions of the previous iteration. Therefore, the complexity of solving the sub-problems **P16** and **P19** can be quantified to determine the complexity of solving **P14**. The computational complexity of the general interior-point methods is given by [144–146]

$$\mathcal{O} = \sqrt{\varrho}\ln(1/c) + n\left((n^2 + n\sum_{i=1}^{A} m_i^2 + \sum_{i=1}^{A} m_i^3) + \sum_{i=1}^{B} m_i^2\right), \tag{6.71}$$

where the first term is related to the iteration complexity, and c represents the iteration accuracy, ϱ is the barrier parameter associated with the cone [145, 146]. The second term is called the per-iteration computational cost, where n, A, and B denote the number of variables, LMI constraints, and SOC constraints, respectively. Therefore, the computational complexity of the proposed methods can be written as the following

form based on the above discussion.

$$\mathcal{O} = \mathcal{O}_{\mathbf{w}_k} + \mathcal{O}_{\mathbf{\Theta}}, \tag{6.72}$$

where $\mathcal{O}_{\mathbf{w}_k}$ and $\mathcal{O}_{\mathbf{\Theta}}$ are the computational complexity associated with the beamforming and phase shift design, respectively. Now, consider the problem in **P16**, which has $3K$ LMI constraints of size $(N+1)$, $(K+2)$ LMI constraints of size N, K LMI constraints of size 1, and $(K+1)$ SOC constraints. Hence, the complexity of solving problem **P16** is given by

$$\begin{aligned}\mathcal{O}_{\mathbf{w}_k} =&\sqrt{\varrho_1}\ln(1/c) + n_1(n_1^2 + 3K((N+1)^3 + n_1(N+1)^2) \\ &+ (K+2)(N^3 + n_1N^2) + 2K + n_1K + 1),\end{aligned} \tag{6.73}$$

where $\varrho_1 = 4KN + 2N + 6K + 2$, $n_1 = \mathcal{O}(KN^2)$. Similarly, the problem in **P19** consists of $3K$ LMI constraints of size $(N+1)$, 1 LMI constraints of size N, $(K+N+1)$ LMI constraints of size 1, and K SOC constraints. Therefore, the complexity of solving problem **P19** is given by

$$\begin{aligned}\mathcal{O}_{\mathbf{\Theta}} =&\sqrt{\varrho_2}\ln(1/c) + n_2(n_2^2 + 3K((N+1)^3 + n_2(N+1)^2) \\ &+ N^3 + n_2N^2 + (K+N+1)(n_2+1) + K),\end{aligned} \tag{6.74}$$

where $\varrho_2 = 3KN + 2N + 6K + 1$, $n_2 = \mathcal{O}(N^2)$.

6.4 Simulation Results

In this section, numerical results evaluate the performance of the proposed robust design, and the proposed AO algorithm. Considering the system setup shown in Fig. 6.2, the BS is assumed to be centered at (0 m, 0 m), whereas the IRS is placed at (120 m, 20 m). While the analysis provided in this chapter is valid for any number of users, the case of two users, i.e., $K = 2$ is considered, which are located at (100 m, 0 m) and (150 m, 0 m), respectively. The channel between the BS and the IRS is $\mathbf{G} = d^{-loss}\mathbf{g}$ [147], where d is the distance in meters between the BS and IRS, and $loss = 2.2$ is the path loss exponent. $\mathbf{g}$ follows a Rayleigh distribution, which are complex Gaussian random variables with zero mean and unit variance modelling the small-scale fading [70]. The channel between the IRS and the users is $\mathbf{h}_l = {d'}^{-loss'}g'$, where d' is the distance in meters between IRS and users, and $loss' = 2.5$. Other parameters are set as $\sigma_l^2 = -80$ dBm, $\epsilon = 0.6$ and the convergence tolerance is 10^{-4} [70, 141, 144].

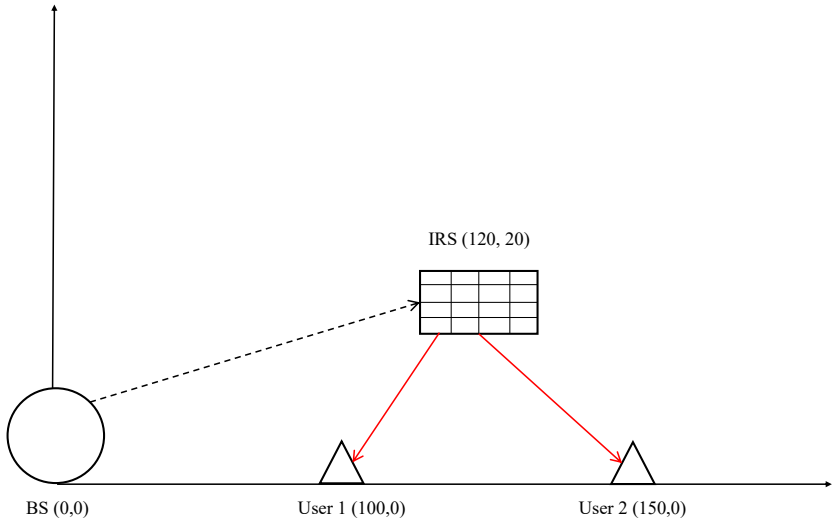

Fig. 6.2 The simulated IRS-assisted NOMA system setup.

