Resource Allocation Techniques for Non-Orthogonal Multiple Access Scheme for 5G and Beyond Wireless Networks

Seyedeh Faezeh Alavi

Ph.D.

University of York

Electronic Engineering

October 2018

Abstract

The exponential growth of wireless networks and the number of connected devices as well as the emergence of new multimedia based services have resulted in growing demands for high data-rate communications, and a spectrum crisis. Hence, new approaches are required for better utilization of spectrum and to address the high data-rate requirements in future wireless communication systems. Non-orthogonal multiple access (NOMA) has been envisioned as a promising multiple access technique for 5G and beyond wireless networks due to its potential to achieve high spectral efficiency (SE) and energy efficiency (EE) as well as to provide massive connectivity in supporting proliferation of Internet of Things. In NOMA, multiple users can share the same wireless resources by applying superposition coding (SC) and power domain multiplexing at the transmitter and employing successive interference cancellation (SIC) technique at the receiver for multi-user detection. NOMA outperforms conventional orthogonal multiple access (OMA) by simultaneously sharing the available communication resources between all users via the power domain multiplexing which offers a significant performance gain in terms of SE.

In this thesis, several resource allocation problems have been addressed in NOMA based communication systems, in order to improve network performance in terms of power consumption, fairness and EE. In particular, the NOMA scheme has been studied in multiple-input-single-output transmissions where transmit beamformers are designed to satisfy quality of service using convex optimization techniques. To incorporate the channel uncertainties in beamforming design, robust schemes are proposed based on the worst-case design and the outage probabilistic-based design. Finally, the EE is investigated for non-clustering and clustering NOMA schemes with imperfect channel state information. To eliminate the interference between different clusters, zero-forcing beamformers are employed at the base station. Theoretical analysis and algorithmic solutions are derived and the performance of all these schemes has been verified using simulation results.

List of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Kanapathippillai Cumanan, for his precious time, patient encouragement, and invaluable advice. I have learned many valuable lessons from such an outstanding researcher who always selflessly shares his research experience and expertise with me. This gratitude also goes to my co-supervisor, Prof. Alister G. Burr, for his kind support and guidance. Indeed, the accomplishment of this study would not have been feasible without their kind support and cooperation.

Last and most importantly, I would like to extend my heartfelt gratitude to my ever faithful parents, who have always been a source of moral and emotional support for me throughout my life and from start to the end of this thesis. I would like to sincerely appreciate my sister, Atefeh who has never left my side and is very special. Finally, my special feeling of gratitude goes out to my husband, Milad, for all of his always encouragement, support, and love. All that I am or hope to be I owe to my family, and I feel deeply indebted to them. Indeed, without them, I would never have gotten to this point and this thesis would never have been written.

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. The research presented in this thesis features in a number of the author's publications listed on the next page.

Seyedeh Faezeh Alavi
October 2018

List of publications

1. F. Alavi, K. Cumanan, M. Fozooni, Z. Ding and A. G. Burr, “Energy-Efficient Resource Allocation in NOMA System with Imperfect CSI," to be submitted in *IEEE Transactions on Wireless Communications*.

2. 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," submitted in *IEEE Transactions on Communications*.

3. F. Alavi, K. Cumanan, Z. Ding and A. G. Burr, “Beamforming Techniques for Non-Orthogonal Multiple Access in 5G Cellular Networks," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9474-9487, October 2018.

4. F. Alavi, K. Cumanan, Z. Dingl and A. G. Burr, “Outage constraint based robust beamforming design for non-orthogonal multiple access in 5G cellular networks," *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC, Canada, 2017, pp. 1-5.

5. F. Alavi, K. Cumanan, Z. Ding and A. G. Burr, “Robust Beamforming Techniques for Non-Orthogonal Multiple Access Systems with Bounded Channel Uncertainties," in *IEEE Communications Letters*, vol. 21, no. 9, pp. 2033-2036, September 2017.

6. F. Alavi, N. M. Yamchi, M. R. Javan and K. Cumanan, “Limited Feedback Scheme for Device-to-Device Communications in 5G Cellular Networks with Reliability and Cellular Secrecy Outage Constraints," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8072-8085, September 2017.

7. M. R. Javan, N. Mokari, F. Alavi and A. Rahmati, “Resource Allocation in Decode-and-Forward Cooperative Communication Networks With Limited Rate Feedback Channel," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 256-267, January 2017.

8. N. Mokari, F. Alavi, S. Parsaeefard and T. Le-Ngoc, “Limited-Feedback Resource Allocation in Heterogeneous Cellular Networks," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2509-2521, April 2016.

9. F. Alavi and H. Saeedi, “Radio resource allocation to provide physical layer security in relay-assisted cognitive radio networks," in *IET Communications,*, vol. 9, no. 17, pp. 2124-2130, 2015.

10. F. Alavi, N. Mokari and H. Saeedi, “Secure resource allocation in OFDMA-based cognitive radio networks with two-way relays," in *23rd Iranian Conference on Electrical Engineering*, Tehran, Iran, 2015, pp. 171-176.

11. F. Alavi, N. Mokari and H. Saeedi, “Secure Resource Allocation in Cooperative Cognitive Radio Systems," in *17th Iranian Student Conference on Electrical Engineering (ISCEE)*, Sharif University of Technology-International Campus-Kish Island, Iran, December 2014.

Chapter 1

Introduction

1.1 Background and Motivations

Over the past few decades, mobile devices and wireless technologies have transformed people's lives in significant way. They continue to underpin people's lives and the infrastructure of the networked society that people live in today. As such, they play a crucial role in people's day-to-day life, services and the means of our interactions. The first generation (1G) of cellular wireless systems was developed in the 1980s based on analog technology and provided basic voice services [1]. To handle the increasing demand for data transmission and high quality of communications, the digital cellular technology was integrated into the second generation (2G) mobile telecommunication networks by the early 1990s. The 2G systems provided more capacity, enhanced sound quality and offered more data services than the 1G systems [2]. Subsequently, the introduction of the general packet radio service (GPRS) became the major step in the evolution of wireless networks towards the third generation (3G) telecommunication technology which was introduced in 1998 [3]. The 3G networks enabled users to access the Internet over mobile devices. Therefore, video calls, mobile TV, online games and location-based services were developed through different applications by devoting more bandwidth, and consequently, higher data transmission. In March 2008, the International Telecommunications Union-Radio (ITU-R) communications sector specified the International Mobile Telecommunications (IMT)-Advanced requirements for the fourth generation (4G) mobile telecommunication technology standards [4]. A 4G system not only supports voice and other 3G services but also provides a wide range of exciting applications such as fast bulky file transfer, high definition (HD) mobile TV, multimedia teleconferencing and online game services. Given a historical 10-year

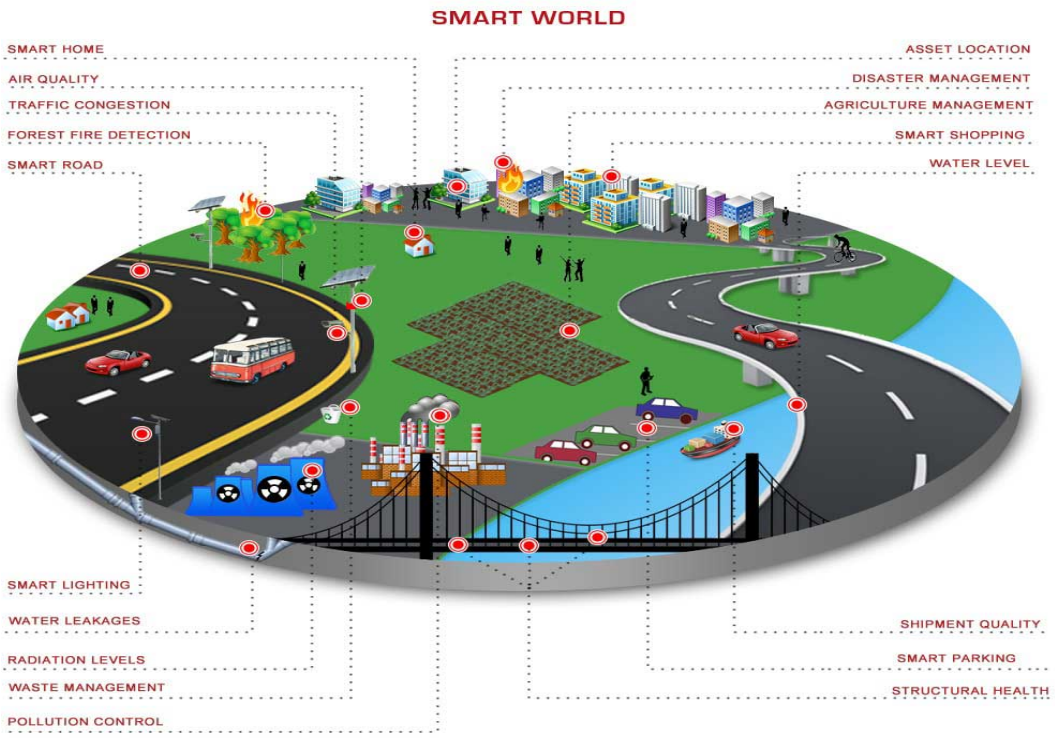


Fig. 1.1 Smart world [6].

cycle for every generation of cellular networks, it is expected that the fifth generation (5G) wireless networks will roll out by 2020 [5]. However, the demanding requirements for high quality new services with stringent delay requirements should be met in 5G and beyond wireless networks.

In the near future, people are expected to completely rely on different services and applications provided by the smart world, as illustrated in Fig.1.1 where connectivity will become ubiquitous [7]. This smart world is a result of the growing interest in area of machine-to-machine (M2M) communications and Internet-of-Things (IoT) which have stringent requirements for reliability, latency and seamless connectivity [8, 9]. The autonomous vehicles which operate in real time environments to prevent accidents and industrial manufacturing processes controlled through wireless networks are only a few examples of ultra-reliable and low-latency communications in IoT networks. On the other hand, it is expected in the next ten years of wireless communications evolution, the continual growth of mobile devices including smartphones and other data-consuming wireless network devices will introduce a $1,000$-fold increase on the volume of mobile data traffic [10]. In order to handle this explosive growth of data traffic, future wireless networks are also expected to provide a $1,000$-fold capacity growth

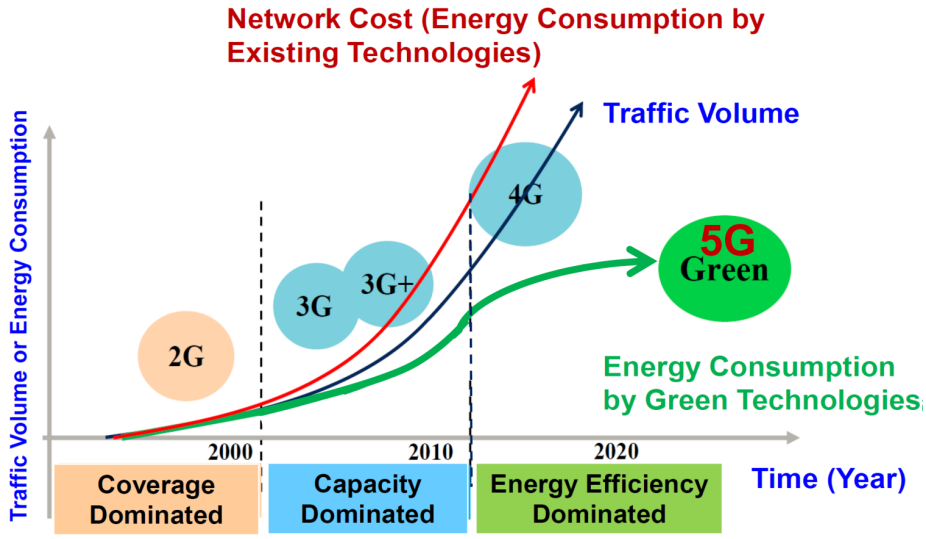


Fig. 1.2 Sustainable green communications evolution for 5G by 2020 [16].

compared to the current generation of wireless networks. These requirements unfold many new challenges, such as the massive increase in the number of connected devices with diverse service requirements, the various applications, and the volume of traffic demands and also uniform quality of user experience anywhere and any time [11–13]. Furthermore, the radio spectrum has been heavily occupied and become overcrowded [14, 15]. Thus, the wireless networks encounter a scarcity of radio resources to support this massive connectivity with high data rate applications. In order to address these issues, future wireless systems should be developed with extended capabilities to efficiently use the available limited resources to satisfy a massive number of users with high data rate requirements. Furthermore, this explosive growth of data traffic, as illustrated in Fig.1.2, has triggered a rapid increase in energy consumption. Statistics show that the information and communication technology infrastructures consume more than 4% of the world-wide energy [17–19]. This energy consumption can indirectly increase the amount of greenhouse gas emission levels. Beyond environmental contamination, the cost of high energy consumption imposes further financial pressure on the network operators. Thus, improving energy efficiency (EE) will be another major challenge that should be appropriately considered in designs of future wireless networks. According to diverse set of requirements in future wireless networks, the most relevant performance targets to 5G, depicted in Fig.1.3, can be summarized as follows:

- Massive device connectivity

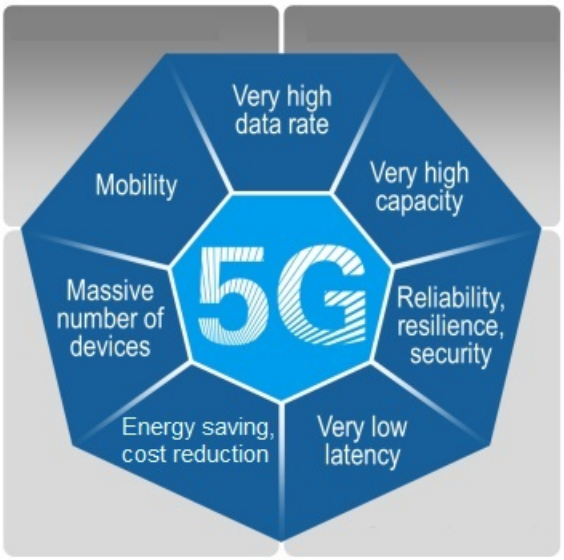


Fig. 1.3 5G performance targets [2].

- Higher system capacity
- Higher data-rate
- Low latency
- Energy saving and cost reduction

As the current wireless communication technologies do not have the capabilities to handle the future wireless networks requirements, advanced technologies and intelligent radio resource management techniques have to be developed for 5G and beyond wireless networks [20]. To further meet the demanding requirements of data rates and massive connectivity, different potential candidate technologies have been proposed in recent years such as massive multiple-input multiple-output (MIMO) [21], millimeter wave (mm-Wave) communications [22], device-to-device (D2D) communications [23], and non-orthogonal multiple access (NOMA) [24].

Employing multiple antennas at the transceivers is a proposed technique to address the increased data rate demand in future wireless networks [25–28]. MIMO, using multiple antennas at both transmitter and receiver, is a well known technique for multiplying the capacity of a radio link and offers generous array and multiplexing gains by leveraging the spatial selectivity [29–32]. MIMO techniques can further improve SE and transmission reliability by simultaneously transmitting multiple data streams in wireless communication systems. Hence, this technology can support a high-speed and reliable link through the additional degrees of freedom introduced in the propa-

gation channel. A MIMO system with an especially high number of antennas called massive MIMO is a promising technique for the next generation of wireless communication systems which has the potential capabilities to address many of the future wireless networks'challenges [33–36]. Deploying BSs with very large numbers of antennas can effectively increase both the beamforming gain and diversity gain, which consequently, can increase the SE [37, 38]. The BSs can concentrate their radiated energy on particular directions by forming very narrow beams since a large number of antennas is utilized in massive MIMO. Through this benefit, massive MIMO can enhance the EE in the order of 100 times compared to conventional MIMO systems [39]. Massive MIMO makes use of favorable propagation channel and the channel hardening property [40]. Hence, very simple and linear signal processing techniques achieve the same performance of the complicated non-linear signal processing methods. Latency is one of the key challenges of current wireless communication systems. It usually occurs in multipath environments with strong destructive interference while the channel experiences deep fading. This phenomenon is more prohibitive in slow fading channels, when the channel may experience a deep fading and the receiver has to wait for a longtime for the channel to recover its good condition and provide a reasonable gain again. However, thanks to the law of large numbers, massive MIMO makes use of the channel hardening with less gain fluctuation compared to the conventional MIMO channels. Thus, massive MIMO can naturally contribute to design of low-latency wireless links as it is unlikely that a destination terminal is trapped in a deep fading. Therefore, massive MIMO can bring improvements in SE, EE, complexity of signal processing, and latency for 5G and beyond wireless networks [21].

mm-Wave communication is another advanced physical layer technology which is expected to address the challenges of providing high-rate services and reducing latency in 5G and beyond wireless networks. Network capacity enhancement is one of the major challenges in future wireless networks and including more spectrum is a straightforward method to increase the capacity in future wireless networks [2]. Most current wireless networks operate in low frequency bands below 6 GHz which have a wide area coverage and low penetration losses however they are already licensed for different services and communication technologies. On the other hand, much wider spectrum is available in the untapped mm-Wave frequency bands, between 30 GHz and 300 GHz, which can potentially offer more bandwidth to provide peak data rates beyond 10 Gbits/s. The short propagation range in this frequency bands can be an advantage which allows the frequency reuse in smaller distances than at lower frequencies and provides improvements in SE. This technique also offers spatial degrees

of freedom with very high-dimensional antenna arrays due to the smaller size of antenna elements at higher frequencies. Furthermore, smaller antenna size also allows the integration of multiple arrays with mobile devices which provides more throughput enhancements for mobile devices and maintains reliable connectivity even if the signal from one array is blocked [22]. mm-Wave communication has several merits such as extremely wide bandwidths, small element sizes, and narrow beams compared to the existing wireless technologies. Due to its potential benefits of multi-gigabit and low latency wireless links, mm-Wave communication is expected to play a crucial role in 5G and beyond wireless networks.

The next proposed technique to achieve higher SE and reduce latency in future wireless networks is enabling direct communications between devices, referred to D2D communications in the literature [23, 41]. In this technique, different mobile devices can directly establish communication links without involving network infrastructure which paves the way to efficiently reuse the spectrum. Through these D2D communications, the overall system capacity can be improved by accommodating more connected devices. In fact, leveraging additional D2D links can potentially scale the network with the number of user devices contributing improvement in both throughput and quality of service (QoS). Augmenting wireless networks with D2D communication can provide robustness as users can continue to function even in the absence of infrastructure. Also, as users in the D2D link communicate over a short distance, the capacity of the direct link tends to be higher even at the low transmit power and even more importantly the same time and frequency resources can be reused in the network. Hence, both energy and spectral resources can also be efficiently exploited by direct communication [42, 43]. Accordingly, D2D communication is considered as a promising technology to provide low-power, high data rate and low-latency services as well as to address the massive connectivity issues in the 5G and beyond wireless networks.

Multiple access techniques play a key role in handling data traffic in multi-user systems as it directly determines the throughput performance by efficiently accommodating multiple users with the available resources [44–46]. Consequently, a standard multiple access scheme is introduced as a specific feature for each generation of wireless network such as time division multiple access (TDMA) in 2G, code division multiple access (CDMA) in 3G, and orthogonal frequency division multiple access (OFDMA) in 4G. These conventional schemes employ orthogonal multiple access (OMA) techniques in which orthogonal resources such as time, frequency and code are assigned to different users to avoid mutual interference between them. In current wireless systems, orthogonal frequency division multiplexing (OFDM) has been widely employed

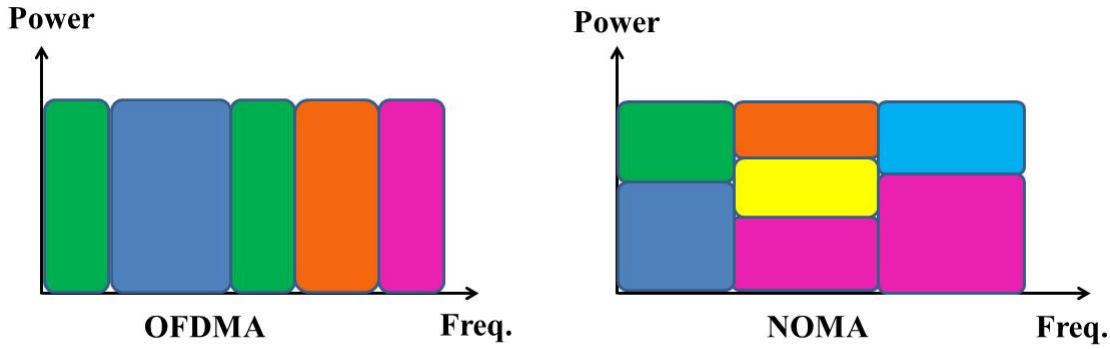


Fig. 1.4 An illustration of OMA and NOMA schemes [52].

due to its ability to cope with the frequency selective fading channel by dividing it into a number of narrow-band flat fading sub-channels [47, 48]. One multiple access scheme based on OFDM is OFDMA where each subcarrier is allocated to only one user. Through this subcarrier assignment, the system throughput is maximized by utilizing multi-user diversity gain [49–51]. However, the networks with massive connectivity and high data-rate requirements may suffer from the performance of OFDMA since it is not allowed to reuse frequency within one cell which significantly limits cell throughput. Although this approach allows simple transceiver implementations, the low complexity of the system comes at the cost of low spectral efficiency (SE) and low EE [45].