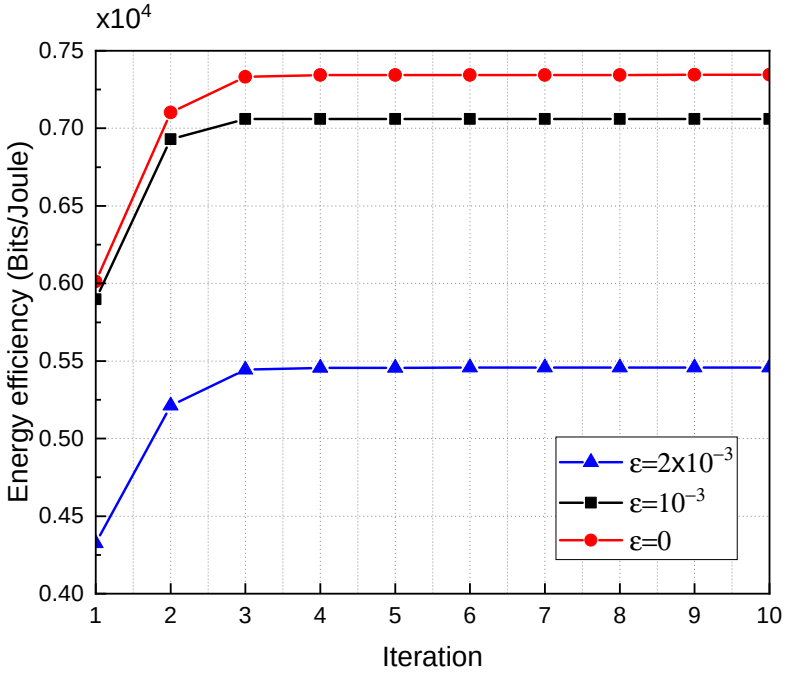

Fig. 6.3 The convergence of the beamforming optimization for different estimation error values, ε.

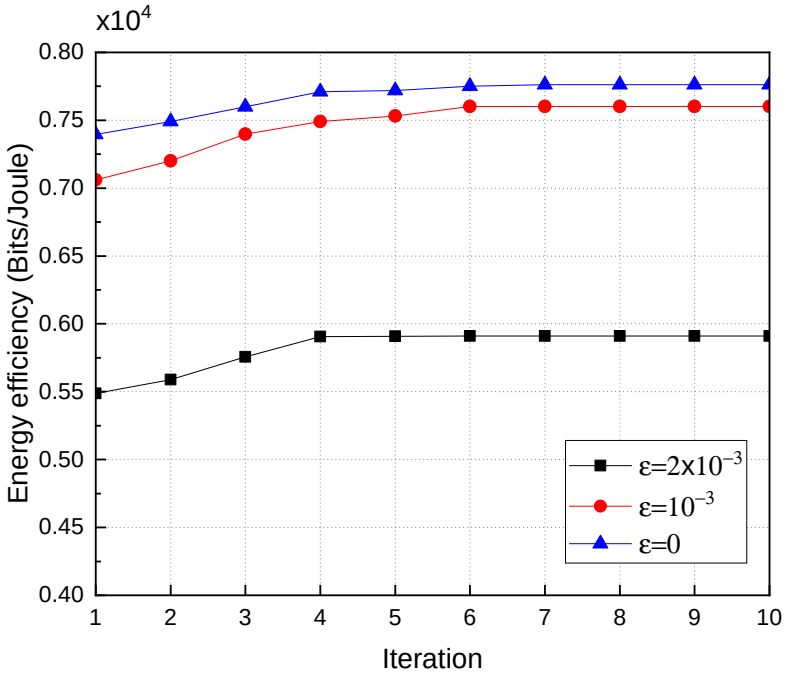

Fig. 6.4 The convergence of the phase shift optimization for different estimation error values, ε.

Fig. 6.3 and Fig. 6.4 show the convergence of the proposed inner iterative algorithms for determining the beamforming vectors and reflection matrix, respectively. In particular, the maximum transmit power $P^{max} = 5$ W and the minimum data rate requirements of all users are assumed to be the same, that is, $R_k^{min} = 2$ bits/s/Hz. It can be observed that at each outer iteration, the inner iterative algorithm of the beamforming optimization and the phase shift optimization tend to converge after a few iterations. Furthermore, it can be seen that the performance of EE decreases with increasing the estimation error ε; $\varepsilon = 0$ reduces to a scenario with perfect CSI between the IRS and users.

To gain insight into the performance of the proposed scheme, Fig. 6.5 shows the achieved EE versus different transmit power and estimation errors with NOMA and with conventional OMA scheme, when $M = 5$ and $N = 4$. The conventional OMA transmission scenario, namely OFDMA, is considered, where the bandwidth is divided into K sub-channels to cover K users. As can be observed, the achieved EE increases at first until it reaches a certain value, and then it remains constant with the maximum value over the different maximum transmit power P^{max}. The EE-Max design with perfect CSI outperforms the imperfect CSI for the same condition. Both beamforming

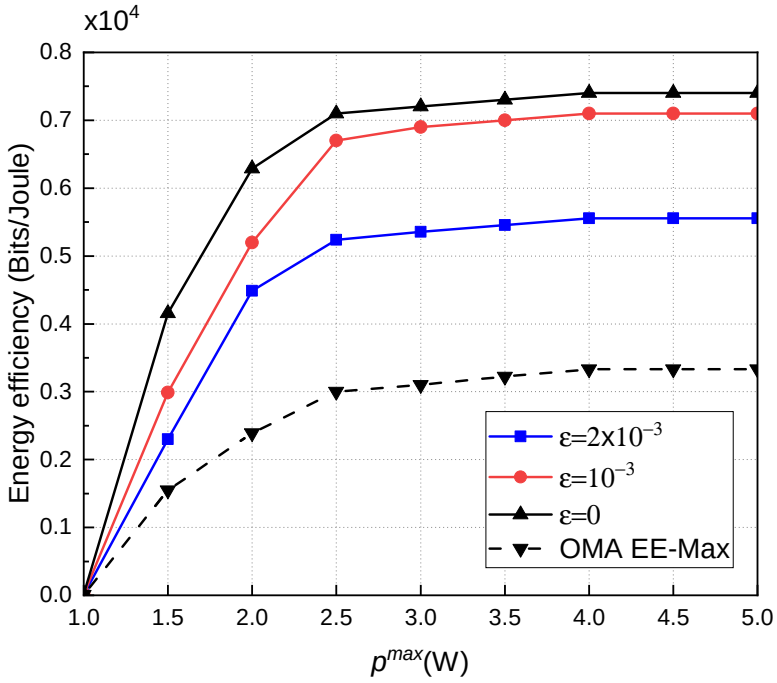

Fig. 6.5 Achieved EE of the proposed algorithm with different transmit power values, P^{max}.

vectors and phase shift are utilized efficiently to achieve the best EE for a given system. In addition, it can be observed that the proposed NOMA schemes provide better EE than the conventional OMA scheme.

In addition, Fig. 6.6 shows the achieved EE for the proposed EE-Max design with respect to the numbers of transmit antennas. The figure indicates that the increase in the number of the transmit antennas does not always provide positive degrees of freedom. The achieved EE is not monotonically increasing with the number of transmit antennas. However, when the number of transmit antennas is larger than a certain threshold, then the performance of EE decreases, even though there are more transmit antennas. The reason can be explained as follows. When M is relatively small, a number of antennas can provide a higher array gain of receiving from the BS, thus improve the performance of the achieved sum rate, which is greater than that of energy consumption related to M. However, when M is relatively large, the energy consumption also increases with the increase of M, which dominates the system performance. It indicates that more transmit antenna is not necessarily better and thus leads to energy cost and limit the overall performance. More importantly, it can be seen that the maximum EE can be achieved with a best M.

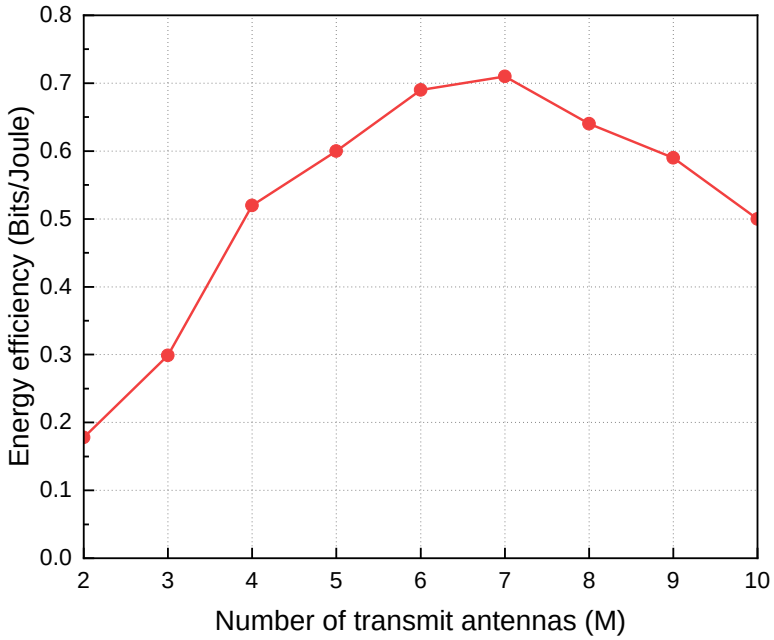

Fig. 6.6 EE versus the number of transmit antenna.

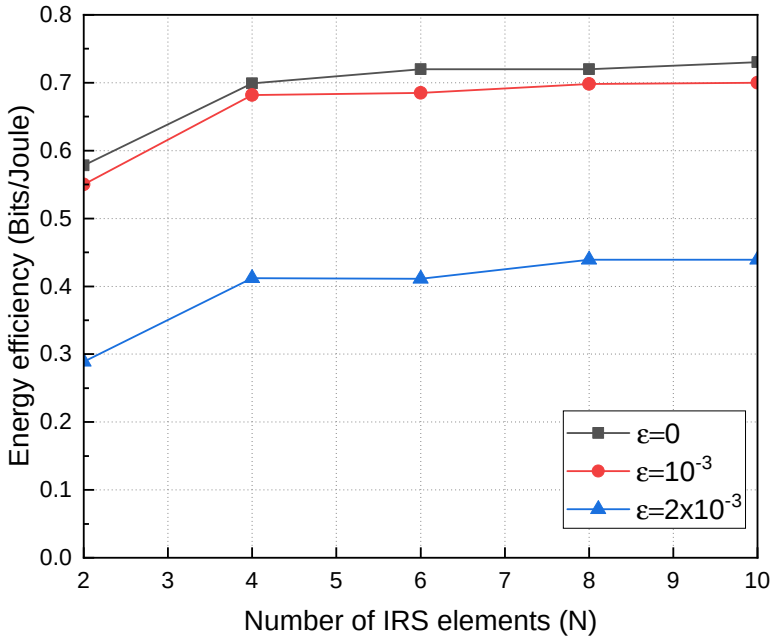

Fig. 6.7 EE versus the number of IRS elements.

The EE of the proposed algorithms versus the number of reflection elements at the IRS is studied in Fig. 6.7. Specifically, it is set $K = 2$, $R_k^{min} = 2$ bit/s/Hz, and $M = 5$. Firstly, it can be observed that the robust algorithms with imperfect CSI model attain much less EE than those under the perfect CSI model. Secondly, the scheme with smaller estimation error achieves more EE than with larger estimation error, i.e., ε. Thirdly, the additional spatial degrees of freedom offered by the increased number of elements of the IRS provides a high flexibility in the beamforming design to enhance the channel quality of the link for improving the system EE. More interestingly, the system EE gradually saturates as the number of IRS elements N increases.