Recently, NOMA has been envisioned as a promising multiple access candidate to address these high data rate requirements as well as to support the massive connectivity in 5G and beyond networks [53]. In contrast to OFDMA, as shown in Fig.1.4, NOMA can simultaneously allocate an available radio resource to more than one user which significantly enhances the system throughput due to frequency reuse within a cell [54]. In this scheme, multiple users are allowed to efficiently and simultaneously share time and frequency resources via the power or code domain. Hence, NOMA can offer different advantages including improved SE, higher cell-edge throughput, and low transmission latency. The available NOMA techniques can broadly be divided into two categories, namely, power-domain and code-domain NOMA:

- **Power-domain NOMA:** Through this approach, different users are served at different transmit power levels according to their channel conditions to obtain the maximum gain in system performance. So that the user with lower channel

gain is served with higher transmit power whilst less transmit power is allocated to the user with high channel gain. To carry out this power domain multiplexing, the base station (BS) transmits a linear superposition of the signals of the users and at the receiver sides, multiuser detection algorithms such as successive interference cancellation (SIC) are utilized to detect the desired signals [55, 56]. In particular, the NOMA scheme allows controllable interference and allocates non-orthogonal resources to increase system throughput and serve more users while introducing a reasonable additional complexity at the receivers [24].

- **Code-domain NOMA:** Unlike power-domain NOMA, which attains multiplexing in power domain, code-domain NOMA achieves multiplexing in code domain. Similar to the basic CDMA systems, code-domain NOMA shares the entire available resources in time and frequency. In contrast, code-domain NOMA utilizes user-specific spreading sequences that are either sparse sequences or non-orthogonal cross-correlation sequences of low correlation coefficient [56]. This can be further divided into a few different classes, such as low-density spreading CDMA (LDS-CDMA) [57], low-density spreading-based OFDM (LDS-OFDM) [24], and sparse code multiple access (SCMA) [58]. The use of low-density spreading sequences helps LDS-CDMA to limit the impact of interference on each chip of basic CDMA systems. LDS-OFDM can be thought of as an amalgamation of LDS-CDMA and OFDM, where the information symbols are first spread across low-density spreading sequences and the resultant chips are then transmitted on a set of subcarriers. SCMA is a recent code-domain NOMA technique based on LDS-CDMA. In contrast to LDS-CDMA, the information bits can be directly mapped to different sparse codewords, because both bit mapping and bit spreading are combined. In comparison with LDS-CDMA, SCMA provides a low complexity reception technique and offers improved performances.

There exist some other multiple access techniques, which are also closely-related to NOMA, including multi-user shared access (MUSA) [59], pattern division multiple access (PDMA) [60] and spatial division multiple access (SDMA) [61]. Theses non-orthogonal schemes share the same idea that the multiple users can simultaneously use the same subchannels [24].

Furthermore, in comparison with conventional user scheduling techniques which tend to allocate more transmit power to the users with stronger channel gain to improve the overall system throughput but exacerbate unfairness, NOMA enables a more flexible management of the achievable rate of users and maintains better fairness by

simultaneously serving all users. In other words, NOMA, in fact, facilitates a balanced tradeoff between system throughput and user fairness [62]. In addition to the higher SE realised in NOMA, it is also compatible with other disruptive communication technologies that proposed for future wireless communications such as multiple antenna [63], distributed antenna systems [64–66] and heterogeneous networks [67–69] without any additional hardware requirements. In particular, without requiring any modifications to the Long Term Evolution (LTE) resource blocks (i.e., OFDMA subcarriers), by using the NOMA scheme, different users can be simultaneously served on the same OFDMA subcarrier. Combination of NOMA with other communication technologies will also provide the additional frequency reuse gain to further improve the performance of wireless networks. The two advantages of reusing the limited spectrum and flexibility to be implemented with other technologies have enabled NOMA as a potential technology to develop for future wireless communications [70]. It is unlikely that a single technology would have the potential capabilities to address the exponential growth in the volume of data traffic and the number of connected devices in future wireless networks. Hence, a combination of different communication technologies need to be introduced in order to enable high data rate communications and to meet other unprecedented requirements in 5G and beyond wireless networks [71–73].

1.2 Thesis Contributions and Organization

Motivated by the challenges of increasing SE and EE in 5G and beyond wireless networks, this thesis studies NOMA transmissions and investigates its performance for future wireless networks. This thesis focuses on the power-domain NOMA that superposes multiple users in power domain and exploits the channel gain difference between multiplexed users. At the transmitter side, signals from various users are superposed and the resulting signal is then transmitted over the same time and frequency resources. At the receiver sides, SIC is utilized to efficiently remove interference between multi-user signals. The main aim of this thesis is to explore different power allocation techniques in multiple-input single-output (MISO) NOMA systems to improve system performance in 5G communication systems. In particular, beamforming designs are proposed for power minimization and user fairness problems as well as robust EE design by incorporating channel uncertainties. The scope of the thesis includes mathematical modeling of the radio resource optimization problems, analysis of the problems'complexities, algorithm development, as well as theoretical and nu-

merical analysis. This thesis consists of seven chapters and the main contributions of each chapter are summarised as follows:

In Chapter 2, fundamental concepts of superposition coding (SC) with SIC for NOMA communications are presented. In particular, the downlink channel capacity of NOMA for the general K user system and some implementation issues of SIC are represented. Furthermore, possible extensions of NOMA are also discussed applying multi-antenna techniques to achieve further performance improvement in 5G networks. In addition, this chapter provides a detailed literature review related to different resource allocation techniques proposed for NOMA.

In Chapter 3, mathematical optimization which is the main approach for addressing the radio resource allocation problems is investigated. The basic concepts of convex optimization theory are outlined and then the most generic classes of convex problems, namely, linear programming (LP), quadratic programming (QP), quadratic constrained quadratic programming (QCQP), second-order cone programming (SOCP) and semidefinite programming (SDP) are described with necessary details.

In Chapter 4, the system model is introduced for the MISO-NOMA system with the assumption of perfect channel state information (CSI) at the transmitter. Two main radio resource allocation problems are investigated; sum-power minimization and max-min fairness. Since low energy consumption is one of the key requirements in future wireless networks, a beamforming design is first provided to minimize the total transmit power with the minimum rate requirement constraints at each user. To solve this optimization problem, Convex-Concave Procedure (CCP) is employed based on two different approaches; Taylor series approximation and semidefinite relaxation (SDR) to design the beamformers. The performance of proposed approaches are evaluated in terms of power consumption and computational complexity while comparing the performance with the conventional OMA scheme. Next, a max-min fairness problem is defined to provide fairness between users in which the minimum rate between all users is maximized while satisfying the transmit power constraint. This problem is solved by utilizing the bisection method to obtain the optimal solution and simulation results have been provided to demonstrate the effectiveness of this max-min beamforming design [74].

In the previous resource allocation problems, it is assumed that perfect CSI is available at the transmitter. However, in wireless transmissions, channel uncertainties are inevitable due to quantization and channel estimation errors, limited training sequences and feedback delays. In particular, due to ambiguities introduced in SIC through the user decoding order and SC at the transmitter in NOMA, these uncertain-

ties can greatly degrade the overall system performance. To cultivate the desirable benefits offered by NOMA, these channel uncertainties should be accounted for the design of resource allocation techniques; this is the main focus of Chapter 5. In particular, Chapter 5 introduces the concept of robust beamforming and the corresponding performance analyses are also presented. Two approaches are studied to incorporate the channel uncertainties; worst-case design which assumes the CSI errors belong to some known bounded uncertainty sets, and outage probabilistic based design where the channel errors are random with a certain statistical distribution and constraints can be satisfied with certain outage probabilities. By exploiting the SDR approach and S-procedure, the original non-convex constraints are converted into linear matrix inequality (LMI) form. Numerical results are provided to show that the proposed robust schemes outperform the non-robust scheme in terms of the achieved rates and rate satisfaction ratio at each user. In particular, they offer a better performance than the non-robust scheme by satisfying the signal-to-interference-noise ratio (SINR) requirement at each user all the time regardless of associated channel uncertainties [75, 76].

Due to the limited available spectrum and increasing throughput requirements, SE has been considered as the sole performance metric in the designs of resource allocation in the NOMA based systems. However, with the immense increase of the traffic data and number of mobile devices, the energy consumption has increased and become an important issue in the green cellular network. Accordingly, the designs of future wireless networks should be not only spectral efficient but also energy efficient. Hence, EE has become one of the key performance metrics in the development of 5G and beyond wireless networks. To this end, Chapter 6 investigates energy efficient resource allocation techniques for MISO NOMA system with imperfect CSI. Two approaches are introduced for NOMA systems: non-clustering and clustering schemes. In the first scheme, NOMA is employed to share the radio resources between all users and each user has its own beamforming vector. However, in the clustering scheme, the users in a cell are grouped into different clusters and the users in the same cluster are supported by the NOMA scheme. In this scheme, transmit beamforming vectors are generated in the same manner as in the conventional multi-user beamforming systems. Each of these transmit beamforming vectors can support a group of two or more users in the same cluster, rather than only a single user. To remove the interference between different clusters, two different types of zero-forcing (ZF) designs, namely, hybrid-ZF and full-ZF are employed at the BS. The full-ZF scheme completely removes the inter-cluster interference at the cost of more number of antennas at the transmitter whereas the hybrid-ZF scheme partially cancels the inter-cluster interference. Moreover, by

leveraging the norm-bounded channel uncertainty model, the worst-case EE is consider for both NOMA schemes while each user enjoys a minimum required QoS. To solve the problem, an iterative scheme is developed by exploiting Dinkelbach's algorithm. The simulation results reveal that hybrid-ZF outperforms the full-ZF scheme with a few clusters, while full-ZF shows a better performance with a higher number of clusters. In addition, the numerical results confirm that the proposed robust schemes outperform the non-robust scheme in terms of the rate satisfaction ratio at each user.

Finally, Chapter 7 draws conclusions of this thesis and identifies interesting future research directions.

Chapter 2

Fundamental Concepts and Literature Review

In this chapter, the fundamental transmission principles of the NOMA systems are presented. First, SC with SIC is discussed which are underlying notions behind this novel multiple access technique. Then, the downlink channel capacity is provided for NOMA system in comparison with that of the conventional OMA techniques. Finally, the literature on NOMA is reviewed which provides recent work to address the challenges in future wireless networks.

2.1 Non-orthogonal Multiple Access Technology Towards 5G and Beyond

With the complementation of 4G standards and the development of LTE networks, the interests of the wireless research community is moving towards 5G. The capabilities of 5G and beyond wireless networks should extend far beyond previous generations of wireless communications in order to enable massive connectivity with diverse and stringent requirements. These capabilities will include massive system capacity, high data rates at any time and anywhere, very low latency, ultra-high reliability, very low device cost and energy-efficient techniques. The current mobile networks will not be able to handle these demanding requirements which imposes the necessity of different disruptive technologies. Recently, NOMA has been proposed as a key technique to achieve high SE and considerable performance improvements in system throughput. In addition to the higher SE realised in NOMA, it is also flexible to implement with other proposed technologies for 5G and beyond networks which provides the additional fre-

quency reuse gain to further improve the performance of wireless communication systems. Unlike OMA that restricts the bandwidth assignment to the weaker users in order to achieve a higher throughput, all users can use the overall transmission bandwidth in non-orthogonal access regardless of their channel conditions. The fundamental concept of NOMA is sharing the same wireless resources between multiple-users in the code or power domain, resulting in non-orthogonality among user access. This scheme applies SC [54] which is a summation of all users'signal to superpose multiple user at the transmitter. By relying on the receivers with equipped processors with different capabilities, multi-user detection and SIC are applied to separate multi-user signal at the receiver end.

2.1.1 Superposition Coding with Successive Interference Cancellation

Power domain multiplexing means that multiple users'signals are multiplexed through different power levels based on their channel conditions while the same time-frequency-code resources are shared among multiple users. This scheme applies SC to encode multiple user's signals at the transmitter, and performs SIC at the receiver to decode multi-user signals. Since the SC scheme has the capability to achieve the channel capacity, it is an efficient technique to increase capacity in the NOMA system. This thesis has focused on a NOMA scheme in the power domain [46, 52, 77, 78], and hereafter, NOMA refers the power-domain NOMA scheme. In NOMA, it is expected that the weaker users are allocated with a higher transmit power than that of the strong users as they are already experiencing a poor channel conditions. In other words, a stronger user is allowed to access the slot which is used by a weaker user without degrading the performance of the weaker user [79]. As they share the same radio resources, the signal related to the weaker user received at the strong user has a higher SINR than that at the weaker user. Hence, the strong user is able to employ SIC to decode its own signal by first decoding and removing the weaker user's signal. On the other hand, the weaker user decodes its own signal without using SIC as the strong user's signal is negligible at the weaker user due to its transmit power is lower than that of the weak user.

Here, an example is provided to further exemplify the basic concepts of SC and SIC as shown in Fig. 2.1. Assume a downlink transmission scenario with two users where a BS serves them by using the same channel. The users are geographically distributed where one of them is the near user and the other one is the far user. The

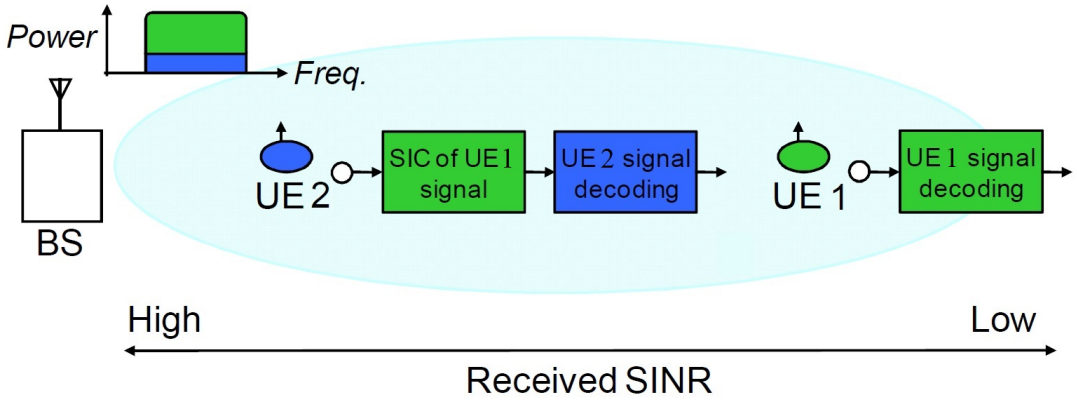


Fig. 2.1 SC and SIC receiver [52].

channel gains from the BS to two users are denoted by h_1 and h_2. The near user is considered as user 2, U_2, which has a better channel condition than that of the far user, i.e., user 1, U_1, hence, $\|h_1\| \leq \|h_2\|$. NOMA allocates more transmit power to the user with poor channel condition whereas the user with better channel condition is served with less transmit power. Consider x_1 and x_2 as signals intended for U_1 and U_2, respectively. At the transmitter, the BS broadcasts a superposition coded information to two users which is the sum of users'signals, i.e., $x = x_1 + x_2$. Hence, the received signals at U_1 and U_2 can be written by

$$y_1 = h_1 x + n_1, \tag{2.1}$$

$$y_2 = h_2 x + n_2, \tag{2.2}$$

where n_k is Additive White Gaussian Noise (AWGN) for $U_k,\ k = 1, 2$, with zero mean and variance σ_k^2.

For multi-user detection, the SIC is employed at the receivers. In SIC technique, by assuming that $\|h_1\| \leq \|h_2\|$, the signal which can be decoded at the weaker user, i.e., U_1, can be also decoded by the strong user, i.e., U_2 [80, 81]. Therefore, at the receiver end, U_2 can perform SIC and first decodes the signal intended to U_1, i.e., x_1. After subtracting x_1 from the received signal y_2, U_2 can decode its own signal, i.e., x_2 from $h_2 x_2 + n_2$. At U_1's receiver, x_2 is treated as noise and U_1 decodes its own signal, x_1, directly from y_1 [82] as illustrated in Fig. 2.2. This process can be extended for systems with more number of users. Assume the scenario with K users where x_k is supposed to be transmitted to $U_k,\ \forall k$. The users are sorted based on their channel strengths i.e., $\|\mathbf{h}_1\| \leq \|\mathbf{h}_2\| \leq \ldots \leq \|\mathbf{h}_K\|$. The BS transmits a superposition coded signal, i.e., $x = \sum_{k=1}^{K} x_k$, to all users. At the receiver end based on this user ordering,

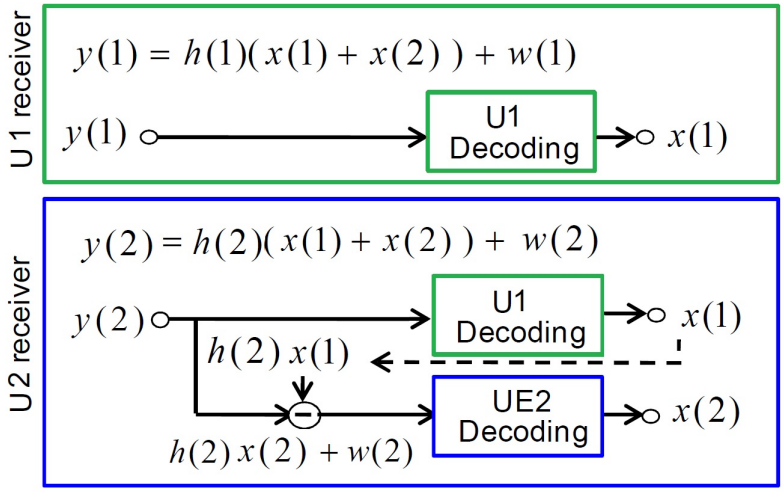


Fig. 2.2 The receivers for NOMA with two users scenario.

each user U_k can detect and remove the first $k-1$ users'signals in a successive manner whereas the messages of the other users, i.e., from U_{k+1} to U_K, are treated as noise. From an information-theoretic perspective, it is well-known that non-orthogonal user multiplexing using SC at the transmitter and SIC at the receiver not only outperforms orthogonal multiplexing, but also is optimal in the aspect of achieving the capacity region of the downlink broadcast channel [55, 82].

2.1.2 Downlink Channel Capacity

A channel in wireless communications is defined in terms of the bandwidth and time. The fundamental metric to measure the performance of a wireless channel is its capacity which is defined as the maximum data rate of communication over this wireless channel for which arbitrarily small error probability can be achieved at the receiver end [82]. The previous example of a two-user NOMA system is considered where U_2 with better channel condition performs SIC, whereas the U_1 only decodes its own data while U_2's signal is treated as noise. According to the Shannon's capacity definition, the data rate normalized to the bandwidth for U_1 and U_2 can be written by

$$R_1 = \log_2\left(1 + \frac{P_1|h_1|^2}{P_2|h_1|^2 + \sigma_1^2}\right), \tag{2.3}$$

$$R_2 = \log_2\left(1 + \frac{P_2|h_2|^2}{\sigma_2^2}\right), \tag{2.4}$$

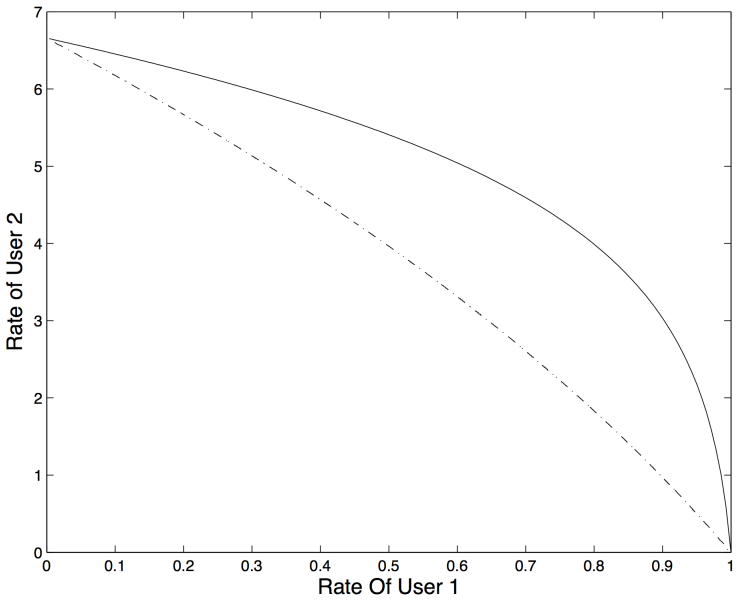


Fig. 2.3 The rate regions of the two-user downlink broadcast channel of NOMA and orthogonal schemes with the users' signal-to-noise ratio (SNR) equal to 0 and 20 dB (i.e., $\frac{P|h_1|^2}{\sigma_1^2} = 1$ and $\frac{P|h_2|^2}{\sigma_1^2} = 100$). In the orthogonal schemes, both the power split $P = P_1 + P_2$ and split in degrees of freedom α are jointly optimized to compute the boundary [82].

where P_1 and P_2 denote the transmit power for the U_1 and U_2, respectively, and $P = P_1 + P_2$ is the total transmit power at the BS.