6.5 Summary

In this chapter, the robust design of an IRS-assisted NOMA DL transmission system is studied. The objective of the design is to maximize the EE by jointly optimizing the beamforming vector and the reflection matrix under the assumption of imperfect CSI. To overcome the non-convex nature of this problem, the AO algorithm has been proposed based on the SDP and *S*-procedure. It is shown that the EE performance in terms of the maximum transmit power and the number of elements of the IRS under the perfect CSI model is higher than that under imperfect CSI. Moreover, the additional spatial degrees of freedom offered by the increased number of elements of the IRS provide a high flexibility in beamforming to enhance the channel quality of the link, and thus, improve the system EE.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

NOMA has recently been identified as a promising multiple access technique for the 5G and beyond wireless networks due to its potential benefits including superior SE and user fairness. Unlike the conventional OMA schemes, such as TDMA and OFDMA, multiple users in NOMA based DL transmission share the same orthogonal radio resources, i.e., time and frequency, by exploiting power-domain multiplexing at the transmitter. This multiplexing is referred to as SC, in which signals intended to different users are encoded with different power levels that are inversely proportional to the channel strengths of the users. In particular, serving multiple users simultaneously within the same resource block through NOMA supports the proliferation on IoTs by offering massive connectivity. At the receiver end, SIC technique is utilized at stronger users to decode the signals intended to the weaker users prior to decoding their own signals. Furthermore, NOMA has been recently combined with other multiple access techniques. These include NOMA with multiple antenna, and conventional OMA techniques. These strategic combinations offer additional degrees of freedom, and hence to cultivate its underlying potential benefits. Resource allocation also plays a crucial role in wireless systems and networks design as it enables an efficient use of resources. Therefore, this thesis focused on different resource allocation techniques for different types of NOMA systems.

In Chapter 4, the GEE-Max problem with joint power-time resource allocation for a DL hybrid TDMA-NOMA system was investigated. In particular, the overall EE of a hybrid TDMA-NOMA system has been maximized subject to pre-defined users' minimum rate requirements and the power budget constraint at the BS. An optimization framework has been developed, which allocates the available transmit power at the

BS between the users and opportunistically assigns the available time for transmission between the clusters. However, the formulated GEE-Max optimization problem was non-convex in nature, and could not be solved directly using available standard optimization software. Hence, an iterative approach was proposed for solving the original GEE-Max problem with a novel SOC approach and approximations. Furthermore, an alternative approach based on Dinkelbach's algorithm has been proposed to solve this non-convex problem. Specifically, a new non-negative variable has been introduced to transform the fractional objective function into a non-fractional one. Simulation results have demonstrated the effectiveness of the proposed schemes.

In Chapter 5, by exploiting the performance advantages of the hybrid TDMA-NOMA system in different scenarios, two resource allocation designs were proposed. In the first design, an SE-EE trade-off based resource allocation technique has been investigated. In particular, the original problem has been formulated as an MOO problem with the conflicting objective functions SE and EE. Then, a weighted-sum approach has been utilized to convert the MOO framework into an SOO problem, and thus to obtain the Pareto-optimal solutions. However, the SOO problem has turned out to be a non-convex problem. Therefore, an iterative algorithm has been developed to deal with the non-convexity issues. In the second design, a max-min problem has been considered, in which the aim was to maximize the minimum per user rate for a hybrid TDMA-NOMA system, while satisfying the relevant constraints on the system. This has been achieved through developing an optimization framework to allocate the available transmit power efficiently among users and opportunistically assign the available time for transmission between the clusters (i.e., groups of users). The formulated optimization problem is non-convex in nature, and cannot be solved via available software. Hence, an iterative algorithm has been developed by exploiting SCA. Furthermore, a novel form of a SOC was utilized to cast some of non-convex constraints as SOCs. A number of performance comparisons have been demonstrated the advantages of the proposed hybrid NOMA-TDMA technique over the conventional schemes with equal time allocations.

In Chapter 6, a DL transmission of an IRS-assisted NOMA system was investigated. In contrast to the conventional stand-alone NOMA or IRS design, an IRS has been deployed to support the communication between a multi antenna BS and multiple single-antenna users for a NOMA system, taking into account the imperfect CSI. The EE maximization problem was designed by jointly optimizing the transmit beamforming at BS and the phase shift of IRS. To address this non-convex optimization problem, an iterative algorithm based on the alternating optimization was proposed, where

the beamforming and the phase shift design were alternatively solved by utilizing the S-procedure and SDP relaxation technique. Results also illustrated that the proposed scheme outperformed OMA scheme in terms of the EE. Also, the performance of the proposed robust IRS-assisted NOMA design was provided by comparing that of the design with perfect CSI.

7.2 Future Work

In this section, the potential extensions of the current works in this thesis are described.

7.2.1 Imperfect CSI

The current works in Chapter 4 and 5 in this thesis have assumed that the BS has perfect CSI. The results obtained with this assumption serve as the performance upper bounds which can be also considered as the benchmark schemes to evaluate the performance of other similar schemes. In practice, the perfect CSI assumption might not be realistic due to the significant system overhead and the errors from the channel estimations and quantization. Furthermore, these channel uncertainties might significantly degrade the overall system performance, especially in NOMA systems as the decoding order of the received signal intended for different users plays a crucial role in SIC. Therefore, it is important to incorporate these channel uncertainties in the design to explore the full potential benefits offered by NOMA. In particular, the robust design can be dealt with either the worst-case design with the bounded channel uncertainties, or with the outage probability based design with the statistics of channels and uncertainties. Robust designs to deal with imperfect CSI will be an interesting future direction to extend our work.

7.2.2 Machine Learning-based Resource Allocation Techniques for NOMA Systems

Machine learning algorithms are promising techniques to deal with the computational complexity of solutions. In fact, deep learning and reinforcement techniques have been recently utilized to obtain reasonable solutions for various highly complex optimization problems. In particular, the deep learning-based solutions require a set of reliable training data to train the developed model, which cannot be obtained without solving the original problems using standard optimization techniques. The solution provided in this thesis can be viewed as an approach to generate training data for different deep

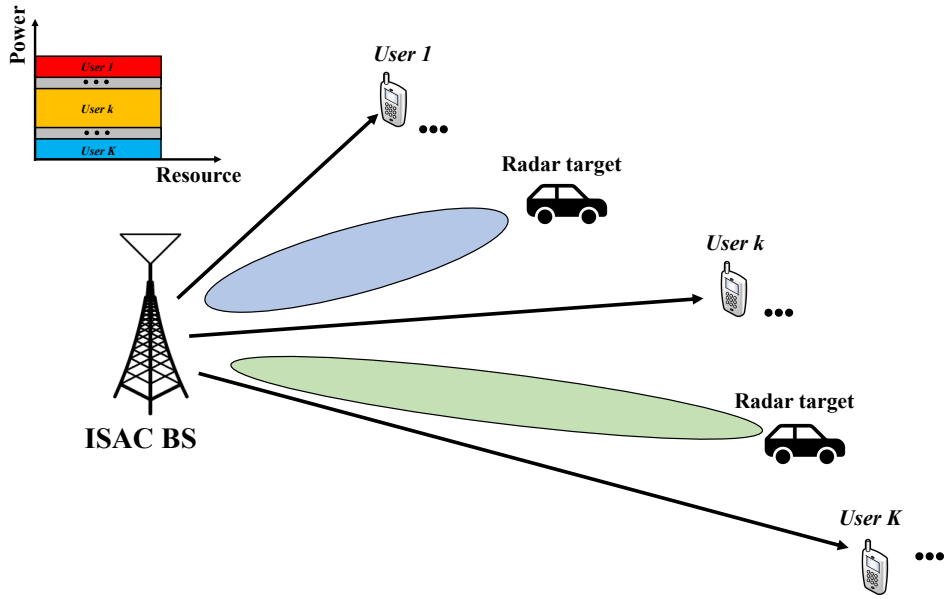

Fig. 7.1 Illustration of the NOMA-ISAC system.

learning models and to validate their performance. Reinforcement learning techniques do not require an external data set during training and they can be exploited to improve the computational complexity while meeting the delay requirements. For the dynamic problem in NOMA systems, the traditional approaches could not provide long-term performance outcomes and extract knowledge from any given problem. Therefore, the reinforcement learning-based approaches are capable to handle this type of issues, where the objective can be chosen as the rewards, the constraints can be assumed to be the states, and thus the feasible region of the constraints are the actions. The optimal action can be determined by the sum of iterative computations. Therefore, this is one of the promising research directions that would be explored in future work resource allocation techniques for NOMA-based systems.

7.2.3 Integrated Sensing and Communications (ISAC) for NOMA Systems

Integrated sensing and communications (ISAC) refers to a design paradigm and corresponding enabling technologies that integrate sensing and communication functionalities to achieve efficient usage of wireless resources and to mutually benefit from each other. ISAC captures two main advantages over dedicated sensing and communication functionalities, which meets the requirements of an unprecedented proliferation of new IoT

services. These IoT services include digital twins, smart city, vehicle-to-everything, and remote sensing. However, striking a good balance between the two functionalities, such as communication performance metrics and sensing and communication performance metrics, is a challenging task when designing ISAC. This is due to the fact that the ISAC may suffer from severe inter-functionality interference in terms of sharing the same spectrum and infrastructure. On the other hand, the ISAC system is more likely to encounter the overloaded regime issue due to the massive device connectivity in the future wireless network. As a remedy, NOMA allows multiple communication devices to be served over the same radio resources, thus significantly enhancing the connectivity and improving the resource efficiency. This can be achieved by employing power-domain multiplexing and the inter-user interference can be mitigated by exploiting SIC in NOMA systems. Motivated by this discussion, a DL NOMA-ISAC system is viewed as an interesting research direction, which consists of a dual-functional BS equipped with N antennas, K single-antenna users, and M radar targets, as shown in Fig.7.1. In other words, the dual-functional BS serves multiple communication users employing NOMA, while the superimposed NOMA communication signal is simultaneously exploited for radar target sensing. Different resource allocation problems could be investigated for this NOMA-ISAC system in the future research.

7.2.4 Age of Information (AoI)-based Resource Allocation Techniques for NOMA Systems

Age of information (AoI) is defined as the time elapsed since the generation time of the latest received status update at the destination, which can quantify the freshness of information. AoI has been attracted more attention in practical cases, such as sensor data needs to be gathered and analyzed to detect the surrounding environment, camera images from the drone need to be collected to generate point clouds that describe road information, and video streams with informative labels need to be updated based on the customer's behaviours. Inspired by the wide range of applications in AoI, it could be applied in the IRS-assisted NOMA system, which has been investigated in Chapter 6. Generally, different approaches can be selected based on the specific use-cases, including time-average AoI, peak AoI, stochastic hybrid systems for AoI, and nonlinear AoI. Hence, the aim is to minimize the average peak AoI for a given set of constraints in the IRS-assisted NOMA system. This would be another interesting future research direction in resource allocation techniques for NOMA systems.

References

[1] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[2] M. Series, "IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU*, vol. 2083, p. 0, 2015.

[3] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 5–36, 2021.

[4] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[5] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE network*, vol. 33, no. 4, pp. 70–75, 2019.

[6] P. Sharma, "Evolution of mobile wireless communication networks-1G to 5G as well as future prospective of next generation communication network," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 8, pp. 47–53, 2013.

[7] L. J. Vora, "Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G," *International journal of modern trends in engineering and research*, vol. 2, no. 10, pp. 281–290, 2015.

[8] H. Honkasalo, K. Pehkonen, M. T. NieMi, and A. T. Leino, "WCDMA and WLAN for 3G and beyond," *IEEE Wireless Communications*, vol. 9, no. 2, pp. 14–18, 2002.

[9] Y. K. Kim and R. Prasad, *4G Roadmap and Emerging Communication Technologies (Universal Personal Communications)*. Artech House, Inc., 2006.

[10] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 668–695, 2021.

[11] J. Wills, "5G technology: Which country will be the first to adapt?" [Online]. Available: https://www.investopedia.com/articles/markets-economy/090916/5g-technology-which-country-will-be-first-adapt.asp.