On the other hand, consider an orthogonal scheme which allocates a fraction α of the degrees of freedom to U_1 and the rest $(1-\alpha)$ to U_2. Hence, orthogonal schemes can achieve following rates, for each power split $P = P_1 + P_2$ and degree-of-freedom split $\alpha \in [0\ 1]$:

$$R_1 = \alpha \log_2 \left(1 + \frac{P_1|h_1|^2}{\alpha\sigma_1^2}\right), \tag{2.5}$$

$$R_2 = (1-\alpha) \log_2 \left(1 + \frac{P_2|h_2|^2}{(1-\alpha)\sigma_2^2}\right). \tag{2.6}$$

Here, α represents the fraction of the bandwidth devoted to users. In the multi-user systems, as the channel should be shared or divided among different users, the rate region is utilized to characterize the multi-user channel capacity. This region demonstrates the tradeoff between the rates of the concurrent streams in the multi-user channel [83]. To compare the performance of NOMA and orthogonal schemes, the rate region can be obtained for both schemes. The achievable boundaries of the rate regions with

NOMA and the orthogonal scheme for the downlink channel is shown in Fig.2.3 with $\text{SNR}_1 = 0$ dB and $\text{SNR}_2 = 20$ dB. It is observed that NOMA has a better performance than orthogonal scheme (except for the two corner points where only one user is being communicated to). For any rate pair achieved by orthogonal scheme there is a power split for which the NOMA scheme achieves the rate pairs that are larger than that of orthogonal scheme. This performance gap is more pronounced when the difference between the users'channel gain increases. In Fig.2.3, for example, while maintaining the rate of the weaker user R_1 at 0.9 bps/Hz, NOMA can provide a rate of around $R_2 = 3$ bps/Hz to the strong user while an orthogonal scheme can provide a rate of only around 1 bps/Hz. Intuitively, the strong user, being at high SNR, is degree of freedom limited and NOMA allows it to use the full degrees of freedom of the channel while being allocated only a small amount of transmit power, thus causing small amount of interference to the weak user. In contrast, an orthogonal scheme has to allocate a significant fraction of the degrees of freedom to the weak user to achieve near single user performance, and this causes a large degradation in the performance of the strong user. As result, by allowing to send the information to two users over the same frequency band and time by NOMA, the channel capacity boundary shown in Fig. 2.3 can be potentially reached. Different from orthogonal scheme where channel gain difference is expressed as a multi-user diversity gain via frequency-domain scheduling, the channel gain difference in NOMA is expressed as multiplexing gains by superposing the signals of multiple users in the power-domain. In NOMA, both users with high and low channel gains are in a win-win setup [46]. In other words, the users with better channel gain lose a little by being allocated less transmit power, but gain more by being allocated more bandwidth, while weaker users also lose a little because of the interference from the strong users but gain more by using more bandwidth [52, 77, 78]. Another benefit of SIC based receiver is that by allocating more transmit power to weak users, they can get a much higher data rate than that of the conventional schemes. This means that users with better channel conditions can be allowed to take advantage of the strong channel and be transmitted with a higher rate while not degrading the performance of the cell-edge users by controlling the transmit power. In the conventional receiver, this is not possible due to the near-far problem. With the OMA techniques, the good performance for the weak user can be attained by nearly sacrificing all the rate of the stronger user. With the SIC, the near-far problem is turned into a near-far advantage be realizing better rates for both users.

These results have natural extensions to the general K user downlink channel. In general with the ordering $\|\mathbf{h}_1\| \leq \|\mathbf{h}_2\| \leq \ldots \leq \|\mathbf{h}_K\|$, the data rate for U_k is achieved

by

$$R_k = \log_2 \left(1 + \frac{p_k |h_k|^2}{\sum_{i=k+1}^{K} p_i |h_k|^2 + \sigma_k^2} \right), \quad k = 1, 2, \ldots, K. \tag{2.7}$$

where p_k denotes the transmit power for the U_k, $\forall k$, and $P = \sum_{k=1}^{K} p_k$ is the total transmit power at the BS. The optimal points are achieved by SC at the transmitter and SIC at each of the receivers. The interference cancellation order at every receiver is always to decode the weaker users before decoding its own signal. By applying SIC technique at the receivers, multiple users with different transmit power levels can be multiplexed on the same frequency band, providing higher sum rate than that of conventional orthogonal schemes.

2.1.3 Implementation Issues of SIC

In the previous section, the advantages of SIC technique have been discussed which provides significant performance gain over the conventional OMA techniques. It takes advantage of the strong channel of the nearby user to achieve higher throughput while providing the weak user with the best possible performance. However, there are some potential practical issues in iterative SIC such as the complexity, error propagation, analog-to-digital (A/D) quantization error and imperfect CSI due to estimation errors [82]:

- In the downlink, by employing SIC at the receivers, the stronger users have to decode information intended for the weaker users. It means that the complexity of SIC scales with the number of users which shares the same bandwidth. However, the SC with SIC has the largest performance gain when the users have different quality of channels from the BS. This can be a practical way to decrease the complexity by grouping the users in a cell into different clusters with a small number of users. The users in each cluster are supported by the SC-SIC scheme to share the same time-frequency block. Hence, clustering approach can significantly contribute to the performance gain of the overall system with low complex SIC.

- In capacity analysis in section 2.1.2, the error-free decoding is assumed, however, errors are inevitable with practical codes. Since the decoding in the SIC is sequential, if an error occurs for one user, it is more likely that the next users in

the SIC decoding order are decoded incorrectly. In other words, the error propagates in the successive decoding. It is shown if ρ_i is the probability of decoding the U_i incorrectly, assuming that all the previous users are decoded correctly, then the error probability for the U_k using SIC is $\sum_{i=1}^{k} \rho_i$ [82]. Hence, the error propagation degrades the error probability by a factor of the number of users K when all users are coded with the same target error probability. Hence, If number of users is reasonably small, this effect can easily be compensated by using a slightly stronger code.

- When the range of received powers is very large, an A/D converter with a large dynamic range is essential. For example, if the power disparity is 20dB, even 1-bit accuracy for the weak signal would require an 8-bit A/D converter. This may well pose an implementation constraint on how much gain SIC can offer [82].

- To apply SIC and to remove the signals of weaker users from the received signal, they must be reconstructed from the decoded signal and also the impulse response of the channel. Imperfect estimate of the channel will lead to residual cancellation errors. One concern is that, if the range of the users'received power is very large, the residual error from cancelling the weaker user can degrade the strong users signal. On the other hand, it is also much easier to accurately estimate the channel of strong users. Therefore, these two effects compensate each other and the effect of residual errors does not grow with the power disparity [82].

2.2 NOMA with Multiple Antennas

As mentioned in Chapter 1, the requirements in future wireless networks are very demanding and stringent. In particular, to achieve a further enhancement in the SE compared to the LTE baseline, it is required to introduce new technologies. The gain from NOMA using SIC is an improvement of approximately $30 - 40\%$ in the SE [84]. Although the same time-frequency resources can be shared by multiple users in the basic NOMA employing SIC, the improvement of SE still remains limited. One solution to address this problem is the extension of the NOMA by incorporating it with multiple-antennas techniques [85, 86]. Hence, in this section, possible extensions of NOMA are discussed by applying multi-antenna techniques through which multiple degrees

of freedom can be introduced to achieve further performance improvement in 5G and beyond wireless networks. Despite of potential benefits by applying multiple antennas in NOMA, numerous open research challenges need to be addressed. In contrast to the single-input single-output (SISO) NOMA, design of MIMO NOMA is more challenging. The main problem in SISO-NOMA is determining optimal power allocation between users, while by the extra spatial degrees of freedom in MIMO-NOMA, it is possible to cancel the user interference via beamforming in both the power and spatial domains. This makes the design of MIMO-NOMA more complicated. In general, the introduction of multi-antenna NOMA brings up two major challenges. First, although it is obvious that MIMO-NOMA outperforms MIMO-OMA [87], it is not clear that SC with a SIC in MIMO can achieve the optimal performance in term of multi-user capacity region. It is shown in [88] that if the users'channels are quasi-degraded, the use of MIMO-NOMA can achieve the optimal performance. However, the extension from quasi-degraded channels to general ones remains as an open problem. Secondly, user ordering in MIMO-NOMA scenarios is a NP hard problem as they are determined through the design parameters that in the form of matrices or vectors. In the SISO case, since the users'channels are scalars, the SIC order usually depends directly on channel gains. It is possible for the receiver to decode a signal intended to the weaker user that has a lower channel gain, and hence the optimal decoding order is in the order of the increasing/decreasing channel gain. However, this property is not guaranteed in multiple antenna systems where separate beamformer is designed for each user [89]. In MIMO-NOMA, the effective channel gains are related to particular beamforming designs, which makes the design of beamforming and SIC ordering coupled [90]. Different techniques are proposed for ordering in MIMO NOMA scenarios. The random beamforming is considered in [85] where the BS orders the users according to quality of feedback channels. Another proposed technique is the large scale path loss based user ordering [91, 79, 92] where user located far from the BS is treated as the weak user and its signal is decoded first at all the receivers. Another effective way is to pair users into clusters, and assign the same beamforming vector to the users in the same cluster. This decomposes the MIMO-NOMA channels into multiple parallel SISO-NOMA subchannels [87, 93].

In this thesis, the SC with a SIC receiver is employed for two different categories of MISO-NOMA scheme: the beamformer-based scheme [79] and the cluster-based scheme [77, 94, 95]. In the beamformer-based scheme, there is no clustering and user ordering is based on the large scale path loss. Then, SC based SIC is applied between all users served by the BS where each user is supported by its own beamforming vector.

However, in the cluster-based scheme, the users are grouped into different clusters with at least two users in each cluster. Then, a transmit beamforming vector is designed to support each cluster through conventional multi-user beamforming designs. The users in each cluster are supported by a NOMA based SC-SIC approach where multiple user signals are superposition coded within each beam. The key difference between these two schemes is whether one beamformer serves one user or multiple users. It is worth to note here that the ordering which used in this thesis may not be optimal, and better rates may be achievable for different decoding order. However, the work in this thesis does not focus on the optimal decoding order problem, but in the radio resource allocation and design of the beamforming vectors to achieve the best performance by taking realistic constraints into account.

2.3 Literature Review

The NOMA scheme has recently received considerable attention in research community due to its potential benefits in 5G and beyond networks. Although OFDMA is widely adopted as the multiple access technique in current wireless communication systems by providing concurrent interference-free transmission to multiple users, it does not generally achieve the highest possible rate [96–99]. Hence, NOMA has become as one of the key enabling techniques for the design of radio access techniques for the 5G wireless networks to address the future wireless network challenges [24, 100]. The performance of NOMA scheme is investigated in [101–105] where the BS and users are equipped with a single antenna. In [101], the NOMA scheme is studied for downlink transmission in a cellular system with randomly deployed users whereas the design of uplink NOMA schemes has been proposed in [102]. The impact of user pairing on a hybrid multiple access system is studied in [103] where NOMA is combined with conventional multiple access. In [104], a novel cooperative NOMA scheme has been proposed with the derivations of the outage probability and diversity order. Joint power allocation and relay beamforming design is investigated in [105] for a NOMA based amplify-and-forward relay network where the achievable rate of the destination with the best channel condition is maximized with rate requirements at other destinations and individual transmit power constraints.

A general framework for a MIMO NOMA system has been studied in [93] for both downlink and uplink transmission to further improve the performance in which the inter-cluster interference is eliminated by exploiting signal alignment. In [106],

a design for precoding and detection has been developed for MIMO-NOMA by deriving the outage probabilities for two different power allocation schemes whereas the MIMO-NOMA network with limited feedback is considered in [107]. The sum rate maximization problem for a MISO NOMA has been investigated in [79] through the minorization maximization algorithm. The optimal and low complexity power allocation schemes have been proposed in [91] for a two-user MIMO-NOMA system. Authors in [52, 85] provided system-level performance comparisons between NOMA and OFDMA based cellular downlink transmissions, where [52] considered single-antenna transceivers, and [85] extended performance comparison to MIMO technology with random beamforming. They showed that NOMA improves both the overall cell throughput and cell-edge user throughput performances over OFDMA. It is shown that the performance of NOMA without MIMO is similar to that of OFDMA with 2×2 random beamforming [85], which suggested that NOMA has a similar effect to spatial multiplexing with random beamforming. For the MIMO NOMA system, the secrecy rate maximization problem is solved in [108] where it is demonstrated that the NOMA scheme outperforms the conventional OMA scheme in terms of achieved sum secrecy rate by efficiently utilizing available bandwidth. Joint optimization with beamforming design and power allocation with clustering in MISO-NOMA systems is considered in [109] where an iterative algorithm is proposed based on SDR to minimize power, showing that this algorithm requires less power than for power allocation and beamforming considered separately. In [110], a secure beamforming design is proposed for a MISO NOMA system by grouping the users into clusters.

In wireless transmissions, channel uncertainties are inevitable due to quantization and channel estimation errors, limited training sequences and feedback delays. To circumvent the inevitable channel uncertainties, robust design is a well-known approach in the literature, which can be classified into two groups, the worst-case robust design [111–114], and outage probability based design [115, 116]. In the worst-case design, it is assumed that the CSI errors belong to some known bounded uncertainty sets and robust beamforming design is proposed to tackle the worst error whereas in the second approach, the channel errors are random with a certain statistical distribution and constraints can be satisfied with certain outage probabilities. In [111], the robust beamforming design has been developed for providing secure communication in wireless networks with imperfect CSI. By incorporating the bounded channel uncertainties, the robust sum power minimization problem is investigated in [112] for a downlink multicell network with the worst-case SINR constraints whereas the robust weighted sum-rate maximization was studied for multicell downlink MISO systems in

[113]. In [114], a robust minimum mean square error based beamforming technique is proposed for multi-antenna relay channels with imperfect CSI between the relay and the users. An outage-constrained distributed robust beamforming scheme is developed for multicell networks using second order statistical CSI in [115]. Authors in [116] consider robust multi-cell coordinated beamforming design for downlink wireless systems by assuming that the CSI errors are complex Gaussian distributed and formulate a chance-constrained robust design problem which guarantees that the mobile stations can achieve the desired SINR requirements with a high probability. In most existing NOMA schemes, it is assumed that perfect CSI is available at the transmitter, however, it is not a practical assumption. In particular, due to ambiguities introduced in SIC through user decoding order and SC at the transmitter in NOMA, these uncertainties can greatly degrade the overall system performance [117]. Therefore, to cultivate the desirable benefits offered by NOMA, these channel uncertainties should be accounted for in the design of resource allocation techniques. In the context of NOMA, a robust design with the norm-bounded channel uncertainties is investigated in [95] where a clustering scheme is studied to maximize the worst-case achievable sum rate with a total transmit power constraint.

Besides, all of the above work in the literature on NOMA mainly focus on improving the SE of NOMA systems [79, 118–120]. However, there is a dearth of literature considering the EE which has been identified as one of the key performance metrics in future wireless networks. The EE for NOMA systems was investigated in [121] for a given statistical CSI at the transmitter. A crucial step forward followed in [122] to maximize the EE of downlink NOMA systems by recalling a non-linear fractional programming method. Also, the authors in [123] proposed a power allocation and subchannel assignment to maximize the EE in NOMA networks by assigning only two users per subchannel. The joint user scheduling and power allocation in this context was further explored in [124, 125] under the assumption of imperfect CSI. In [124], it was assumed that only two users can be multiplexed on each subchannel whereas a general case with more number of users on same subchannel was developed in [125]. Their results confirmed that the NOMA system can achieve a better performance in terms of sum rate and EE compared to the conventional OMA systems. In [126], the authors proposed two user scheduling schemes combined with a power allocation scheme to enhance the EE in the MIMO NOMA system.

Radio resource allocation techniques play a key role in performance improvement in cellular networks. The goal of resource allocation is to optimize the assignment of the limited resources to achieve the best performance by satisfying the required

constraints. The benefits of optimization in cellular networks include network performance improvement, satisfying QoS requirements, saving energy, as well as reducing the operation expenditure. Therefore, it is important to investigate the various resource allocation problems in NOMA transmission to analyse and validate its performance through better utilization of available radio resources. Since most research works on NOMA systems focuses on the case that the BS has the perfect knowledge of the CSI and the SE design, this thesis has focused on the robust SE design as well as robust energy efficient one. Accordingly, in this thesis, it is aimed to find optimal solutions, or develop near-optimal solutions for different resource allocation problems in the context of MISO NOMA which are not addressed in the literature. In particular, it is focused on three different resource allocation problems i.e., beamforming design with perfect CSI, robust beamforming design, and robust energy-efficient beamforming design.

2.4 Summary

In this chapter, an overview on fundamental concepts of NOMA systems has been provided. Furthermore, the concept of SC with SIC is presented and some implementation issues of SIC have been discussed. Then, the downlink channel capacity for NOMA system has been defined while comparing the rate region of the NOMA system with the OMA system. Moreover, the extension of the NOMA to incorporate multiple-antennas techniques has been introduced to enable further performance improvement in 5G and beyond wireless networks. Finally, a detailed literature review has been provided related to different resource allocation techniques available for NOMA.

Chapter 3

Mathematical Optimization

Optimization techniques play a vital role in the deployment and operation of most types of networks by providing a powerful set of tools for the design and analysis of communication systems and signal processing algorithms [127–129]. Many communication problems can be appropriately formulated into different optimization problems with a set of constraints. These problems are either convex in nature or can be converted into convex forms through some mathematical manipulations and approximations. All standard convex optimization problems, can be efficiently solved using interior point methods [130, 131]. Furthermore, the algorithms and software tools for convex optimization problems have advanced significantly over the last decades which enables solving many engineering problems through convex optimization framework. However, most optimization problems are non-convex in nature and formulating those problems into a convex forms is the major challenge in the application of convex optimization. In this section, first, the general structure of radio resource allocation problems is briefly presented. Then, different kind of convex optimization problems are reviewed.

3.1 Radio Resource Optimization in Wireless Networks

Radio resource allocation has become increasingly important in wireless systems and networks design as it enables an efficient use of radio resources, offering additional benefits under the same infrastructure [132–134]. In general, optimizing the assignment of the limited resources (i.e., frequency, power or time) to achieve the best performance with a given set of constraints is the main aim of resource allocation techniques. There are typically three main components in a radio resource allocation problem: 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 physical limitations in wireless networks. Any optimization problem can be feasible or infeasible according to the feasible solution region defined based on the constraints and optimization variables. In other words, resource allocation aims to find the optimal solutions from the feasible region, or develop near-optimal solutions [135, 136]. Some classic objective functions used in this thesis are summarized as follows:

- Power minimization: Power consumption is an important performance metric and imposes demanding requirements in future green wireless networks. A typical objective function is the sum of the transmit power allocated for the users served in the network as $\sum_k P_k$ where P_k is the transmit power allocated for the k^{th} user U_k.

- Max-min fairness: In some application scenarios, maintaining fairness between users in terms of achieved throughput is more important than maximum throughput. Hence, the corresponding utility function is expressed as the minimum achievable rate between all users as $\min_k R_k$ where R_k is the data rate of U_k.

- Spectrum efficiency: This utility function is used to quantify throughput in unit bandwidth and expressed in bits per second per Hz (bps/Hz). In this class of optimization, the objective is the sum of the users' data rate (sum-rate), i.e., $\sum_k R_k$, normalized to the bandwidth where the data rate is usually computed by well-known Shannon's channel capacity definition in bits per second (bps).

- Energy efficiency: To strike a good balance between the achievable data rate and power consumption, EE is another performance metric defined as the number of bits that can be reliably transmitted per Joule of energy consumption. The function is expressed by $\frac{\sum_k R_k}{\sum_k P_k}$ in bit per Joule.

In the following sections, an overview of mathematical optimization is provided focusing on the special role of various convex optimization techniques. Furthermore, a general optimization problem is formulated and the different canonical optimization problems are discussed at the end of this chapter.

3.2 Fundamentals of Convex Optimization

This section briefly introduces the fundamentals of convex optimization.

3.2.1 Convex Sets

A set $C \in \mathbb{R}^n$ is convex if for any $\mathbf{x}, \mathbf{y} \in C$ and any $\theta \in [0, 1]$, it holds [127]:

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C, \quad \forall \mathbf{x}, \mathbf{y} \in C. \tag{3.1}$$

In other words, a set can be classified as a convex set if the line segment between any two points from set lies in the same set.

3.2.2 Convex Cones

A set C is called a convex cone, if it is convex and for every $\mathbf{x} \in C$ and $\theta \geq 0$, it holds $\theta\mathbf{x} \in C$. This can be mathematically expressed as [127]:

$$\theta_1\mathbf{x} + \theta_2\mathbf{y} \in C, \quad \forall \mathbf{x}, \mathbf{y} \in C, \;\; \theta_1, \theta_1 \geq 0. \tag{3.2}$$

Points of this form can be described geometrically as forming the two-dimensional pie slice with apex 0 and edges passing through $\mathbf{x}$ and $\mathbf{y}$. Convex cones arise in the various forms in engineering applications. The most common convex cones which are used in this thesis are second-order cone (SOC) $C = \{ (t, \mathbf{x}), \; t \geq \| \mathbf{x} \| \}$, and positive semidefinite matrix cone $C = \{\mathbf{X} | \mathbf{X} \text{ symmetric and } \mathbf{X} \succeq \mathbf{0}\}$.

3.2.3 Convex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if **dom** f is a convex set and if for all $\mathbf{x}, \mathbf{y} \in$ **dom**f, it holds

$$f\big(\theta\mathbf{x} + (1-\theta)\mathbf{y}\big) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}). \tag{3.3}$$

In other words, the line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ lies above the graph of f. The function f is concave if $-f$ is convex. A convex function is continuous on the relative interior of its domain and it can have discontinuities only on its relative boundary.