[12] U. Report, "Connected nations 2020," [Online]. Available: https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2020.

[13] I. Deng, "China targets 2 million installed 5G base stations this year, expanding world's biggest next-generation mobile network, as 6G preparations push ahead," [Online]. Available: https://www.scmp.com/tech/policy/article/3169832/china-targets-2-million-installed-5g-base-stations-year-expanding.

[14] H. Ullah, N. G. Nair, A. Moore, C. Nugent, P. Muschamp, and M. Cuevas, "5G communication: an overview of vehicle-to-everything, drones, and healthcare use-cases," *IEEE Access*, vol. 7, pp. 37 251–37 268, 2019.

[15] M. Series, "Minimum requirements related to technical performance for IMT-2020 radio interface (s)," *Report*, pp. 2410–0, 2017.

[16] G. Liu, Y. Huang, N. Li, J. Dong, J. Jin, Q. Wang, and N. Li, "Vision, requirements and network architecture of 6G mobile network beyond 2030," *China Communications*, vol. 17, no. 9, pp. 92–104, 2020.

[17] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[18] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.

[19] I. A. Alimi, R. K. Patel, N. J. Muga, A. N. Pinto, A. L. Teixeira, and P. P. Monteiro, "Towards enhanced mobile broadband communications: A tutorial on enabling technologies, design considerations, and prospects of 5G and beyond fixed wireless access networks," *Applied Sciences*, vol. 11, no. 21, p. 10427, 2021.

[20] N. Alliance, "5G white paper," *Next generation mobile networks, white paper*, vol. 1, no. 2015, 2015.

[21] H. Kim, *Design and optimization for 5G wireless communications*. John Wiley & Sons, 2020.

[22] 3rd Generation Partnership Project (3GPP), "Study on scenarios and requirements for next generation access technologies," *Version 14.2.0*, 2017.

[23] T. Fehrenbach, R. Datta, B. Göktepe, T. Wirth, and C. Hellge, "URLLC services in 5G low latency enhancements for LTE," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–6.

[24] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 426–471, 2019.

[25] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[26] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[27] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1486–1497, 2017.

[28] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A prospective look: Key enabling technologies, applications and open research topics in 6G networks," *IEEE access*, vol. 8, pp. 174 792–174 820, 2020.

[29] M. Liu, J. Zhang, K. Xiong, M. Zhang, P. Fan, and K. B. Letaief, "Effective user clustering and power control for multi-antenna uplink NOMA transmission," *IEEE Transactions on Wireless Communications*, 2022.

[30] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[31] X. Xie, F. Fang, and Z. Ding, "Joint optimization of beamforming, phase-shifting and power allocation in a multi-cluster IRS-NOMA network," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7705–7717, 2021.

[32] S. Basharat, S. A. Hassan, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surface-assisted backscatter communication: A new frontier for enabling 6G IoT networks," *IEEE Wireless Communications*, 2022.

[33] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[34] 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.

[35] 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.

[36] S. R. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 40–47, 2018.

[37] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[38] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficient beamforming design for MISO non-orthogonal multiple access systems," *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4117–4131, 2019.

[39] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation for MIMO-NOMA with multiple users in a cluster," *IEEE Access*, vol. 6, pp. 5170–5181, 2018.

[40] H. Al-Obiedollah, K. Cumanan, A. G. Burr, J. Tang, Y. Rahulamathavan, Z. Ding, and O. A. Dobre, "On energy harvesting of hybrid TDMA-NOMA systems," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[41] F. Fang, H. Zhang, J. Cheng, and V. C. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3722–3732, 2016.

[42] X. Zhang and F. Wang, "Resource allocation for wireless power transmission over full-duplex OFDMA/NOMA mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 2, pp. 327–344, 2018.

[43] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.

[44] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α-fairness," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2357–2369, 2017.

[45] M. T. Le, G. C. Ferrante, T. Q. Quek, and M.-G. Di Benedetto, "Fundamental limits of low-density spreading NOMA with fading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4648–4659, 2018.

[46] S. Islam, M. Zeng, and O. A. Dobre, "NOMA in 5G systems: Exciting possibilities for enhancing spectral efficiency," *arXiv preprint arXiv:1706.08215*, 2017.

[47] 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*, vol. 2018, 2018.

[48] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE communications surveys & tutorials*, vol. 11, no. 2, pp. 87–105, 2009.

[49] N. Costa and S. Haykin, *Multiple-input multiple-output channel models: theory and practice*. John Wiley & Sons, 2010.

[50] Q. Yang, H.-M. Wang, D. W. K. Ng, and M. H. Lee, "NOMA in downlink SDMA with limited feedback: Performance analysis and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2281–2294, 2017.

[51] S.-M. Tseng, C.-S. Tsai, and C.-Y. Yu, "Outage-capacity-based cross layer resource management for downlink noma-ofdma video communications: Non-deep learning and deep learning approaches," *IEEE Access*, vol. 8, pp. 140 097–140 107, 2020.

[52] H. B. Salameh, R. Tashtoush, H. Al-Obiedollah, A. Alajlouni, and Y. Jararweh, "Power allocation technique with soft performance guarantees in hybrid ofdma–noma cognitive radio systems: Modeling and simulation," *Simulation Modelling Practice and Theory*, vol. 112, p. 102370, 2021.

[53] H. Al-Obiedollah, K. Cumanan, H. B. Salameh, G. Chen, Z. Ding, and O. A. Dobre, "Downlink multi-carrier NOMA with opportunistic bandwidth allocations," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2426–2429, 2021.

[54] J. Luo, J. Tang, D. K. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT," *IEEE Access*, vol. 7, pp. 17 450–17 460, 2019.

[55] Z. Li and J. Gui, "Energy-efficient resource allocation with hybrid TDMA–NOMA for cellular-enabled machine-to-machine communications," *IEEE Access*, vol. 7, pp. 105 800–105 815, 2019.

[56] A. B. Rozario and M. F. Hossain, "Hybrid TDMA-NOMA based M2M communications over cellular networks with dynamic clustering and 3D channel models," in *2019 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*. IEEE, 2019, pp. 1–6.

[57] Z. Hadzi-Velkov, S. Pejoski, N. Zlatanov, and R. Schober, "UAV-assisted wireless powered relay networks with cyclical NOMA-TDMA," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2088–2092, 2020.

[58] R. Khan, I. Ali, M. A. Jan, M. Zakarya, M. A. Khan, S. S. Alshamrani, and M. Guizani, "A hybrid approach for seamless and interoperable communication in the Internet of Things," *IEEE Network*, vol. 35, no. 6, pp. 202–208, 2021.

[59] L. Bai, L. Zhu, X. Zhang, W. Zhang, and Q. Yu, "Multi-satellite relay transmission in 5G: Concepts, techniques, and challenges," *IEEE Network*, vol. 32, no. 5, pp. 38–44, 2018.

[60] B. P. Rimal, D. P. Van, and M. Maier, "Mobile edge computing empowered fiber-wireless access networks in the 5G era," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 192–200, 2017.

[61] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[62] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7170–7183, 2020.

[63] A. Zappone, E. Jorswieck *et al.*, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[64] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE communications magazine*, vol. 49, no. 8, pp. 55–62, 2011.

[65] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *IEEE wireless communications*, vol. 18, no. 5, pp. 40–49, 2011.

[66] B. Xu, Y. Chen, J. R. Carrión, and T. Zhang, "Resource allocation in energy-cooperation enabled two-tier NOMA HetNets toward green 5G," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2758–2770, 2017.

[67] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2082–2085, 2015.

[68] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5119–5131, 2019.

[69] M. Zeng, N.-P. Nguyen, O. A. Dobre, Z. Ding, and H. V. Poor, "Spectral-and energy-efficient resource allocation for multi-carrier uplink NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9293–9296, 2019.

[70] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14 088–14 092, 2020.

[71] J. Wang, H. Xu, L. Fan, B. Zhu, and A. Zhou, "Energy-efficient joint power and bandwidth allocation for NOMA systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 780–783, 2018.

[72] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2852–2857, 2016.

[73] H. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficiency fairness beamforming designs for MISO NOMA systems," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[74] K. Cumanan, Z. Ding, B. Sharif, G. Y. Tian, and K. K. Leung, "Secrecy rate optimizations for a MIMO secrecy channel with a multiple-antenna eavesdropper," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 4, pp. 1678–1690, 2013.

[75] J. Cui, Z. Ding, and P. Fan, "Power minimization strategies in downlink MIMO-NOMA systems," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[76] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *2015 IEEE international conference on communications (ICC)*. IEEE, 2015, pp. 2668–2673.

[77] H. Al-Obiedollah, K. Cumanan, H. B. Salameh, S. Lambotharan, Y. Rahulamathavan, Z. Ding, and O. A. Dobre, "A joint beamforming and power-splitter optimization technique for SWIPT in MISO-NOMA system," *IEEE Access*, vol. 9, pp. 33 018–33 029, 2021.

[78] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, L. Hanzo, and P. Xiao, "On the performance of cell-free massive MIMO relying on adaptive NOMA/OMA mode-switching," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 792–810, 2019.

[79] A. J. Muhammed, Z. Ma, P. D. Diamantoulakis, L. Li, and G. K. Karagiannidis, "Energy-efficient resource allocation in multicarrier NOMA systems with fairness," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8639–8654, 2019.

[80] X. Wei, H. Al-Obiedollah, K. Cumanan, M. Zhang, J. Tang, W. Wang, and O. A. Dobre, "Resource allocation technique for hybrid TDMA-NOMA system with opportunistic time assignment," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.

[81] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, J. Tang, A. G. Burr, Z. Ding, and O. A. Dobre, "Spectral-energy efficiency trade-off-based beamforming design for MISO non-orthogonal multiple access systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6593–6606, 2020.

[82] Z. Liu, W. Du, and D. Sun, "Energy and spectral efficiency tradeoff for massive MIMO systems with transmit antenna selection," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4453–4457, 2016.

[83] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1119–1123, 2020.

[84] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *IEEE Communications Letters*, vol. 25, no. 1, pp. 234–238, 2020.

[85] Z. Zhang, L. Lv, Q. Wu, H. Deng, and J. Chen, "Robust and secure communications in intelligent reflecting surface assisted NOMA networks," *IEEE Communications Letters*, vol. 25, no. 3, pp. 739–743, 2020.

[86] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Communications Letters*, vol. 24, no. 4, pp. 753–757, 2020.

[87] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDMA: Joint resource allocation and passive beamforming optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 760–764, 2020.

[88] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on selected areas in communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

[89] Z.-Q. Luo, "Applications of convex optimization in signal processing and digital communication," *Mathematical programming*, vol. 97, no. 1, pp. 177–207, 2003.

[90] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[91] Z. Han and K. R. Liu, *Resource allocation for wireless networks: basics, techniques, and applications*. Cambridge university press, 2008.

[92] V. G. Douros and G. C. Polyzos, "Review of some fundamental approaches for power control in wireless networks," *Computer Communications*, vol. 34, no. 13, pp. 1580–1592, 2011.

[93] Z. Wei, D. W. K. Ng, and J. Yuan, "Power-efficient resource allocation for MC-NOMA with statistical channel state information," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–7.

[94] S. Zeb, Q. Abbas, S. A. Hassan, A. Mahmood, R. Mumtaz, S. H. Zaidi, S. A. R. Zaidi, and M. Gidlund, "NOMA enhanced backscatter communication for green IoT networks," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2019, pp. 640–644.