Suppose f is differentiable, then f is convex if for all $\mathbf{x}, \mathbf{y} \in$ **dom**f the following inequality holds

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}). \tag{3.4}$$

Moreover, if f is twice continuously differentiable, then the convexity of f is equivalent to

$$\nabla^2 f(\mathbf{x}) \succeq 0, \quad \forall \mathbf{x}. \tag{3.5}$$

It means f is convex if and only if its Hessian is positive semidefinite on its domain [127]. This reduces to the simple condition $f(\mathbf{x})'' \geq 0$ for a function on $\mathbb{R}$ which means that the derivative is non-decreasing. Similarly, f is concave if and only if $\nabla^2 f(\mathbf{x}) \preceq 0$, or the simple condition $f(\mathbf{x})'' \leq 0$ for a function on $\mathbb{R}$.

Quasiconvex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex if its domain and all its sublevel sets $S_\alpha = \{\mathbf{x} \in \mathbf{dom} f \mid f(\mathbf{x}) \leq \alpha\}$ for $\alpha \in \mathbb{R}$, are convex. A function f is quasiconcave if $-f$ is quasiconvex, it means every superlevel set $\{\mathbf{x}|f(\mathbf{x}) \geq \alpha\}$ is convex. It is shown that the quasiconvexity is a generalization of convexity and quasiconvex functions hold many of the properties of convex functions [127]. There is a variation on Jensen's inequality that characterizes quasiconvexity: A function f is quasiconvex if and only if $\mathbf{dom} f$ is convex and for any $\mathbf{x}, \mathbf{y} \in \mathbf{dom} f$ and $\theta \in [0\ 1]$,

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}. \tag{3.6}$$

This inequality is sometimes called Jensen's inequality for quasiconvex functions.

3.3 Convex optimization Problems

3.3.1 Optimization problems

A general mathematical formulation to describe the problem of finding a solution $\mathbf{x}$ that minimizes $f_0(\mathbf{x})$ while satisfying the set of conditions can be defined in the following standard form:

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p, \\ &&& h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, q, \end{aligned} \tag{3.7}$$

where the vector $\mathbf{x} = (x_1, \ldots, x_n)$ is the optimization variable of the problem, the function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective function or cost function, the functions f_i are the inequality constraints, and the constants h_j are the equality constraints. If there are no constraints, then the problem can be classified as an unconstrained problem. The domain of this optimization problem is the set of points for which the objective and the constraints are defined and is denoted as

$$D = (\bigcap_{i=1}^{p} \mathbf{dom}\ f_i) \bigcap (\bigcap_{j=1}^{q} \mathbf{dom}\ h_j). \tag{3.8}$$

A point $\mathbf{x} \in D$ is feasible, if it satisfies all the constraints f_i and h_j and problem (3.7) is said to be feasible if there exists a feasible point, otherwise it is called infeasible. The solution of the optimization problem is achieved at the optimal point $\mathbf{x}^*$ if and only if

$$f_0(\mathbf{x}^*) \leq f_0(\mathbf{x}), \quad \forall \mathbf{x} \in D. \tag{3.9}$$

It means if vector $\mathbf{x}^*$ has the smallest objective value among all vectors that satisfy the constraints, it is called optimal or a solution of the problem (3.7) [127]. Different classes of optimization problems are generally characterized based on the forms of the variables, objective and constraint functions.

Maximization problems

Some problems are defined in the form of the maximization problem as

$$\begin{aligned} &\text{maximize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) \leq 0, \ \ i = 1, \ldots, p, \\ &&& h_j(\mathbf{x}) = 0, \ \ j = 1, \ldots, q. \end{aligned} \tag{3.10}$$

This problem can be solved by minimizing the function $-f_0$ subject to the constraints. Hence, all the terms defined for a minimization problem can also be defined for the maximization problem. The optimal value of the problem in (3.10) is defined as $sup\{f_0(\mathbf{x}) | f_i(\mathbf{x}) \leq 0, \forall i, \ h_j(\mathbf{x}) = 0, \forall j\}$. The solution of the optimization problem is achieved at the optimal point $\mathbf{x}^*$ if and only if $f_0(\mathbf{x}^*) \geq f_0(\mathbf{x})$ for all $\mathbf{x}$ in the domain of this problem. For the maximization problem, the objective is also called the utility instead of the cost.

Equivalent problems

Two problems are equivalent if a solution of one can be easily found from a solution of the other. Consider the following problem as an example

$$\begin{aligned} &\text{minimize} && \tilde{f}(\mathbf{x}) = \alpha_0 f_0(\mathbf{x}) \\ &\text{subject to} && \tilde{f}_i(\mathbf{x}) = \alpha_i f_i(\mathbf{x}) \leq 0, \;\; i = 1, \ldots, p, \\ &&& \tilde{h}_j(\mathbf{x}) = \beta_j h_j(\mathbf{x}) = 0, \;\; j = 1, \ldots, q, \end{aligned} \tag{3.11}$$

where $\alpha_i > 0, \forall i$, and $\beta_j \neq 0, \forall j$. It is obtained from the standard form problem (3.7) by scaling the objective and inequality constraint functions by positive constants, and scaling the equality constraint functions by non-zero constants. The feasible sets of the original problem and equivalent problem are identical. A point $\mathbf{x}^*$ is optimal for the original problem if and only if it is optimal for the scaled problem.

In the following, some general transformations to obtain the equivalent problems are described.

- **Change of variables:** Suppose $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is one-to-one, with image covering the domain of original problem D, i.e., $\psi(\mathbf{dom}\psi) \supseteq D$. The following problem with variable z can be defined as

$$\begin{aligned} &\text{minimize} && \tilde{f}_0(z) \\ &\text{subject to} && \tilde{f}_i(z) \leq 0, \;\; i = 1, \ldots, p, \\ &&& \tilde{h}_j(z) = 0, \;\; j = 1, \ldots, q, \end{aligned} \tag{3.12}$$

  where the functions $\tilde{f}_i(z) = f_i\big(\psi(z)\big), \forall i$ and $\tilde{h}_j = h_j\big(\psi(z)\big), \forall j$. The original problem in (3.7) and the problem in (3.12) are related by the change of variable or substitution of variable $\mathbf{x} = \psi(z)$ [127]. These two problems are equivalent, hence, if z solves the problem in (3.12), then $\mathbf{x} = \psi(z)$ solves the problem in (3.7) and also if $\mathbf{x}$ solves the problem in (3.7), then $z = \psi^{-1}(\mathbf{x})$ solves the problem in (3.12).

- **Objective and constraint functions transformation:** Suppose $\phi_0 : \mathbb{R} \rightarrow \mathbb{R}$ is monotone increasing, $\phi_1, \ldots, \phi_p : \mathbb{R} \rightarrow \mathbb{R}$ satisfy $\phi_i(u) \leq 0$ if and only if $u \leq 0$, and $\phi_{p+1}, \ldots, \phi_{p+q} : \mathbb{R} \rightarrow \mathbb{R}$ satisfy $\phi_j(u) = 0$ if and only if $u = 0$.

The equivalent problem can be obtained as

$$\begin{aligned} &\text{minimize} && \tilde{f}_0(\mathbf{x}) \\ &\text{subject to} && \tilde{f}_i(\mathbf{x}) \leq 0, \;\; i = 1, \ldots, p, \\ &&& \tilde{h}_j(\mathbf{x}) = 0, \;\; j = 1, \ldots, q, \end{aligned} \tag{3.13}$$

where $\tilde{f}_i(\mathbf{x}) = \phi_i\big(f_i(\mathbf{x})\big), i = 0, \ldots, p$, and $\tilde{h}_j(\mathbf{x}) = \phi_{p+j}\big(h_j(\mathbf{x})\big), j = 1, \ldots, q$, are defined as the compositions. The problem in (3.13) and the original problem in (3.7) are equivalent, hence, the feasible sets and also the optimal points are identical.

- **Slack variables:** By using the simple transformation that $f_i(\mathbf{x}) \leq 0$ if and only if there is an $s_i \geq 0$ that satisfies $f_i(\mathbf{x}) + s_i = 0$, the following problem can be defined

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && s_i \geq 0, \;\; i = 1, \ldots, p, \\ &&& f_i(\mathbf{x}) + s_i = 0, \;\; i = 1, \ldots, p, \\ &&& h_j(\mathbf{x}) = 0, \;\; j = 1, \ldots, q, \end{aligned} \tag{3.14}$$

where the variables are $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{s} \in \mathbb{R}^p$. The new variable s_i is called the slack variable associated with the original inequality constraint $f_i(\mathbf{x}) \leq 0$. The problem in (3.14) is equivalent to the original problem in (3.7), hence, if $(\mathbf{x}, \mathbf{s})$ is feasible for the problem in (3.14), then $\mathbf{x}$ is feasible for the original problem in (3.7) and conversely, if $\mathbf{x}$ is feasible for the original problem, then $(\mathbf{x}, \mathbf{s})$ is feasible for the problem in (3.14). Moreover, $\mathbf{x}$ is optimal for the original problem in (3.7) if and only if $(\mathbf{x}, \mathbf{s})$ is optimal for the problem in (3.14).

3.3.2 Convex Optimization

A convex optimization problem is used for the general optimization problem in (3.7) when the objective function, f_0, and inequality constraint functions, $f_i, \; \forall i$, are convex, and the equality constraint functions $h_j, \; \forall j$, are affine [127]. Recall that a function is affine if it is a sum of a linear function and a constant, i.e., $\mathbf{a}_j^T\mathbf{x} = b_j, \; \forall j$. In convex optimization problems, any locally optimal point is also globally optimal. In general there is no analytical formulation for the solution of convex optimization prob-

lems, however, they can be solved efficiently by some well-known algorithms such as interior-point algorithm.

3.3.3 Linear optimization problems

The optimization problem is called a LP when the objective and constraint functions are all affine. The problem is non-linear program (NLP) if the objective or some constraint functions are non-linear. A standard form of LP problem can be expressed as follows

$$\begin{aligned} \text{minimize} \;\; & \mathbf{c}^T\mathbf{x} + d \\ \text{subject to} \;\; & \mathbf{G}\mathbf{x} \leq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.15}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{h} \in \mathbb{R}^p$, and $\mathbf{b} \in \mathbb{R}^q$ are column vectors and $\mathbf{G} \in \mathbb{R}^{p\times n}$ and $\mathbf{A} \in \mathbb{R}^{q\times n}$ are matrices. LPs are convex optimization problems. It is common to eliminate the constant d in the objective function, since it does not affect the optimal solution.

3.3.4 Quadratic optimization problems

The convex optimization problem in (3.7) is called a QP when the objective function is quadratic and the constraint functions are affine. The standard form of QP problem is shown below

$$\begin{aligned} \text{minimize} \;\; & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r \\ \text{subject to} \;\; & \mathbf{G}\mathbf{x} \leq \mathbf{h}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.16}$$

where $\mathbf{P} \in \mathbb{S}^n_+$, $\mathbf{G} \in \mathbb{R}^{p\times n}$ and $\mathbf{A} \in \mathbb{R}^{q\times n}$. When both objective and the inequality constraint functions are quadratic, the convex optimization problem is called a QCQP. This has the form

$$\begin{aligned} \text{minimize} \;\; & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}_0^T\mathbf{x} + r_0 \\ \text{subject to} \;\; & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} + \mathbf{q}_i^T\mathbf{x} + r_i \leq 0, \;\; i = 1, \ldots, p, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.17}$$

where $\mathbf{P}_i \in \mathbb{S}^n_+,\ \forall i$. QPs include LPs as a special case, by taking $\mathbf{P} = \mathbf{0}$ in (3.16). Also, QCQP includes QP (and therefore also LPs) as a special case, by taking $\mathbf{P}_i = \mathbf{0},\ i \geq 1$ in (3.17).

3.3.5 Second-Order Cone Programming

A SOCP can be written as [127, 137]

$$\begin{aligned} &\text{minimize } \mathbf{f}^T\mathbf{x} \\ &\text{subject to } \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\| \leq \mathbf{c}_i^T\mathbf{x} + d_i,\ \ i = 1, \ldots, p, \\ &\qquad\qquad F\mathbf{x} = \mathbf{g}, \end{aligned} \tag{3.18}$$

where $\mathbf{A}_i \in \mathbb{R}^{n_i \times n}$ and $\mathbf{F} \in \mathbb{R}^{q \times n}$. The first constraint in (3.18) is called a second order cone constraint. This class of optimization problems is closely related to QP. If $\mathbf{c}_i = \mathbf{0},\ \forall i$, a standard form of QCQP will be obtained by squaring both sides of the constraints. Similarly, if $\mathbf{A}_i = \mathbf{0},\ \forall i$, then the SOCP reduces to a LP.

3.3.6 Semidefinite Programming

The most general form for SDP can be written as [127, 138]

$$\begin{aligned} &\text{minimize } \mathbf{c}^T\mathbf{x} \\ &\text{subject to } \mathbf{x}_1F_1 + \mathbf{x}_2F_2 + \ldots + \mathbf{x}_nF_n + G \preceq \mathbf{0}, \\ &\qquad\qquad \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{3.19}$$

where $\mathbf{G}, \mathbf{F}_l, \ldots, \mathbf{F}_n \in \mathbb{S}^k$ and $\mathbf{A} \in \mathbb{R}^{q \times n}$. The inequality constraints in (3.19) are also called LMI. An SDP reduces to LP if the matrices $\mathbf{G}, \mathbf{F}_l, \ldots, \mathbf{F}_n$ are all diagonal.

3.4 Summary

In this chapter, a general introduction to the optimization problems is provided along with the related technologies and mathematical tools and the various convex optimization problems have also been briefly discussed. In general, the aim of optimization problem is finding optimal solutions from the feasible region, or developing near-optimal solutions. Convex problems are usually solved efficiently using interior point methods, but some original problems may not have a convex form due to inappropriate

formulations. However, for some of them it is possible to transform them to convex problems. For the addressed radio resource optimization problems in this thesis, it is a difficult and challenging task to solve the problems. Hence, some widely used mathematical tools are presented in this chapter which is utilized in the next chapters to solve the addressed optimization problems in NOMA-based systems.

Chapter 4

Resource Allocation Techniques for MISO NOMA Systems

4.1 Introduction

NOMA introduces a paradigm shift to offer high SE compared to that of the conventional multiple access schemes. Since available resources are limited, an efficient resource utilisation is required in the system design to meet the demanding requirements. Resource allocation is a series of processes required to determine the amount of related resources to be efficiently allocated to each user in order to achieve the best performance by taking realistic constraints into account [139]. Therefore, optimum power allocation among users in NOMA is required to provide the best performance with the minimum amount of resources. Early works on NOMA system mainly focus on the resource allocation to improve the performance of the wireless networks with single antenna transceiver [101–105]. Recently, some attempts are made to utilize multiple antenna in NOMA system to achieve high SE. In [46, 52], the random opportunistic beamforming is first proposed for the MIMO NOMA systems, where the transmitter generates multiple beams and superposes multiple users within each beam. In this chapter, beamforming design for a NOMA based downlink system is investigated, where a BS is equipped with multiple antennas serving a number of geographically distributed users. It is assumed that the CSI between the transmitter and the users is perfectly known at the transmitter. Then, two resource allocation problems are defined in MISO-NOMA systems to design the beamformers for power minimization and user fairness. Since both optimization problems are non-convex, Taylor series approximation and SDR approach are used to transform and approximate the original

non-convex problems to convex optimization problems. The problems are solved by iteratively solving the convex subproblems. Simulation results show that the NOMA system yields better performance in terms of power consumption and fairness than the conventional multiple access schemes.

4.2 System Model

A downlink transmission of a MISO NOMA system with K users, U_k, $k \in \{1, \ldots, K\}$ is considered as shown in Fig. 4.1, where BS employs NOMA scheme to transmit signals of different users. The BS is equipped with N transmit antennas, whereas each user equipment consists of a single antenna. The BS transmits a superposition of the individual messages to all users which linearly weighted by the corresponding beamforming vector. Each user receives its signals as well as interference signals from the other users. Hence, the signal received at the k^{th} user, U_k, is given by

$$y_k = \mathbf{h}_k^H \mathbf{w}_k s_k + \sum_{m \neq k} \mathbf{h}_k^H \mathbf{w}_m s_m + n_k, \tag{4.1}$$

where s_k and $\mathbf{w}_k$ are the transmitted information symbol intended for U_k and the corresponding transmit beamforming weight vector, respectively. The symbols are normalized so that they have unit energy and the messages are assumed to be independent, i.e., $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = \mathbf{I}$ where $\mathbf{s} = [s_1 \; s_2 \; \ldots \; s_K]^T$, and $\|\mathbf{w}_k\|_2^2$ represents the transmit power assigned to U_k. In addition, $\mathbf{h}_k \in \mathbb{C}^N$ denotes the complex channel vector between the BS and U_k and n_k represents a zero-mean circularly symmetric AWGN with variance σ_k^2 at U_k, (i.e., $n_k \sim \mathcal{CN}(0, \sigma_k^2)$).

In the NOMA scheme, user multiplexing is performed in the power domain and the SIC approach is employed at the receivers to separate signals between different users [15, 54]. In this scheme, users are sorted based on the norm of their channels, i.e., $\|\mathbf{h}_1\|_2 \leq \|\mathbf{h}_2\|_2 \leq \ldots \leq \|\mathbf{h}_K\|_2$. In this scenario, U_k should be able to decode the signals intended for the users from U_1 to U_k and effectively remove the interference from the first $k-1$ users, whereas the signals intended for the rest of the users, i.e., $U_{k+1}, \ldots, U_K$, are treated as the interference at U_k. Based on this SIC approach, the l^{th} user can decode and remove the k^{th} user's signals for all $1 \leq k < l$ in a successive manner. In other words, the k^{th} user's signal should be decoded by the l^{th} user for all $l \in \{k, k+1, \ldots, K\}$ [46, 79]. Hence, the remaining signal at the l^{th} user to decode

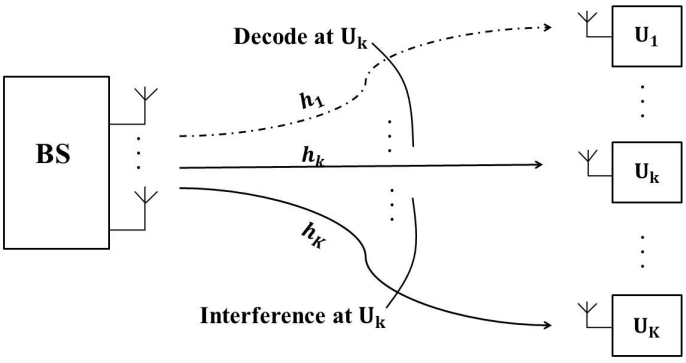


Fig. 4.1 The NOMA based downlink systems. One BS with multiple antennas serves users. The signals of users from U_1 to U_{k-1} are cancelled at k^{th} user, while the signals of users U_{k+1} to U_K are considered as interference.

the k^{th} user's signal is represented as follows:

$$y_l^{(k)} = \mathbf{h}_l^H \mathbf{w}_k s_k + \sum_{m=k+1}^{K} \mathbf{h}_l^H \mathbf{w}_m s_m + n_l, \tag{4.2}$$

where the first term is the desired signal to detect s_k and the second term treated as the interference from the signals intended for the users $U_{k+1}, \ldots, U_K$. Based on these conditions, the achievable rate of U_k can be defined as follows:

$$R_k = \log_2 \left(1 + \min_{l \in \{k, k+1, \ldots, K\}} \text{SINR}_l^{(k)} \right), \tag{4.3}$$

where

$$\text{SINR}_l^{(k)} = \frac{|\mathbf{h}_l^H \mathbf{w}_k|^2}{\sum_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2}, \tag{4.4}$$

represents the SINR of the signal intended for U_k, at the l^{th} user.

Moreover, the following conditions should be satisfied in NOMA scheme to guarantee the intended ordering of SIC in decoding the signals of the weaker users and to allocate non-trivial data rates to the weaker users [79]:

$$|\mathbf{h}_k \mathbf{w}_1|^2 \geq \ldots \geq |\mathbf{h}_k \mathbf{w}_{k-1}|^2 \geq |\mathbf{h}_k \mathbf{w}_k|^2 \geq |\mathbf{h}_k \mathbf{w}_{k+1}|^2 \geq \ldots \geq |\mathbf{h}_k \mathbf{w}_K|^2, \quad \forall k. \tag{4.5}$$

The sequence of above inequalities helps boost the SINRs needed to decode other users' messages. Also, in SIC based receivers, each user decodes its own message after decoding the messages of weaker users and successfully removing their interference. In order to facilitate this SIC technique, the received power of the signals to be decoded should be made greater than the received powers of the other users' signals. Hence, the above inequalities help to implement SIC by increasing the power of the signals intended for the weaker users. Through imposing these conditions, the users located far from the BS receive more signal power than that of the users near to the BS.

4.3 Power Optimization Framework

Low energy consumption is one of the key requirements in future wireless networks. Hence, a resource allocation problem is first considered to minimize the required total transmit power while satisfying a predefined QoS at each user. This scenario could arise in a network consisting of users with delay-intolerant real-time services (real-time users) [140]. These users should achieve their required QoS at all times, regardless of their channel conditions. In this section, a power minimization problem is considered to satisfy QoS at each user with perfect CSI assumption. Accordingly, the beamformers are designed to minimize the total transmit power while imposing constraints on the rate requirement at each user. This power minimization problem can be formulated into the following optimization framework:

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{4.6a}$$

$$\text{subject to} \quad \log_2\left(1 + \min_{l\in\{k,k+1,\ldots,K\}} \text{SINR}_l^{(k)}\right) \geq R_k^{\min}, \qquad \forall k, \tag{4.6b}$$

where the constraint in (4.6b) is considered to guarantee minimum data rate requirement $R_k^{\min}$ at U_k. The power allocation problem considered in (4.6) is non-convex in terms of $\mathbf{w}_k$. To handle this issue, first, some transformations are applied to simplify the constraints and then, different approaches are exploited to approximate the equivalent problem. Since $\log(.)$ is a non-decreasing function, the following equivalent constraint can be considered instead of (4.6b):

$$\min_{l\in\{k,k+1,\ldots,K\}} \text{SINR}_l^{(k)} \geq \gamma_k^{\min}, \qquad \forall k, \tag{4.7}$$

where $\gamma_k^{\min} = 2^{R_k^{\min}} - 1$ is the minimum SINR requirement at U_k. Then, (4.7) can be rewritten as follows:

$$\begin{cases} \text{SINR}_k^{(k)} \geq \gamma_k^{\min}, \\ \text{SINR}_{k+1}^{(k)} \geq \gamma_k^{\min}, \\ \vdots \\ \text{SINR}_K^{(k)} \geq \gamma_k^{\min}, \end{cases} \Leftrightarrow \begin{cases} \gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_k^H \mathbf{w}_m|^2 + \sigma_k^2 \right) \leq |\mathbf{h}_k^H \mathbf{w}_k|^2, \\ \gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_{k+1}^H \mathbf{w}_m|^2 + \sigma_{k+1}^2 \right) \leq |\mathbf{h}_{k+1}^H \mathbf{w}_k|^2, \\ \vdots \\ \gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_K^H \mathbf{w}_m|^2 + \sigma_K^2 \right) \leq |\mathbf{h}_K^H \mathbf{w}_k|^2, \end{cases}$$

$$\Leftrightarrow \gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2 \right) \leq |\mathbf{h}_l^H \mathbf{w}_k|^2, \qquad \forall k,\ l = k, \ldots, K. \tag{4.8}$$

Finally, the equivalent formulation of the original power minimization problem in (4.6) can be reformulated as

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N \times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{4.9a}$$

$$\text{subject to} \quad \gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2 \right) \leq |\mathbf{h}_l^H \mathbf{w}_k|^2, \quad \forall k,\ l = k, \ldots, K. \tag{4.9b}$$

The above problem is still non-convex due to the quadratic term on the right side of the inequality in (4.9b). To tackle this issue, some approximation methods are applied to convexify the optimization problem. Then, an iterative algorithm is eventually developed to solve the original power minimization problem. This type of algorithms will be initiated with the feasible solution which satisfies the constraints and then, solves the problem iteratively and yields a better solution in each iteration.

4.3.1 Taylor Series Approximation

In this subsection, convex approximations for the non-convex constraint in (4.9b) are provided. To handle these constraints, SOC representation is used to equivalently transform the constraint into a tractable form. In this beamforming design, choosing an arbitrary phase for $\mathbf{w}_k$ has no impact on the optimization and also provides the same solutions. Thus, any arbitrary phase can be selected for this beamformer. Furthermore, this enables to assume that $\mathbf{h}_l^H \mathbf{w}_k > 0$, which makes the square root of $|\mathbf{h}_l^H \mathbf{w}_k|^2$ well-defined [141, 142]. By reshuffling the constraints and taking their square root, the

non-convex constraints can be reformulated as a SOC as follows:

$$\gamma_k^{\min}\left(\sum_{m=k+1}^{K}|\mathbf{h}_l^H\mathbf{w}_m|^2+\sigma_l^2\right)\leq|\mathbf{h}_l^H\mathbf{w}_k|^2\Leftrightarrow\begin{cases}\sqrt{\gamma_k^{\min}}\left\|\begin{array}{c}|\mathbf{h}_l^H\mathbf{w}_{k+1}|\\ \vdots\\ |\mathbf{h}_l^H\mathbf{w}_K|\\ \sigma_l\end{array}\right\|\leq|\mathbf{h}_l^H\mathbf{w}_k|,\\ \\ \Im(\mathbf{h}_l^H\mathbf{w}_k)=0.\end{cases}\tag{4.10}$$

However, it is impossible to have a phase rotation to simultaneously satisfy the following conditions:

$$\Im(\mathbf{h}_k^H\mathbf{w}_k)=\Im(\mathbf{h}_l^H\mathbf{w}_k)=0,\quad\forall l=k+1,\ldots,K.\tag{4.11}$$

Therefore, this phase rotation has been only applied to satisfy $\Im(\mathbf{h}_l^H\mathbf{w}_k)=0$, for $l=k$, and exploited the Taylor series approximation [36, 32, 143] for $l=k+1,\ldots,K$ to convexify the non-convex constraints in (4.9b) based on the following Lemma:

Lemma 1: By using the first order Taylor approximation of the function $f_l(\mathbf{w}_k)$ around $\mathbf{w}_k^{(t)}$ in the t^{th} iteration, the following inequality holds

$$\begin{aligned}|\mathbf{h}_l^H\mathbf{w}_k|^2=\mathbf{w}_k^H\mathbf{h}_l\mathbf{h}_l^H\mathbf{w}_k\triangleq f_l(\mathbf{w}_k)\geq{\mathbf{w}_k^{(t)}}^H\mathbf{h}_l\mathbf{h}_l^H\mathbf{w}_k^{(t)}\\ +2\Re\left[{\mathbf{w}_k^{(t)}}^H\mathbf{h}_l{\mathbf{h}_l}^H(\mathbf{w}_k-\mathbf{w}_k^{(t)})\right]\triangleq g_l\left(\mathbf{w}_k,\mathbf{w}_k^{(t)}\right).\end{aligned}\tag{4.12}$$

This formulation is linear in terms of variable $\mathbf{w}_k$, and will be used instead of the original norm-squared function. Any inequality in constraint (4.9b) can be approximated to a convex one as follows:

$$\gamma_k^{\min}(\sum_{m=k+1}^{K}|\mathbf{h}_l^H\mathbf{w}_m|^2+\sigma_l^2)\leq g_l(\mathbf{w}_k,\mathbf{w}_k^{(t)}).\tag{4.13}$$

Proof: Please refer to Appendix A.1. ■

After applying the SOC representation in (4.10) and the approximation in Lemma 1, the following optimization problem is formulated:

Table 4.1 Proposed algorithm based on Taylor series approximation

Algorithm 1. Proposed Algorithm for solving problem (4.14)
1. **Initialization:** Set $t = 0$ and randomly generate a set of feasible $\mathbf{w}_k^{(0)},\ \forall k$ for problem in (4.14).
2. **repeat**
3. Solve problem (4.14)
4. Update $\{\mathbf{w}_k^{(t+1)}\} = \{\mathbf{w}_k^{(t)}\},\ \forall k$
5. $t \leftarrow t + 1$
6. **until** $\lvert\mathbf{w}_k^{(t+1)} - \mathbf{w}_k^{(t)}\rvert \leq \varepsilon$

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{4.14a}$$

$$\text{subject to} \quad \begin{cases} \sqrt{\gamma_k^{\min}} \left\| \begin{array}{c} |\mathbf{h}_k^H \mathbf{w}_{k+1}| \\ \vdots \\ |\mathbf{h}_k^H \mathbf{w}_K| \\ \sigma_l \end{array} \right\| \leq |\mathbf{h}_k^H \mathbf{w}_k|, \qquad \forall k, \\ \Im(\mathbf{h}_k^H \mathbf{w}_k) = 0, \end{cases} \tag{4.14b}$$

$$\gamma_k^{\min} \left(\sum_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2 \right) \leq g_l \left(\mathbf{w}_k, \mathbf{w}_k^{(t)} \right), \quad \forall k,\ l = k+1, \ldots, K. \tag{4.14c}$$

An iterative algorithm is developed to solve the power minimization problem based on the approximated problem in (4.14) which is summarized in Table 4.1. This algorithm will be initialized with $\mathbf{w}_k^{(t)}$ and the corresponding approximated problem will be solved to obtain the beamforming vector, i.e., $\mathbf{w}_k^{(t+1)}$. In other words, the corresponding initial solution is updated iteratively and the algorithm will be terminated once the required accuracy is achieved.

4.3.2 Semidefinite Relaxation Approach

SDP is a standard subfield of convex optimization problem that consists of a linear objective function over the intersection of the cone of positive semidefinite matrices with an affine space. SDPs can be efficiently solved by interior point methods. By using SDR, the non-convex constraint can be relaxed and the problem can be converted to a SDP problem [140, 141]. This relaxation formulates the non-convex optimization

problem into a convex optimization problem where the cost and constraints are convex in nature. In this section, an approach based on SDR is proposed to solve the original non-convex power minimization problem in (4.9). By considering $\mathbf{H}_k = \mathbf{h}_k\mathbf{h}_k^H$ and $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$, a new matrix variable $\mathbf{W}_k$ is introduced and the original optimization problem in (4.9) can be reformulated into the following optimization framework:

$$\min_{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k) \tag{4.15a}$$

$$\text{subject to} \quad \gamma_k^{\min}\left(\sum_{m=k+1}^{K} \mathrm{Tr}(\mathbf{H}_l\mathbf{W}_m) + \sigma_l^2\right) \leq \mathrm{Tr}(\mathbf{H}_l\mathbf{W}_k),\ \ \forall k,\ l = k,\ldots,K, \tag{4.15b}$$

$$\mathbf{W}_k \succcurlyeq 0,\ \ \forall k, \tag{4.15c}$$

$$\mathrm{rank}(\mathbf{W}_k) = 1,\ \ \forall k. \tag{4.15d}$$

Note that, all the constraints except the rank one constraint in (4.15d) are convex. To obtain a solution, the rank-one constraint is relaxed by exploiting the SDR approach. Without the rank-one constraint, the following optimization problem is solved:

$$\min_{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k) \tag{4.16a}$$

$$\text{subject to} \quad \gamma_k^{\min}\left(\sum_{m=k+1}^{K} \mathrm{Tr}(\mathbf{H}_l\mathbf{W}_m) + \sigma_l^2\right) \leq \mathrm{Tr}(\mathbf{H}_l\mathbf{W}_k),\ \ \forall k,\ l = k,\ldots,K, \tag{4.16b}$$

$$\mathbf{W}_k \succcurlyeq 0,\ \ \forall k. \tag{4.16c}$$

Since (4.16) is a standard SDP, it can be efficiently solved through convex optimization techniques. In general, if the solution of the relaxed problem in (4.16) is a set of rank-one matrices $\mathbf{W}_k$, then it will be also the optimal solution to the original problem in (4.15). Otherwise, the randomization technique can be used to generate a set of rank-one solutions [115]. The beamforming vector $\mathbf{w}_k$ can be obtained from a rank-one $\mathbf{W}_k$ solution, as $\mathbf{w}_k = \sqrt{\lambda_k}\mathbf{v}_k$ where λ_k and $\mathbf{v}_k$ are the maximum eigenvalue and the corresponding eigenvector of $\mathbf{W}_k$, respectively.

4.3.3 Complexity Analysis

Computational complexity theory is a branch of the theory of computation in theoretical computer science that focuses on classifying computational problems according to their inherent difficulty of solving a problem. In fact, the computational complexity helps algorithm developers to identify how difficult for solving a problem is. A computational problem is understood to be a task that is in principle amenable to being solved by mechanical application of mathematical steps, such as an algorithm. This is equivalent to stating that the problem may be solved by a computer. A problem is regarded as inherently difficult if its solution requires significant resources in the algorithm. The theory formalizes this intuition, by introducing mathematical models of computation to study these problems and quantifying their computational complexity, i.e., the amount of resources needed to solve them, such as time and storage [144].

In this section, the complexity of two different approaches developed to transform the original non-convex optimization problem to a convex one is investigated. In the SDR approach, the optimization problem is reformulated in SDP form by relaxing the non-convex rank one constraint. The optimal solution of the original problem can be obtained from this simple SDR method if it yields rank-one solutions. On the other hand, it is possible in some cases for the solution of the relaxed problem to turn out not to be rank-one. In this case, the proposed Taylor series approximation can be employed to convexify the original problem, resulting in a suboptimal solution. The complexity of the proposed algorithms are analyzed by evaluating the computational complexity of each problem based on the complexity of the interior point methods [137, 131]. This complexity can be defined by quantifying the required number of arithmetic operations in the worst-case at each iteration and the required number of iterations to achieve the solutions with a certain accuracy. The computational complexity for each algorithm is defined as follows:

1) In the Taylor series approximation, the beamformer design in the power minimization problem is formulated into a SOCP in problem (4.14). Therefore, the worst-case complexity is determined by the SOCP in each step. It is well known that for general interior-point methods the complexity of the SOCP depends upon the number of constraints, variables and the dimension of each SOC constraint. The total number of constraints in the formulation of (4.14) is $0.5K^2 + 1.5K$. Therefore, the number of iterations needed to converge with ε solution accuracy at the termination of the algorithm is $O(\sqrt{0.5K^2 + 1.5K} \log \frac{1}{\varepsilon})$ [137]. Each iteration requires at most $O\big((NK)^2(0.33K^3 + 0.5K^2 + 1.16K + 1)\big)$ arithmetic operations to solve the SOCP

where NK and $0.33K^3+0.5K^2+1.16K+1$ are the number of optimization variables and the total dimension of the SOC constraints in (4.14).

2) In SDP, the algorithm finds an ε-optimal solution for the semidefinite problem with an n dimensional semidefinite cone in at most $O(\sqrt{n}\log\frac{1}{\varepsilon})$ iterations where $n = N^2$ in the problem in (4.16). Each iteration requires at most $O(mn^3+m^2n^2+m^3)$ arithmetic operations to solve the SDP where m denotes the number of semidefinite constraints [131]. Thus, $O\big(0.5K(K+1)N^6+0.25K^2(K+1)^2N^4+0.125K^3(K+1)^3\big)$ arithmetic operations are required in each iteration of solving the problem in (4.16).

In summary, the first scheme has a lower worst-case complexity than an SDR scheme. In contrast to the semidefinite formulation, there is no need to introduce the additional matrices $\mathbf{W}_k$ for the first scheme and the resulting optimization involves significantly fewer variables. However, first scheme has to deal with an approximation which makes the solution suboptimal. On the other hand, the SDR method can yield the optimal solution if it given a set of rank-one matrices which eliminates the need for the iterative approach as in the Taylor series approximation scheme. Note that the first scheme requires an iterative process, however, as seen in simulation results section, this approach converges with a small number of iterations which has no significant impact on the order of the complexity of the proposed algorithm.

4.4 Max-Min Fairness Problem

For the previously considered power minimization approach, the transmitter requires a certain amount of transmit power to achieve the target rate at each user. However, the maximum available transmit power is generally limited at the transmitter and therefore the power minimization problem might turn out to be infeasible due to insufficient transmit power. In this case, the target rate should be decreased, and optimization should be repeatedly performed until the problem becomes feasible. To overcome this infeasibility issue, a max-min fairness based approach is considered, in which the minimum rate between all users is maximized while satisfying the transmit power constraint [145]. This practical scenario could arise in a network consisting of users with delay-tolerant packet data services (non-real time users) [146, 147], where packet size could be varied according to the achievable rate value. In the context of NOMA systems, there are a number of works that consider the max-min fairness scheme in single antenna NOMA systems [148, 149]. However, this has not been studied for a multi-antenna NOMA system. In this section, a max-min fairness problem for the

NOMA downlink system has been investigated. Since the users' rate in the NOMA scheme can be managed with more flexibility, it may be an efficient way to enhance user fairness for scenarios with strict fairness constraints.

Therefore, to balance the rate between different users in the network, the max-min fairness approach is an appropriate criterion. The corresponding max-min fairness problem can be formulated as follows:

$$\max_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \min_k R_k, \tag{4.17a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \leq P^{\max}, \tag{4.17b}$$

$$|\mathbf{h}_k^H \mathbf{w}_1|^2 \geq \ldots \geq |\mathbf{h}_k^H \mathbf{w}_{k-1}|^2 \geq |\mathbf{h}_k^H \mathbf{w}_k|^2 \geq |\mathbf{h}_k^H \mathbf{w}_{k+1}|^2 \geq \ldots \geq |\mathbf{h}_k^H \mathbf{w}_K|^2, \quad \forall k, \tag{4.17c}$$

where R_k is given in (4.3). The constraint in (4.17b) represents the maximum available total transmit power, i.e., $P^{\max}$ and the constraint in (4.17c) is considered to ensure the given decoding order of SIC. It is difficult to realize the optimal solution for this max-min problem due to its non-convex nature. Hence, the problem is first transformed into a convex one, then utilize a low-complexity polynomial algorithm to find an optimal solution.

Lemma 2: This max-min fairness problem is quasi-concave and can be solved through a bisection method.

Proof: Please refer to Appendix A.2. ■

First, some transformations are considered to simplify the constraints. From the inequalities in (4.17c), it holds that

$$\begin{cases} |\mathbf{h}_k^H \mathbf{w}_2|^2 \leq |\mathbf{h}_k^H \mathbf{w}_1|^2, \\ \vdots \\ |\mathbf{h}_k^H \mathbf{w}_{k+1}|^2 \leq \min_{m\in[1,k]} |\mathbf{h}_k^H \mathbf{w}_m|^2, \\ \vdots \\ |\mathbf{h}_k^H \mathbf{w}_K|^2 \leq \min_{m\in[1,K-1]} |\mathbf{h}_k^H \mathbf{w}_m|^2, \end{cases}$$

$$\Leftrightarrow |\mathbf{h}_k^H \mathbf{w}_n|^2 \leq |\mathbf{h}_k^H \mathbf{w}_m|^2, \quad \forall k,\ n = 2, \ldots, K,\ m = 1, \ldots, n-1. \tag{4.18}$$

Table 4.2 Proposed algorithm based on bisection method

Algorithm 2 Proposed Algorithm for solving problem (4.17)
1. **Initialization:** Set $\alpha_{\min} = 0, \alpha_{\max} = \log_2\left(1 + \frac{P^{\max}\lvert\mathbf{h}_K\rvert^2}{\sigma_K^2}\right)$,
2. **repeat**
3. Set $\alpha = (\alpha_{\max} + \alpha_{\min})/2$ and solve (4.21) to obtain $\mathbf{w}_0$
4. **if** (4.17b) is satisfied **then**
5. Set $\alpha_{\min} = \alpha;\ \mathbf{w}^* = \mathbf{w}_0$;
6. **else**
7. $\alpha_{\max} = \alpha$
8. **until** $(\alpha_{\max} - \alpha_{\min} \leq \varepsilon)$.

After these simplifications, Lemma 1 in Section 4.3.1 can be employed to convexify (4.18) as

$$|\mathbf{h}_k^H \mathbf{w}_n|^2 \leq g_k(\mathbf{w}_m, \mathbf{w}_m^{(t)}), \quad \forall k,\ n = 2, \ldots, K,\ m = 1, \ldots, n-1, \tag{4.19}$$

where $g_k(\mathbf{w}_m, \mathbf{w}_m^{(t)})$ is the Taylor series approximation of the term $|\mathbf{h}_k^H \mathbf{w}_m|^2$ around $\mathbf{w}_m^{(t)}$ in the t^{th} iteration.

Similarly, the equivalent convex formulation for (A.9) can be reformulated as

$$(2^{\alpha} - 1)(\Sigma_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2) \leq g_l(\mathbf{w}_k, \mathbf{w}_k^{(t)}), \quad \forall k,\ l = k, k+1, \ldots, K. \tag{4.20}$$

where $g_l(\mathbf{w}_k, \mathbf{w}_k^{(t)})$ is the Taylor series approximation of the term $|\mathbf{h}_l^H \mathbf{w}_k|^2$ around $\mathbf{w}_k^{(t)}$ in the t^{th} iteration.

In order to solve this problem through a bisection method, assume that R^* denotes the optimal value of the objective function of the problem in (4.17). For a given threshold α, if there exists a set of $\mathbf{w}_0 = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ that satisfies the constraints (4.17b),(4.19) and (4.20), then $R^* \geq \alpha$, otherwise $R^* \leq \alpha$. Equivalently, the following problem can be solved

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{4.21a}$$

$$\text{subject to} \quad \text{(4.19) and (4.20)}, \tag{4.21b}$$

and determined whether the solution satisfies the total power constraint $\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \leq P^{\max}$. By appropriately choosing α through a bisection method, the solution of (4.17), $\mathbf{w}^*$ can be obtained by solving a sequence of feasibility problems of (4.21). Table 4.2

presents the proposed bisection method for realizing the solution for the problem in (4.17).