[95] K. Chi, X. Wei, Y. Li, and X. Tian, "Throughput maximization in wireless powered communication networks with minimum node throughput requirement," *International Journal of Communication Systems*, vol. 31, no. 15, p. e3775, 2018.

[96] B. Radunovic and J.-Y. Le Boudec, "A unified framework for max-min and min-max fairness with applications," *IEEE/ACM Transactions on networking*, vol. 15, no. 5, pp. 1073–1083, 2007.

[97] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 782–785, 2017.

[98] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 30–37, 2011.

[99] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5574–5588, 2015.

[100] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[101] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.

[102] N. J. Higham, *Functions of matrices: theory and computation*. SIAM, 2008.

[103] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Transactions on evolutionary computation*, vol. 8, no. 3, pp. 256–279, 2004.

[104] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and multidisciplinary optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[105] P. J. Copado-Méndez, C. Pozo, G. Guillén-Gosálbez, and L. Jiménez, "Enhancing the ϵ-constraint method through the use of objective reduction and random sequences: Application to environmental problems," *Computers & Chemical Engineering*, vol. 87, pp. 36–48, 2016.

[106] N. T. Do, D. B. da Costa, T. Q. Duong, V. N. Q. Bao, and B. An, "Exploiting direct links in multiuser multirelay SWIPT cooperative networks with opportunistic scheduling," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5410–5427, 2017.

[107] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 76–88, 2015.

[108] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *MILCOM 2013-2013 IEEE Military Communications Conference*. IEEE, 2013, pp. 1278–1283.

[109] 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.

[110] S. Boyd, "Sequential convex programming," [Online]. Available: http://www.stanford.edu/class/ee364b/lectures/seq slides.pdf.

[111] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.

[112] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3711–3724, 2009.

[113] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.

[114] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear algebra and its applications*, vol. 284, no. 1-3, pp. 193–228, 1998.

[115] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1407–1421, 2020.

[116] Y. Huang, S. He, J. Wang, and J. Zhu, "Spectral and energy efficiency tradeoff for massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 6991–7002, 2018.

[117] C. He, B. Sheng, P. Zhu, X. You, and G. Y. Li, "Energy-and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness," *IEEE Journal on Selected areas in Communications*, vol. 31, no. 5, pp. 894–902, 2013.

[118] K. Wang, W. Liang, Y. Yuan, Y. Liu, Z. Ma, and Z. Ding, "User clustering and power allocation for hybrid non-orthogonal multiple access systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12 052–12 065, 2019.

[119] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, vol. 104, no. 2, pp. 154–171, 2013.

[120] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE access*, vol. 4, pp. 6325–6343, 2016.

[121] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, "Multiobjective evolutionary algorithms: A survey of the state of the art," *Swarm and evolutionary computation*, vol. 1, no. 1, pp. 32–49, 2011.

[122] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE signal processing letters*, vol. 22, no. 10, pp. 1647–1651, 2015.

[123] Z. Wei, D. W. K. Ng, J. Yuan, and H.-M. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3944–3961, 2017.

[124] S. Boyd, "CVX: Matlab software for disciplined convex programming," [Online]. Available: http://cvxr.com/cvx/.

[125] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1263–1266, 2016.

[126] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

[127] J. Zhao, "A survey of intelligent reflecting surfaces (IRSs): Towards 6G wireless communication networks," *arXiv preprint arXiv:1907.04789*, 2019.

[128] X. Wei, H. Al-Obiedollah, K. Cumanan, Z. Ding, and O. A. Dobre, "Energy efficiency maximization for hybrid TDMA-NOMA system with opportunistic time assignment," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8561–8573, 2022.

[129] X. Wei, H. Al-Obiedollah, K. Cumanan, W. Wang, Z. Ding, and O. A. Dobre, "Spectral-energy efficiency trade-off based design for hybrid TDMA-NOMA system," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 3377–3382, 2022.

[130] F. Alavi, K. Cumanan, M. Fozooni, Z. Ding, S. Lambotharan, and O. A. Dobre, "Robust energy-efficient design for MISO non-orthogonal multiple access systems," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7937–7949, 2019.

[131] 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.

[132] S. Guo and X. Zhou, "Robust resource allocation with imperfect channel estimation in NOMA-based heterogeneous vehicular networks," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2321–2332, 2018.

[133] J. Cui, Z. Ding, and P. Fan, "Outage probability constrained MIMO-NOMA designs under imperfect CSI," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8239–8255, 2018.

[134] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Beamforming techniques for nonorthogonal multiple access in 5G cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9474–9487, 2018.

[135] Y. Chen, M. Wen, L. Wang, W. Liu, and L. Hanzo, "SINR-outage minimization of robust beamforming for the non-orthogonal wireless downlink," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7247–7257, 2020.

[136] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Robust beamforming techniques for non-orthogonal multiple access systems with bounded channel uncertainties," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2033–2036, 2017.

[137] W. Hao, G. Sun, M. Zeng, Z. Chu, Z. Zhu, O. A. Dobre, and P. Xiao, "Robust design for intelligent reflecting surface-assisted MIMO-OFDMA Terahertz IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 13 052–13 064, 2021.

[138] P. Song, G. Scutari, F. Facchinei, and L. Lampariello, "D3M: Distributed multi-cell multigroup multicasting," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3741–3745.

[139] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, 2016.

[140] Z. Chu, K. Cumanan, M. Xu, and Z. Ding, "Robust secrecy rate optimisations for multiuser multiple-input-single-output channel with device-to-device communications," *IET Communications*, vol. 9, no. 3, pp. 396–403, 2015.

[141] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[142] Z.-Q. 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.

[143] M. Bengtsson and B. Ottersten, "Optimal transmit beamforming using convex optimization," *submitted to IEEE Transactions on Communications*, 1999.

[144] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Transactions on Signal Processing*, vol. 68, pp. 705 092–5106, 2020.

[145] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[146] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5690–5705, 2014.

[147] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6884–6898, 2020.