4.5 Simulation Results

In this section, the performance of the proposed beamforming designs for the NOMA scheme is evaluated through numerical simulations. A single cell downlink transmission is consider where a multi-antenna BS serves single-antenna users which are uniformly distributed within the circle with a radius of 50 meters around the BS, but no closer than 1 meter. The small-scale fading of the channels is assumed to be Rayleigh fading which represents an isotropic scattering environment. The large-scale fading effect is modelled by ${d_k}^{-\beta}$ to incorporate the path-loss effects where d_k is the distance between U_k and the BS, measured in meters and β is the path-loss exponent. More precisely, $\mathbf{h}_k = \chi_k \sqrt{{d_k}^{-\beta}}$ where $\chi_k \sim \mathcal{CN}(0, \mathbf{I})$ refer to small scale fading channels. It should be noted here that in simulations the user distances are fixed and the average is taken over the fast fading component of the channel vectors and the path-loss exponent is considered as $\beta = 3.8$. Moreover, the noise variance at each user is assumed to be 0.01 (i.e., $\sigma_k^2 = 0.01$) and it is assumed that the target rates for all users are the same. For the simulations, the Yalmip package [150] is used and the numerical results are obtained by averaging over different 1000 random channels.

To compare the performance of NOMA with conventional methods, the simulation results are also achieved for OMA and ZF schemes. For OMA scheme, the following optimization problem is solved:

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{4.22a}$$

$$\text{subject to} \quad \frac{1}{K}\log_2\left(1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sigma_k^2}\right) \geq R_k^{\min}, \qquad \forall k, \tag{4.22b}$$

where the OMA and NOMA schemes are compared in terms of required bandwidth. In other words, the OMA scheme requires K sub-channels to cover K users, while, in the NOMA scheme, these users are allowed to simultaneously share one sub-channel via the power domain. Hence, more bandwidth is required in the OMA scheme in comparison with the NOMA scheme.

In the ZF scheme, the beamforming vector is considered as $\mathbf{w}_k = \sqrt{p_k}\mathbf{v}_k$ where p_k and $\mathbf{v}_k$ are power and direction associated to the k^{th} user's data, respectively. Then,

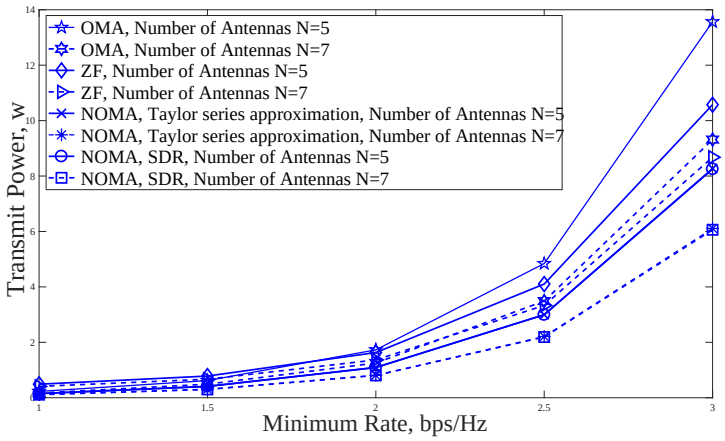


Fig. 4.2 The required total transmit power to achieve different target rates for 3 users in NOMA, ZF and OMA schemes.

the vectors $\mathbf{v}_k, \forall k$, are designed to avoid mutual interference between users. In other words, the following condition should be satisfied:

$$\mathbf{h}_i^H \mathbf{v}_k = 0, \quad \forall i \neq k. \tag{4.23}$$

To design the beamforming vector which satisfy condition in (4.23), the following matrix is first defined:

$$\mathbf{H}_k = [\mathbf{h}_1 \cdots \mathbf{h}_{k-1}\ \mathbf{h}_{k+1} \cdots \mathbf{h}_K], \tag{4.24}$$

The null space of the matrix $\mathbf{H}_k$ in (4.24) can be utilized for the vector $\mathbf{v}_k$ which results in $\mathbf{H}_k^H \mathbf{v}_k = \mathbf{0}$ and satisfies condition in (4.23). Then, the following power allocation problem is solved:

$$\min_{p_k,\, \forall k} \quad \sum_{k=1}^{K} p_k, \tag{4.25a}$$

$$\text{subject to} \quad \log_2\left(1 + \frac{p_k |\mathbf{h}_k^H \mathbf{v}_k|^2}{\sigma_k^2}\right) \geq R_k^{\min}, \qquad \forall k, \tag{4.25b}$$

4.5.1 Total Transmit Power

First, the required total transmit power is evaluated for both power minimization approaches (i.e., Taylor series approximation and SDR) with different system parameters.

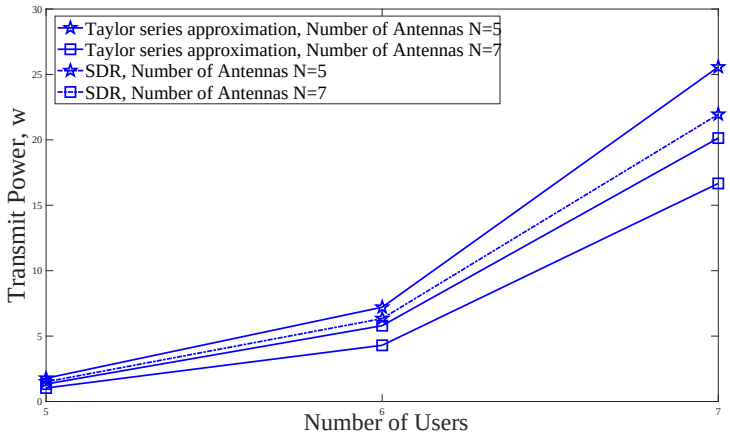


Fig. 4.3 The required total transmit power to achieve $R_k^{\min} = 2$ bps/Hz for different numbers of users by using Taylor series approximation and SDR methods.

The required total transmit power against different target rates is presented in Fig. 4.2 for the NOMA and conventional multiple access schemes with different numbers of transmit antennas. By increasing the minimum required rate at each user, the BS requires more power to satisfy the target rate requirements. For a given target rate, the required total transmit power can be reduced by employing more antennas at the transmitter. As shown in Fig. 4.2, for a specific rate requirement, the conventional multiple access techniques consume more transmit power than the NOMA scheme. This demonstrates that the NOMA scheme outperforms the conventional multiple access in terms of power consumption.

In Fig. 4.3, the required total transmit power for different number of users with different number of transmit antennas is obtained. As the number of antennas increases, the required transmit power decreases due to the spatial diversity gain. However, the BS requires more transmit power as the number of users increases. As shown in Fig. 4.3, both schemes, Taylor series approximation and SDR show a similar performance for a few users due to the small number of approximated terms in the Taylor series approximation. However, the number of approximated terms increases with the number of users. As a result, the performance gap between these two schemes increases and SDR outperforms the Taylor series approximation scheme in terms of required transmit power. The reason is that SDR can provide the optimal solution given that the solution is rank one whereas the other scheme relies on the Taylor series approximation which might lead to a suboptimal solution.

Table 4.3 Comparison of power allocations in two approaches. Taylor series Approximation and SDR method.

	Taylor Approximation				SDR Method			
Channels	U_1 Power (w)	U_2 Power (w)	U_3 Power (w)	Total Power (w)	U_1 Power (w)	U_2 Power (w)	U_3 Power (w)	Total Power (w)
Channel 1	5.3596	0.6702	0.0821	6.1119	5.3593	0.6701	0.0820	6.1114
Channel 2	3.7419	0.4692	0.0636	4.2747	3.7417	0.4690	0.0635	4.2742
Channel 3	7.6185	0.9595	0.1232	8.7012	7.6156	0.9591	0.1230	8.6977
Channel 4	10.2415	1.2811	0.1642	11.6868	10.2367	1.2808	0.1641	11.6814
Channel 5	8.4636	1.0626	0.1468	9.6730	8.4634	1.0625	0.1467	9.6726

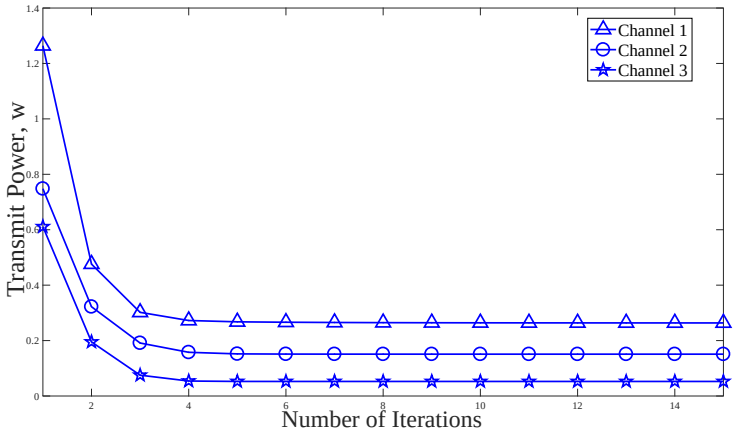


Fig. 4.4 The convergence of the algorithm in Table 4.1 for different set of channels. Number of users$=3$, Number of antennas$=6$, Target rate$=1$.

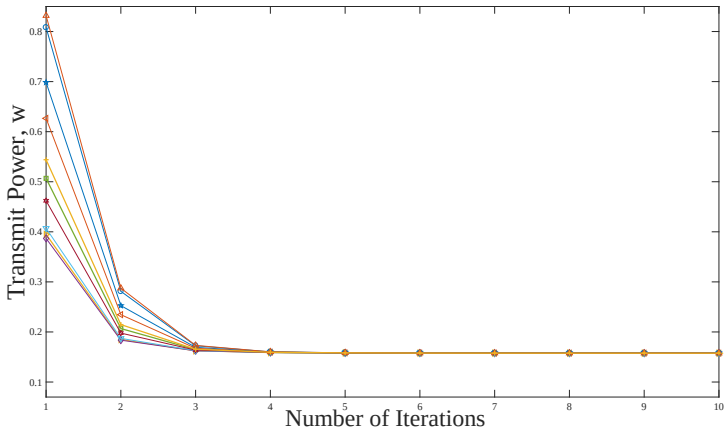


Fig. 4.5 The convergence of the algorithm based on Taylor series approximation for different initializations. Number of users$=3$, Number of antennas$=5$, Target rate$=1$.

Fig. 4.4 depicts the convergence of the algorithm provided in Table 4.1 in terms of transmit power. As shown, this approach converges with a small number of iterations

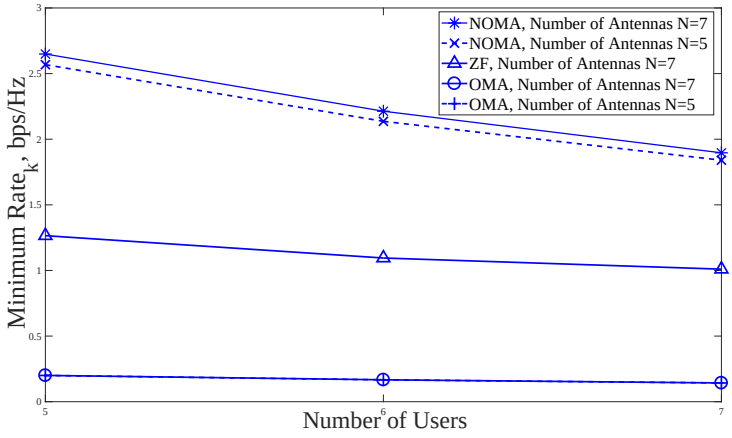


Fig. 4.6 The minimum achieved rate for different numbers of users with $P^{\max} = 10$ watt in NOMA, ZF and OMA schemes.

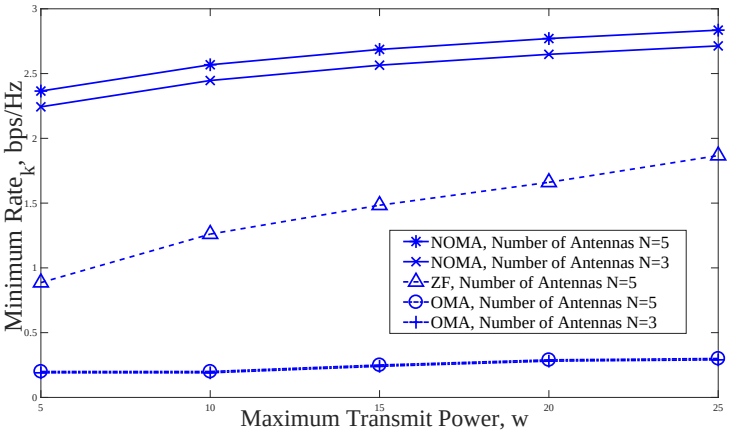


Fig. 4.7 The minimum achieved rate for 5 users with different $P^{\max}$ in NOMA, ZF and OMA schemes.

(most of the time with 3 iterations), which does not have a significant impact on the order of the complexity of the proposed algorithm. Moreover, the impact of the initialization of the algorithm is numerically evaluated on the convergence of the Taylor series approximation. As shown in Fig. 4.5, the Taylor series approximation method converges to the same solution with different initializations. Table 4.3 is provided to compare the required transmit power for each user and the total transmit power obtained through the Taylor series approximation and SDR approaches. As evidenced by these results, there is no significant difference between the two proposed approaches.

4.5.2 Achieved Minimum Rates in Max-min Approach

Next, the performance of the max-min fairness design is studied for both NOMA and conventional multiple access schemes. The balanced rates maintaining fairness between users are demonstrated in Fig. 4.6 and Fig. 4.7, respectively, for different number of users and maximum available transmit power with different number of antennas. As expected, reducing the number of users or increasing the maximum available transmit power threshold improves the achievable fairness rate. Since the fairness rate is a logarithmic function of power, as the power threshold increases the rate improvement is compressed. These simulation results confirm that the QoS based beamforming design satisfies the required rate constraints at each user whereas the rates of the users are balanced in the fairness based approach. As shown in Fig. 4.6 and Fig. 4.7, for a specific available power, the NOMA scheme achieves higher rate than the conventional multiple access techniques.

The power allocations and the balanced rates obtained by solving problem (4.17) are provided for five different random channels in Table 4.4. In order to validate the optimality of the proposed max-min fairness approach, it is compared with the power allocations through the power minimization solution in Section 4.3. In particular, the balanced rates obtained through the fairness approach have been set as the target rates in the power minimization approach for the same set of channels, and the corresponding power allocations are obtained. As seen in Table 4.4 and Table 4.5, both max-min fairness and power minimization approaches utilize the same power allocations to achieve same rates at each user. This confirms the optimality of the proposed max-min fairness based design as the power minimization approach is optimal for a given set of target rates.

Table 4.4 Power allocations and achieved rates obtained through max-min fairness approach.

Channels	U_1 Power (w)	U_2 Power (w)	U_3 Power (w)	Total Power (w)	U_1 Rate (bps/Hz)	U_2 Rate (bps/Hz)	U_3 Rate (bps/Hz)
Channel 1	13.9225	0.9990	0.0804	15	3.8010	3.8010	3.8010
Channel 2	4.3137	0.5965	0.0929	5	2.8584	2.8584	2.8584
Channel 3	9.1271	0.8013	0.0730	10	3.5140	3.5140	3.5140
Channel 4	6.7110	0.7166	0.0754	7.5	3.2327	3.2327	3.2327
Channel 5	11.5454	0.8814	0.0752	12.5	3.7119	3.7119	3.7119

Table 4.5 Total required transmit power and power allocation at each user for different target rates with $K = 3$.

Channels	Target Rate (bps/Hz)	Total Power (w)	U_1 Power (w)	U_2 Power (w)	U_3 Power (w)
Channel 1	3.8010	15	13.9225	0.9990	0.0804
Channel 2	2.8584	5	4.3137	0.5965	0.0929
Channel 3	3.5140	10	9.1271	0.8013	0.0730
Channel 4	3.2327	7.5	6.7110	0.7166	0.0754
Channel 5	3.7119	12.5	11.5454	0.8814	0.0752

4.6 Summary

In this chapter, the beamforming design has been proposed in MISO-NOMA system with perfect CSI. Low energy consumption is one of the key requirements in future wireless networks. As such, a resource allocation problem to minimize the required total transmit power is first considered while satisfying a predefined QoS at each user. In particular, the power minimization problem is studied, where the beamformers are designed to provide the minimum rate requirements at each user. This optimization problem is non-convex in terms of beamforming vectors. To circumvent this non-convexity issue, CCP is employed based on Taylor series approximation and SDR to design the beamforming. Based on this approximation, an iterative algorithm is proposed to design the beamforming vectors while satisfying the rate constraint. For the previously considered power minimization approach, the transmitter requires a certain amount of transmit power to achieve the target rate at each user. However, the maximum available transmit power is generally limited at the transmitter and therefore the power minimization problem might turn out to be infeasible du to insufficient transmit power. To overcome this infeasibility issue, a max-min fairness based approach is considered, in which the minimum rate between all users is maximized while satisfying the transmit power constraint. Hence, the max-min problem is considered in a MISO NOMA system to maintain fairness between users and solved by exploiting bisection approach. Simulation results have been provided to validate the performance of the proposed schemes in the terms of the required transmit power and balanced rates. It is shown that NOMA can ensure high fairness requirements through appropriate power allocation. These simulation results also demonstrate that NOMA can achieve superior performance in terms of power consumption and system throughput compared to the OMA and ZF schemes and can efficiently utilize the bandwidth resources.

Chapter 5

Robust Beamforming Designs for MISO NOMA Systems with Imperfect CSI

5.1 Introduction

In the wireless systems, the channel parameters are prone to imperfection. Hence, there are practical difficulties in having the perfect CSI at the transmitter due to the channel estimation and quantization errors. Ignoring the effect of CSI uncertainties in the design for wireless networks can lead to solutions that may violate critical constraints and results in a poor performance in realistic channel conditions. In particular, these channel uncertainties might significantly degrade the system performance in NOMA, since the decoding order of the received signal intended for different users plays a crucial role in SIC based NOMA schemes [117]. Therefore, it is important to incorporate these channel uncertainties in the beamforming design to cultivate the desirable benefits offered by NOMA. To circumvent the inevitable channel uncertainties, robust design is a well-known approach, which can be classified into two groups, the worst-case robust design [111–114], and the outage probability based design [115, 116]. In the worst-case design, it is assumed that the CSI errors belong to some known bounded uncertainty sets whereas in the outage probability based design, the channel errors are random with a certain statistical distribution and constraints can be satisfied with certain outage probabilities. In fact, the bounded robust optimization is generally conservative owing to its worst-case criterion while probabilistic SINR constrained beamforming provides a soft SINR control. In the context of NOMA, a ro-

bust design with the norm-bounded channel uncertainties is investigated in [95] where a clustering scheme is studied to maximize the worst-case achievable sum rate with a total transmit power constraint.

In this chapter, both methods are employed to design the robust beamforming in a MISO-NOMA system with imperfect CSI between the BS and the users. First, the robust power minimization problem is investigated where the beamformers are designed by incorporating norm-bounded channel uncertainties to provide the required QoS at each user. The constraints are formulated based on the worst-case performance optimization framework. By exploiting S-Procedure, the original non-convex problem is converted into LMI and solved by using interior point methods. Next, an outage probabilistic based robust scheme is considered where the total transmit power is minimized while satisfying the outage probability constraint at each user. The problem is transformed into an LMI form through exploiting SDR approach to obtain the solution of the original non-convex problem. Simulation results show that proposed robust schemes outperform the non-robust scheme in terms of the rate satisfaction ratio at each user.

5.2 Worst-Case Robust Design

In the worst-case performance optimization, the maximum/minimum value of the objective function over all possible errors should be minimized/maximized using a set of constraints on the worst possible errors [111–114]. In this section, the worst-case robust design is investigated to incorporate imperfect CSI in the beamforming design. This technique ensures that the constraint associated with the QoS is satisfied all the time regardless of the channel uncertainties. Consider the same system model as in Chapter 4 where a BS equipped with N antennas sends information to K single antenna users, i.e., $U_1, U_2, \ldots, U_K$ with imperfect CSI assumption as shown in Fig. 5.1. In particular, the norm-bounded channel uncertainties are incorporated in the beamforming design to propose a robust design for the downlink MISO-NOMA systems. In this model, it is assumed that the channel estimation error is bounded in a known region and the true CSI can be expressed as

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \Delta\hat{\mathbf{h}}_k, \tag{5.1}$$

$$\|\Delta\hat{\mathbf{h}}_k\|_2 = \|\mathbf{h}_k - \hat{\mathbf{h}}_k\|_2 \leq \varepsilon_k, \tag{5.2}$$

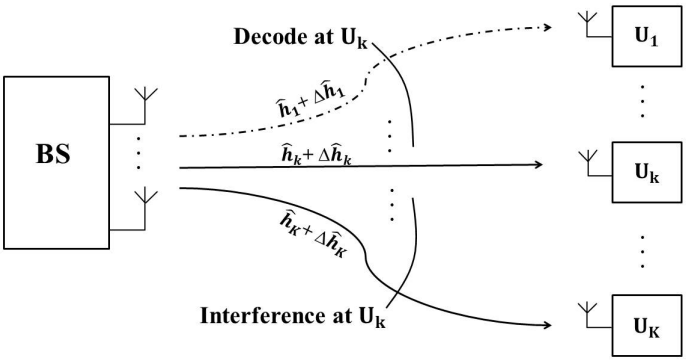


Fig. 5.1 The NOMA based downlink system where imperfect CSI is available at the transmitter. One BS with N antennas serves K users. The signals of users from U_1 to U_{k-1} are cancelled at k^{th} user, while the signals of users U_{k+1} to U_K are considered as interference.

where $\hat{\mathbf{h}}_k$, $\Delta\hat{\mathbf{h}}_k$ and $\varepsilon_k \geq 0$ denote the estimate of $\mathbf{h}_k$, the norm-bounded channel estimation error and the channel estimation error bound, respectively. The received signal at U_k is given by

$$y_k = (\hat{\mathbf{h}}_k + \Delta\hat{\mathbf{h}}_k)\mathbf{w}_k s_k + \sum_{m\neq k}(\hat{\mathbf{h}}_k + \Delta\hat{\mathbf{h}}_k)\mathbf{w}_m s_m + n_k, \qquad \forall k. \tag{5.3}$$

By applying the SIC at receivers, the signal at the l^{th} user after removing the first $k-1$ users' signals to detect the k^{th} user is represented as

$$y_l^{(k)} = (\hat{\mathbf{h}}_l + \Delta\hat{\mathbf{h}}_l)\mathbf{w}_k s_k + \sum_{m=1}^{k-1}\Delta\hat{\mathbf{h}}_l\mathbf{w}_m s_m + \sum_{m=k+1}^{K}(\hat{\mathbf{h}}_l + \Delta\hat{\mathbf{h}}_l)^H\mathbf{w}_m s_m + n_l,$$
$$\forall k,\ \ l = k, k+1, \ldots, K, \tag{5.4}$$

where the first term is the desired signal to detect s_k and other terms treated as the interference. The second term is due to imperfect CSI during the SIC process where the signals intended for the users $U_1, \ldots, U_{k-1}$ cannot be completely removed by the l^{th} user. The third term is the interference introduced by the signals intended to the users $U_{k+1}, \ldots, U_K$. Therefore, the SINR of the signal intended for U_k at the l^{th} user

can be written as

$$\text{SINR}_l^{(k)} = \frac{(\hat{\mathbf{h}}_l + \Delta\hat{\mathbf{h}}_l)^H \mathbf{w}_k \mathbf{w}_k^H (\hat{\mathbf{h}}_l + \Delta\hat{\mathbf{h}}_l)}{\sum_{m=1}^{k-1} \Delta\hat{\mathbf{h}}_l^H \mathbf{w}_m \mathbf{w}_m^H \Delta\hat{\mathbf{h}}_l + \sum_{m=k+1}^{K} (\hat{\mathbf{h}}_k + \Delta\hat{\mathbf{h}}_k)^H \mathbf{w}_m \mathbf{w}_m^H (\hat{\mathbf{h}}_k + \Delta\hat{\mathbf{h}}_k) + \sigma_l^2}. \tag{5.5}$$

For this network setup, the beamformers are designed based on the worst-case performance optimization framework to tackle the norm-bounded channel uncertainties. This robust beamforming design can be formulated as

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\, \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{5.6a}$$

$$\text{subject to} \quad \min_{l\in\{k,k+1,\ldots,K\}} \Big(\min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \le \varepsilon_l} \text{SINR}_l^{(k)} \Big) \ge \gamma_k^{\min},\ \forall k, \tag{5.6b}$$

where $\gamma_k^{\min} = (2^{R_k^{\min}} - 1)$ is the minimum required SINR to achieve a target rate $R_k^{\min}$ at U_k and the worst-case SINR denotes lower bound of SINR over the channel uncertainties $\Delta\hat{\mathbf{h}}_l$, i.e., $\min_{\|\Delta\hat{\mathbf{h}}_l\|_2} \text{SINR}_l^{(k)}$. It is obvious that the problem formulation in (5.6) is non-convex in terms of $\mathbf{w}_k$ and channel uncertainties, and cannot be solved directly. Thus, the equivalent transformation of the constraint in (5.6b) is first considered by introducing a new matrix variable $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$ as follows:

$$\begin{aligned}
&\min_{l\in\{k,k+1,\ldots,K\}} \Big(\min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \le \varepsilon_l} \text{SINR}_l^{(k)} \Big) \ge \gamma_k^{\min} \\
&\quad\Leftrightarrow \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \le \varepsilon_l} \big(\text{SINR}_l^{(k)} \big) \ge \gamma_k^{\min},\ \forall l = k, k+1, \ldots, K, \\
&\quad\Leftrightarrow \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \le \varepsilon_l} \left(\frac{\mathbf{h}_l^H \mathbf{W}_k \mathbf{h}_l}{\sum_{m=1}^{k-1} \Delta\hat{\mathbf{h}}_l^H \mathbf{W}_m \Delta\hat{\mathbf{h}}_l + \sum_{m=k+1}^{K} \mathbf{h}_l^H \mathbf{W}_m \mathbf{h}_l + \sigma_l^2} \right) \ge \gamma_k^{\min}, \\
&\qquad\qquad \forall l = k, k+1, \ldots, K.
\end{aligned} \tag{5.7}$$

Without loss of generality, the original optimization problem in (5.6) can be reformulated into the following optimization framework:

$$\min_{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k), \tag{5.8a}$$

$$\text{subject to} \quad \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \le \varepsilon_l} \left(\frac{\mathbf{h}_l^H \mathbf{W}_k \mathbf{h}_l}{\sum_{m=1}^{k-1} \Delta\hat{\mathbf{h}}_l^H \mathbf{W}_m \Delta\hat{\mathbf{h}}_l + \sum_{m=k+1}^{K} \mathbf{h}_l^H \mathbf{W}_m \mathbf{h}_l + \sigma_l^2} \right) \ge \gamma_k^{\min}, \quad \forall k,\ l = k, k+1, \ldots, K, \tag{5.8b}$$

$$\mathbf{W}_k \succeq 0, \qquad \forall k, \tag{5.8c}$$

$$\mathrm{rank}(\mathbf{W}_k) = 1, \qquad \forall k. \tag{5.8d}$$

The reformulated problem in (5.8) is still non-convex due to the rank-one constraint and unknown channel uncertainties, i.e., $\Delta\hat{\mathbf{h}}_k$. The inclusion of estimation uncertainties leads to the problem (5.8) intractable since the exact statistical information of $\Delta\hat{\mathbf{h}}_k$ is unknown. The rank-one constraint in (5.8d) can be relaxed by exploiting SDR. To remove the unknown channel uncertainties, S-procedure can be employed to reformulate the constraints as LMIs [151]. This is possible since the error bound of the channel estimation is known.

Lemma 3: By relaxing the rank-one constraints on $\mathbf{W}_k$, the original problem in (5.8) can be recast into the following convex problem:

$$\min_{\substack{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k \\ \mu_{kl} \ge 0}} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k), \qquad , \forall k,l \tag{5.9a}$$

$$\text{subject to} \quad C_{kl} = \begin{bmatrix} \mu_{kl}\mathbf{I} + \varphi_k & \phi_k \hat{\mathbf{h}}_l \\ \hat{\mathbf{h}}_l^H \phi_k & \hat{\mathbf{h}}_l^H \phi_k \hat{\mathbf{h}}_l - \sigma_k^2 - \mu_{kl}\varepsilon_l^2 \end{bmatrix} \succeq 0, \quad \forall k,\ l = k, k+1, \ldots, K, \tag{5.9b}$$

$$\mathbf{W}_k \succeq 0, \quad \forall k, \tag{5.9c}$$

where $\varphi_k = \mathbf{W}_k - \gamma_k^{\min} \sum_{m\neq k} \mathbf{W}_m$ and $\phi_k = \mathbf{W}_k - \gamma_k^{\min} \sum_{m=k+1}^{K} \mathbf{W}_m$.

Proof: Please refer to Appendix B.1. ■

The problem in (5.9) is a standard SDP and can be efficiently solved using interior-point methods. If the solution of the relaxed problem in (5.9) is a set of rank-one matrices $\mathbf{W}_k$, the optimal solution for the original problem in (5.6) can be obtained through extracting the eigenvector corresponding to the maximum eigenvalue of the rank-one solution of (5.9). Otherwise, the randomization technique can be used to generate a set of rank-one solutions.

5.3 Outage Probabilistic based Robust Design

In the probabilistic approach, the channel errors are random with a certain statistical distribution [115, 116]. In this model, the constraints assure that the desired SINR targets are delivered to the users under predefined outage probabilities. In this section, the outage probabilistic based approach is investigated by incorporating channel uncertainties in the beamforming design. In particular, the power minimization problem is considered with the outage probabilistic based target rate constraints. In this scheme, it is assumed that an imperfect estimate of channel covariance matrix is available at the BS and the required QoS should be satisfied with a predefined probability. Consider $\hat{\mathbf{C}}_k = \mathbb{E}(\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H) \in \mathbb{C}^{N\times N}$ indicate the estimated channel covariance matrix of U_k and the corresponding uncertainty matrix is denoted by $\mathbf{\Delta}_k \in \mathbb{C}^{N\times N}$. The $(ij)^{th}$ entries of $\mathbf{\Delta}_k$ are independently and identically distributed (i.i.d) as $[\mathbf{\Delta}_k]_{ij} \sim \mathcal{CN}(0, \sigma_{ij}^2)$. Hence, the actual channel covariance matrix can be modelled as

$$\mathbf{C}_k = \hat{\mathbf{C}}_k + \Delta_k, \qquad \forall k. \tag{5.10}$$

Based on the SIC approach, the SINR of the signal intended for U_k at the l^{th} user can be written as

$$\text{SINR}_l^{(k)} = \frac{\mathbf{w}_k^H(\hat{\mathbf{C}}_l + \Delta_l)\mathbf{w}_k}{\sum_{m=1}^{k-1} \mathbf{w}_m^H \Delta_l \mathbf{w}_m + \sum_{m=k+1}^{K} \mathbf{w}_m^H(\hat{\mathbf{C}}_l + \Delta_l)\mathbf{w}_m + \sigma_l^2}, \tag{5.11}$$

where the first term and the second term in denominator are, respectively, due to imperfect CSI at the receivers during the SIC process and the interference introduced by the signals intended to the stronger users. Hence, the power minimization problem can

be formulated as follows:

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1},\ \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{5.12a}$$

$$\text{subject to} \quad \min_{l\in\{k,k+1,\ldots,K\}} \text{SINR}_l^{(k)} \geq \gamma_k^{\min}, \qquad \forall k, \tag{5.12b}$$

where $\gamma_k^{\min} = 2^{R_k^{\min}} - 1$ is the minimum required SINR at the U_k and the constraint in (5.12b) can be simplified to $\text{SINR}_l^{(k)} \geq \gamma_k^{\min},\ \forall k,\ l = k, k+1, \ldots, K$.

In the outage probability-based design, the QoS constraint for given SINR targets $\gamma_k^{\min}$, is the probability that $\min\limits_{l\in\{k,k+1,\ldots,K\}} \text{SINR}_l^{(k)} \geq \gamma_k^{\min}$ should be greater than $1-\rho_k$, for a pre-specified probability of outage ρ_k. Therefore, given the uncertainty model in (5.10) and a distribution for $[\mathbf{\Delta}_k]_{ij}$, the problem of interest can be written as

$$\min_{\mathbf{w}_k \in \mathbb{C}^{N\times 1}, \forall k} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{5.13a}$$

$$\text{subject to} \quad \Pr\big(\text{SINR}_l^{(k)} \geq \gamma_k^{\min}\big) \geq (1-\rho_k), \quad \forall k,\ l = k, k+1, \ldots, K. \tag{5.13b}$$

The robust problem in (5.13) is NP-hard since the inclusion of CSI uncertainties in probabilistic constraints naturally lead to an infinite number of convex sets. To tackle this issue, the probabilistic constraint of the problems in (5.13b) is equivalently transformed into a tractable form. In order to realize the solution for the problem in (5.13), first, a new matrix variable $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ is introduced and the constraints in (5.13b) can be rewritten as follows:

$$\begin{aligned}
&\Pr\left(\frac{\text{Tr}\big(\mathbf{W}_k(\hat{\mathbf{C}}_l + \Delta_l)\big)}{\sum\limits_{m=1}^{k-1}\text{Tr}(\mathbf{W}_m^H\Delta_l) + \sum\limits_{m=k+1}^{K}\text{Tr}\big(\mathbf{W}_m^H(\hat{\mathbf{C}}_l+\Delta_l)\big) + \sigma_l^2} \geq \gamma_k^{\min}\right) \\
&= \Pr\Big(\text{Tr}(-\acute{\mathbf{B}}_k\Delta_l) \leq \text{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma_l^2\Big) \geq (1-\rho_k),\ \ \forall k, l = k, k+1, \ldots, K.
\end{aligned} \tag{5.14}$$

where $\acute{\mathbf{B}}_k = {\gamma_k^{\min}}^{-1}\mathbf{W}_k - \sum_{m\neq k}\mathbf{W}_m$ and $\mathbf{B}_k = {\gamma_k^{\min}}^{-1}\mathbf{W}_k - \sum_{m=k+1}^{K}\mathbf{W}_m$.

Hence, by defining a rank-one positive semidefinite matrix $\mathbf{W}_k$, the problem in (5.13) can be rewritten as

$$\min_{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k), \tag{5.15a}$$

$$\text{subject to} \ \ \Pr\Big(\mathrm{Tr}(-\acute{\mathbf{B}}_k \Delta_l) \leq \mathrm{Tr}(\mathbf{B}_k \hat{\mathbf{C}}_l) - \sigma_l^2\Big) \geq (1-\rho_k),\ \forall k,\ l = k, \ldots, K. \tag{5.15b}$$

Then, a procedure is utilized through the following Lemma to convert probabilistic constraints in (5.15b) into a tractable form to obtain the solution for the robust problem.

Lemma 4: The original robust optimization problem in (5.15) can be reformulated into the following problem:

$$\min_{\mathbf{W}_k \in \mathbb{C}^{N\times N},\ \forall k} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k), \tag{5.16a}$$

$$\text{subject to} \quad \mathbf{C}_{kl} = \begin{bmatrix} \frac{\Phi_{kl}}{\sqrt{2}\mathrm{erf}^{-1}(1-2\gamma_k)}\mathbf{I}_{M^2} & \mathrm{vec}(-\acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}) \\ \mathrm{vec}^H(-\acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}) & \frac{\Phi_{kl}}{\sqrt{2}\mathrm{erf}^{-1}(1-2\gamma_k)} \end{bmatrix} \succeq 0,$$

$$\forall k,\ l = k, k+1, \ldots, K, \tag{5.16b}$$

$$\mathbf{W}_k \succeq 0, \qquad \forall k, \tag{5.16c}$$

$$\mathrm{rank}(\mathbf{W}_k) = 1, \qquad \forall k. \tag{5.16d}$$

Proof: Please, refer to Appendix B.2. ■

The constraints in (5.16b) and (5.16c) are semidefinite in terms of $\mathbf{W}_k$. However, the problem is not convex due to the non-convex rank-one constraint in (5.16d). This non-convex problem can be relaxed into a convex one by exploiting SDR. Therefore, the problem in (5.16) without rank-one constraint is SDP and can be efficiently solved using interior-point methods. Then, the optimal solution for the original problem in (5.15) can be obtained through extracting the eigenvector corresponding to the maximum eigenvalue of the rank-one solution. If the solution of the relaxed problem in (5.16) is not a set of rank-one matrices, the randomization technique can be used to generate a set of rank-one solutions.

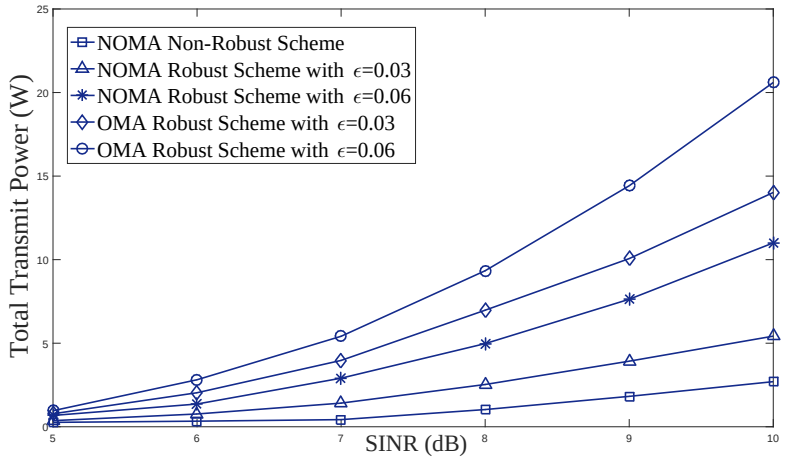


Fig. 5.2 Total transmit power versus different SINR thresholds for the robust and non-robust schemes with different channel estimation error bounds, ε.

5.4 Simulation Results

In this section, the simulation results are provided to evaluate the performance of the proposed robust designs in MISO NOMA systems through numerical simulations. To assess the performance of the proposed robust beamforming approach, the same setup as in Section 4.5 is considered where a multi-antenna BS serves single-antenna users which are uniformly distributed within the circle with a radius of 50 meters around the BS, but no closer than 1 meter. In addition, the channel coefficients between the BS and the users are generated as $\mathbf{h}_k = \chi_k \sqrt{d_k^{-\beta}}$ where $\chi_k \sim \mathcal{CN}(0, \mathbf{I})$, d_k is the distance between U_k and the BS, measured in meters and $\beta = 3.8$ is the path-loss exponent. The noise variance at each user is assumed to be 0.01 (i.e., $\sigma_k^2 = 0.01$). It should be noted that the term non-robust scheme refers to the scheme where the BS has imperfect CSI without any information on the channel uncertainties and the beamforming vectors are designed based on imperfect CSI without incorporating channel uncertainty information. In addition, for the worst-case design the error bounds are considered to be 0.06 (i.e., $\varepsilon_k = \varepsilon = 0.06$). For the outage probabilistic based design, the variance of each entry (i.e., $[\mathbf{\Delta}_k]_{ij}$) of the error covariance matrix $\mathbf{\Delta}_k$ and the predefined outage probability (ρ_k) of the required QoS constraints are set to 0.005 and 0.1, respectively. These numerical results are obtained by averaging over different 1000 random channels. For the simulations the Yalmip package [150] is used.

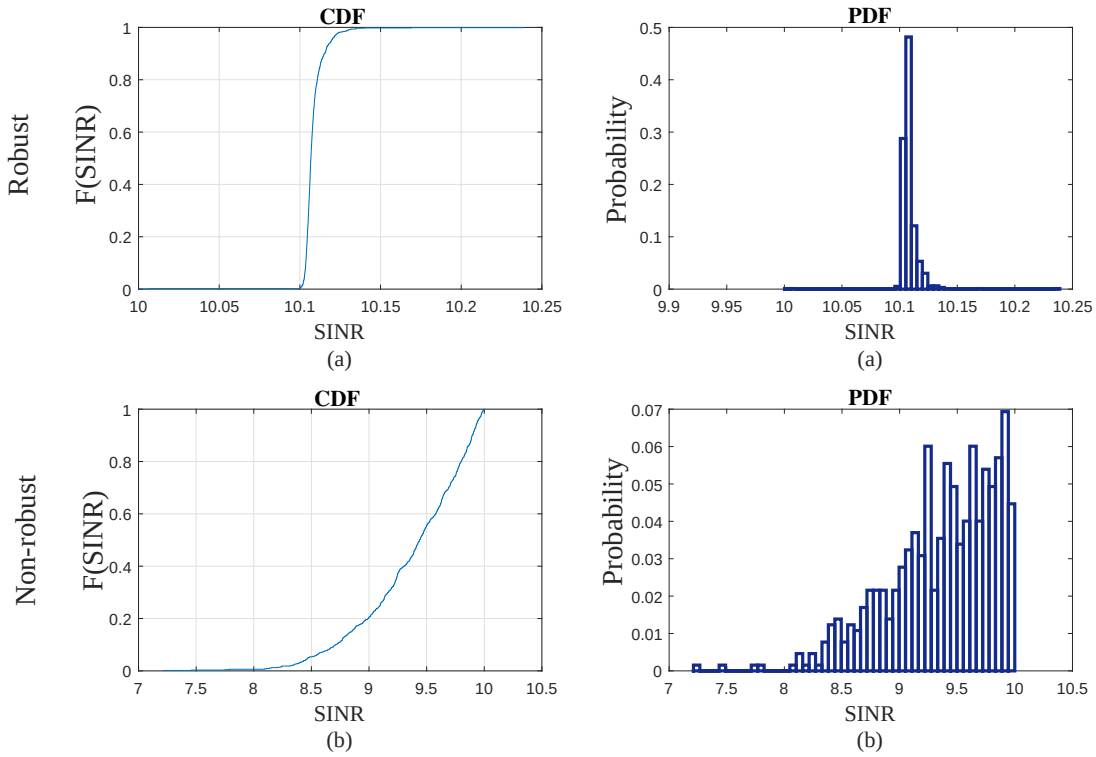


Fig. 5.3 Comparison CDF and PDF of minimum achieved SINR for (a) the robust scheme and (b) the non-robust scheme with $\varepsilon = 0.06$, $\gamma_k^{\min} = 10$ dB.

5.4.1 Performance Study of Worst Case Robust Design

In this subsection, the performance of the worst-case robust design is evaluated. First, the impact of channel uncertainties on the required total transmit power is studied. Fig. 5.2 depicts the required total transmit power against different SINR thresholds for the robust and the non-robust NOMA schemes as well as OMA scheme with different error bounds. As seen in Fig. 5.2, the robust scheme requires more transmit power than that of the non-robust scheme. This is because the robust scheme satisfies the required SINR all the time, at the price of more transmit power at the BS whereas the non-robust scheme does not. The difference between the required transmit power for the robust and the non-robust schemes increases with error bounds. This is because incorporating all possible sets of errors in the beamforming design to satisfy high SINR thresholds requires more transmit power in the robust scheme. Moreover, as seen in Fig. 5.2, the conventional framework OMA requires more transmit power to achieve the same rate in comparison with NOMA scheme.

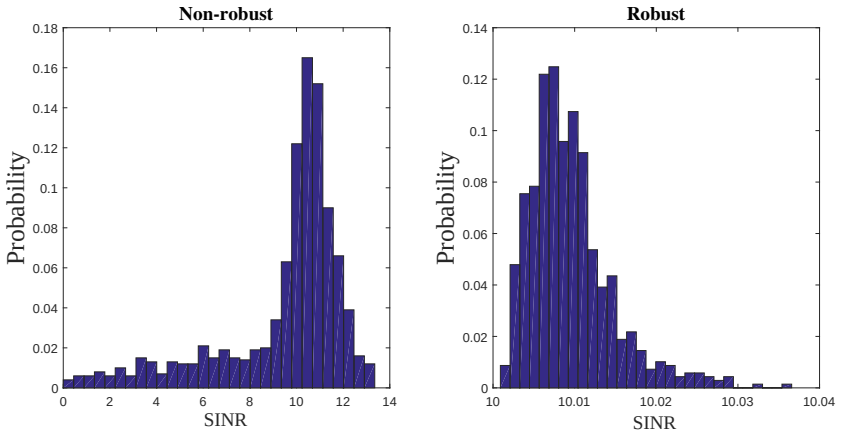


Fig. 5.4 Comparison the performance of the robust and non-robust schemes with equal transmit power.

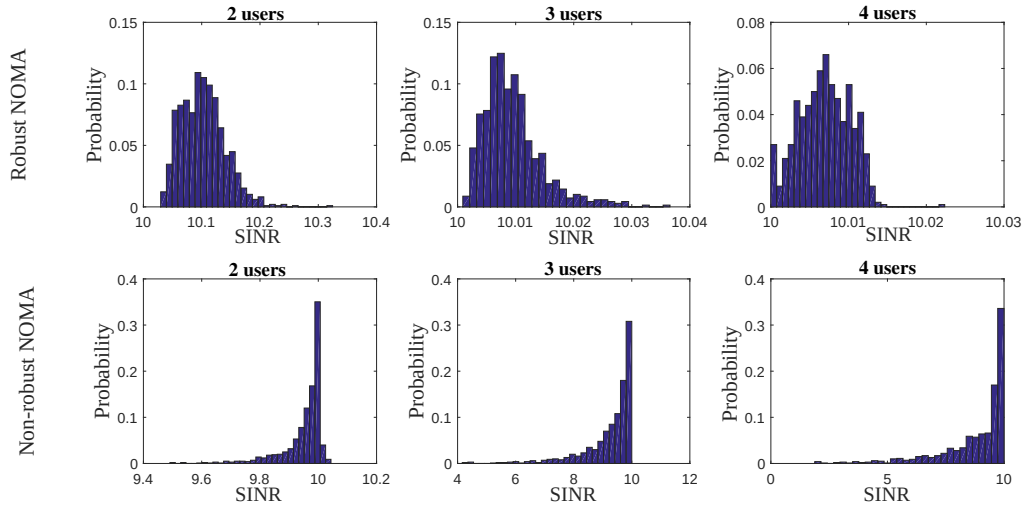
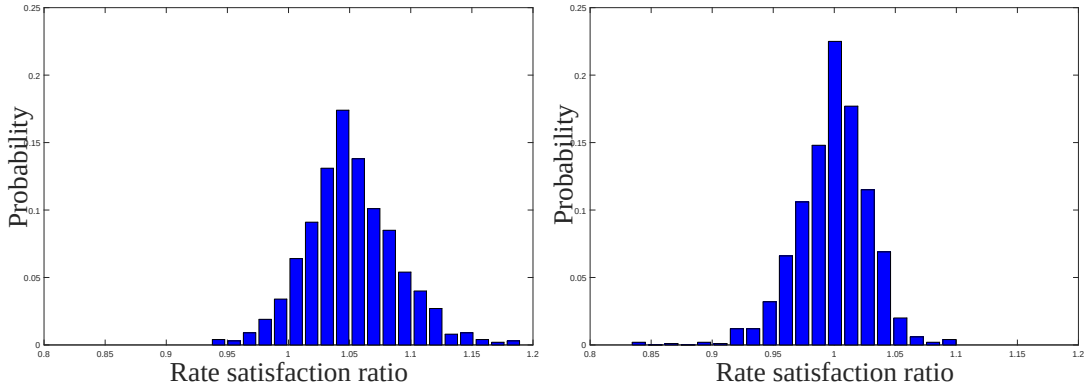


Fig. 5.5 Comparison probability of the achieved SINR for differner number of served users for robust and non-robust schemes.

Next, the performance of the proposed robust and non-robust schemes are evaluated in terms of the minimum achieved SINR between users. Fig. 5.3 provides cumulative distribution function (CDF) and probability density function (PDF) obtained from 1000 random sets of channels with error bounds of 0.06 ($\varepsilon = 0.06$) where the SINR threshold has been set to 10 dB at each user. As evidenced by the results, the robust scheme outperforms the non-robust scheme in terms of minimum achieved SINRs. In addition, the robust scheme satisfies the SINR thresholds all the time regardless of



(a) Robust scheme. (b) Non-robust scheme.

Fig. 5.6 Histogram for rate satisfaction ratio, i.e., η_k, for $R_k^{\min} = 3$bps/Hz.

the channel uncertainties whereas the non-robust design fails to satisfy the minimum SINR requirements.

For a fair comparison, the performance of the robust and the non-robust schemes is compared with equal power as depicted in Fig. 5.4. These simulation results have been generated by considering for 1000 different channels with 30 W transmit powers and 10 dB target SINR constraint. The robust scheme satisfies the minimum SINR requirements (10 dB) all the times regardless of the channel uncertainties. Although the non-robust scheme can achieve the required SINR for some set of channels, however, it fails to satisfy the required QoS all the times.

In order to demonstrate the impact of the number of users on the proposed robust scheme, the performance for different number of served users ie presented in Fig. 5.5. As seen in Fig. 5.5, the robust scheme satisfies the SINR requirement for different number of users all the times regardless of the channel uncertainties. However, the performance of non-robust scheme is degraded as the number of served users increases and fails to provide required SINR for more cases.

5.4.2 Performance Study of Outage Probabilistic based Robust Design

In this subsection, the impact of the outage probabilistic based robust design on the system performance is studied. The performance of the robust and the non-robust scheme is compared through the rate satisfaction ratio η_k, which is defined as the ratio between the achieved rate and the target rate at the user U_k. Hence, $\eta_k \geq 1$ indicates that the rate requirement is satisfied at the user U_k. Fig. 5.6 depicts the histogram of

the rate satisfaction ratio for the robust and the non-robust NOMA schemes with the target rate, $R_k^{\min} = 3$. It can be observed in Fig. 5.6(a) that in the robust design, rate constraint is satisfied in most cases and in only 10% of cases does the rate satisfaction ratio fall below one according to the outage probability requirement. However, as evidenced by result presented in Fig. 5.6(b) the non-robust design cannot satisfy the target rate requirement for approximately 50 percent of the cases since it does not take into account any information regarding channel uncertainties.

In Fig. 5.7, the negative affect on the required transmit power by the channel uncertainty is investigated. As seen, more transmit power is required at the BS as the error in the CSI increases. It means that the channel estimation error will degrade the performance of system in term of power consumption.

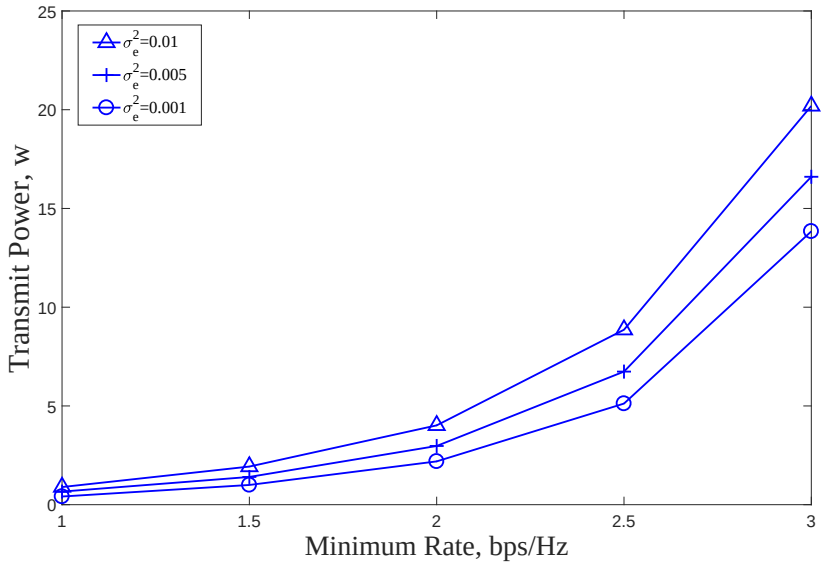


Fig. 5.7 The required total transmit power to achieve different target rates with different channel uncertainties at fix outages $\rho = 0.1$. Number of users = Number of antennas = 3.

5.5 Summary

In this chapter, the practical imperfect CSI scenario is investigated to design a robust beamforming for NOMA based downlink transmission. In the previous chapter, it is assumed that perfect CSI is available at the transmitter. However, in wireless transmissions, channel uncertainties are inevitable due to quantization and channel estimation errors. Particularly, due to ambiguities introduced in SIC through user decoding order

and superposition coding at the transmitter in NOMA, these uncertainties can greatly degrade the overall system performance. Therefore, to cultivate the desirable benefits offered by NOMA, these channel uncertainties should be accounted in the design of resource allocation techniques. Hence, two approaches, the worst-case design and outage probabilistic based design, are considered in this chapter to incorporate the channel uncertainties in the beamforming design for the NOMA-based systems. In the first approach, the beamformers are designed by incorporating norm-bounded channel uncertainties to provide the required QoS at each user. In terms of beamforming vectors, the original robust design is not convex and therefore, the robust beamformers cannot be obtained directly. To overcome this non-convex issue, the original intractable problem is reformulated into a convex problem, where the non-convex constraint is converted into the LMIs by exploiting S-Procedure. In the second approach, the total transmit power is minimized while satisfying these outage probabilities at each user. Through exploiting the SDR approach, the original non-convex problem is reformulated with a LMI form to obtain the solution for the original robust problem. Simulation results are provided to validate the performance of the proposed schemes in the terms of advantages of the robust design. These results confirm that the proposed robust schemes outperform the non-robust scheme in terms of the achieved rates and rate satisfaction ratio at each user. It offers a better performance than the non-robust approach by satisfying the SINR requirement at each user all the time regardless of associated channel uncertainties.

Chapter 6

Energy Efficient Design for MISO NOMA Systems with Imperfect CSI

6.1 Introduction

The exponential growth in volume of data traffic and the number of mobile devices in the next generation cellular networks has led to a rapid increase in energy consumption. This energy consumption can indirectly increase the amount of greenhouse gas emission levels. Beyond environmental contamination, the cost of high energy consumption imposes further financial pressure on the network operators. Hence, providing energy-efficient techniques is necessary to meet the high data-rate requirements in future wireless communication systems. As such, an appropriate performance metric is required to strike a good balance between the achievable data rate and power consumption. To this end, EE has been recently considered as one of the key performance metrics to evaluate the performance of communication networks [152–155]. In this chapter, EE-based resource allocation techniques are investigated for a MISO NOMA system with practical imperfect CSI at the transmitter. Two different system models are considered when designing the beamformers to maximize the worst-case EE. First, the bemformer design is presented for a general system model without any clustering. Then, the users are grouped in different clusters to reduce the complexity of SIC while employing NOMA to share resources between users in each cluster. To remove the interference between different clusters, two ZF techniques are applied in the beamforming design. The original designs are formulated as non-linear fractional programming and the Dinkelbach's algorithm is exploited to solve the corresponding EE maximization problem by converting the problem into a simple subtractive form.

Finally, the performance of the proposed beamforming designs is evaluated through numerical simulations.

6.2 Robust Energy Efficient MISO Transmission for Non-clustering NOMA Scheme

In this section, the energy-efficient design for a NOMA system without clustering is investigated when there are uncertainties in the CSI. The same system model in Chapter 5 is considered in this design where a BS, equipped with N transmit antennas, sends information to K single antenna users. Furthermore, the BS has imperfect CSI of the users. As recalled, it is assumed that the channel estimation error is bounded in a known region and the true CSI is modeled in (5.1)-(5.2), and the corresponding SINR is obtained in (5.5). For this downlink transmission, the worst-case robust design is investigated to maximize the EE while ensuring that each user enjoys a minimum required QoS. For energy-efficient communications, the aim of the design is to maximize the amount of transmitted data bits with a unit energy. Hence, the global EE is defined as the ratio between the achievable sum rate of the system and the total power consumption. Based on this definition, the EE of the NOMA system with worst-case design can be expressed as

$$\text{EE} = \frac{\sum_{k=1}^{K} \log_2 \left(1 + \min\limits_{l\in\{k,k+1,\ldots,K\}} \left(\min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \text{SINR}_l^{(k)} \right) \right)}{\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 + P_{\text{c}}}, \tag{6.1}$$

where $\|\mathbf{w}_k\|_2^2$ represents the transmit power assigned to U_k and P_c is additional circuit power consumption. The beamforming design to maximize the worst-case EE can be formulated under limited total transmit power and the QoS constraint for each user as

$$\max_{\mathbf{w}_k\in\mathbb{C}^{N\times 1},\ \forall k} \quad \text{EE} = \frac{\sum_{k=1}^{K} \log_2 \left(1 + \min\limits_{l\in\{k,k+1,\ldots,K\}} \left(\min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \text{SINR}_l^{(k)} \right) \right)}{\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 + P_{\text{c}}}, \tag{6.2a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \ \leq \ P^{\max}, \tag{6.2b}$$

$$\min_{l\in\{k,k+1,\ldots,K\}} \left(\min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \text{SINR}_l^{(k)} \right) \geq \gamma_k^{\min}, \ \ \forall k, \tag{6.2c}$$

where $P^{\max}$ is the maximum transmit power available at the BS, and $\gamma_k^{\min}$ is the minimum required SINR at each user. This problem cannot be solved directly and different steps are required to turn it into a tractable form. First, by introducing a variable $\gamma_k = \min\limits_{l\in\{k,k+1,\ldots,K\}} \left(\min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon_l} \text{SINR}_l^{(k)} \right)$, problem (6.2) can be equivalently reformulated as

$$\max_{\gamma_k,\mathbf{w}_k,\,\forall k} \quad \frac{\sum_{k=1}^{K}\log_2\left(1+\gamma_k\right)}{\sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2+P_{\text{c}}}, \tag{6.3a}$$

$$\text{subject to} \quad \sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 \leq P^{\max}, \tag{6.3b}$$

$$\gamma_k \geq \gamma^{\min}, \quad \forall k, \tag{6.3c}$$

$$\gamma_k \leq \min_{l\in\{k,k+1,\ldots,K\}} \left(\min_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon_l} \text{SINR}_l^{(k)} \right), \quad \forall k. \tag{6.3d}$$

Note that the constraint (6.3d) can be simplified as $\gamma_k \leq \min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon_l} \text{SINR}_l^{(k)}$, $\forall k,\ l = k, k+1, \ldots, K$. This optimization problem is non-convex in its original form and NP-hard due to the terms on the right-hand side of the constraint in (6.3d) and the fractional expression in objective function, thus, it is challenging to determine the optimal solution.

6.2.1 Low Complexity Beamforming Design

In order to overcome the complexity issues associated with the optimization problem (6.3), a suboptimal algorithm is developed based on the alternating optimization technique, and is derived in conjunction with the sequential convex programming. To this end, the beamforming vector associated with the k^{th} user is defined as

$$\mathbf{w}_k = \sqrt{p_k}\mathbf{v}_k, \tag{6.4}$$

where $p_k = \|\mathbf{w}_k\|_2^2$ and $\mathbf{v}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|_2}$ are power and direction associated to the k^{th} user's data, respectively. Hence, by substituting this expression of $\mathbf{w}_k$, the SINR can be updated as

$$\text{SINR}_l^{(k)} = \frac{p_k|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_k|^2}{\sum_{m=1}^{k-1}p_m|\Delta\hat{\mathbf{h}}_l^H\mathbf{v}_m|^2+\sum_{m=k+1}^{K}p_m|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_m|^2+\sigma_l^2}, \tag{6.5}$$

and the optimization problem (6.3) can be equivalently reformulated as

$$\max_{\gamma_k, p_k, \mathbf{v}_k, \forall k} \quad \frac{\sum_{k=1}^{K} \log_2\left(1+\gamma_k\right)}{\sum_{k=1}^{K} p_k \|\mathbf{v}_k\|_2^2 + P_{\mathrm{c}}}, \tag{6.6a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} p_k \|\mathbf{v}_k\|_2^2 \;\leq\; P^{\max}, \tag{6.6b}$$

$$\gamma_k \;\geq\; \gamma^{\min}, \quad \forall k, \tag{6.6c}$$

$$\gamma_k \;\leq\; \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \mathrm{SINR}_l^{(k)}, \;\; \forall k, \; l = k, k+1, \ldots, K, \tag{6.6d}$$

$$p_k \geq 0, \;\; \|\mathbf{v}_k\|_2^2 = 1, \quad \forall k. \tag{6.6e}$$

The basic idea of this algorithm is to solve the problem with respect to different subsets of variables while the others variables are fixed. In the proposed algorithm, the optimization problem is divided into two subproblems. In the first step, by keeping $\mathbf{v}_k$ fixed, the problem in (6.6) is solved with variables γ_k and p_k. In the next step, γ_k is considered as a fix variable and $p_k, \mathbf{v}_k$ are updated. As both subproblems are non-convex problem, some approximation techniques are applied to convert them to convex problems. The proposed suboptimal algorithm iteratively solves these two subproblems and updates the solution.

Subproblem 1

In this subsection, the first subproblem is considered where the problem is solved with variables γ_k and p_k while the directional vectors $\mathbf{v}_k$ are fixed. This subproblem can be defined as

$$\max_{\gamma_k, p_k, \forall k} \quad \frac{\sum_{k=1}^{K} \log_2\left(1+\gamma_k\right)}{\sum_{k=1}^{K} p_k + P_{\mathrm{c}}}, \tag{6.7a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} p_k \;\leq\; P^{\max}, \tag{6.7b}$$

$$\gamma_k \;\leq\; \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \mathrm{SINR}_l^{(k)}, \;\; \forall k, \; l = k, k+1, \ldots, K, \tag{6.7c}$$

$$\gamma_k \;\geq\; \gamma^{\min}, \;\; p_k \geq 0, \quad \forall k. \tag{6.7d}$$

Since there is a common variable $\Delta\hat{\mathbf{h}}_l$ in both numerator and denominator of the SINR expression, the constraints in (6.7c) are intractable. To tackle this issue, the

lower bound of $\min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon_l} \text{SINR}_l^{(k)}$ is considered to solve the problem through the following Lemma:

Lemma 5: The lower bound of $\min\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon_l} \text{SINR}_l^{(k)}$ can be expressed as

$$\varphi_l^k = \frac{p_k\, f_l^k}{\sum_{m=1}^{k-1}\, p_m\, \bar{g}_l^m + \sum_{m=k+1}^{K}\, p_m\, g_l^m + \sigma_l^2}, \tag{6.8}$$

where

$$f_l^k = \left| \left(\left| \hat{\mathbf{h}}_l^H \mathbf{v}_k \right| - \varepsilon_l \|\mathbf{v}_k\| \right)^+ \right|^2, \tag{6.9}$$

$$g_l^m = \left| \left| \hat{\mathbf{h}}_l^H \mathbf{v}_m \right| + \varepsilon_l \|\mathbf{v}_m\| \right|^2, \tag{6.10}$$

$$\bar{g}_l^m = \varepsilon_l^2\, \|\mathbf{v}_m\|^2. \tag{6.11}$$

Proof: Please refer to Appendix C.1. ■

Therefore, by using a lower bound function φ_l^k in (6.8), the original problem (6.7) can be rewritten as follows:

$$\max_{\gamma_k, p_k,\, \forall k} \quad \frac{\sum_{k=1}^{K} \log_2\left(1+\gamma_k\right)}{\sum_{k=1}^{K} p_k + P_{\text{c}}}, \tag{6.12a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} p_k \;\leq\; P^{\max}, \tag{6.12b}$$

$$\gamma_k \;\leq\; \frac{p_k\, f_l^k}{\sum_{m=1}^{k-1}\, p_m\, \bar{g}_l^m + \sum_{m=k+1}^{K}\, p_m\, g_l^m + \sigma_l^2}, \;\forall k,\; l = k, k+1, \ldots, K, \tag{6.12c}$$

$$\gamma_k \;\geq\; \gamma^{\min}, \;\; p_k \geq 0, \quad \forall k. \tag{6.12d}$$

The problem (6.12) is still a non-convex problem due to constrain (6.12c) and the fractional expression of the objective function. To deal with this fractional expression, the Dinkelbach's algorithm in Table 6.1 is employed which converts a nonlinear fractional optimization problem to an equivalent and a tractable problem. For more details of this algorithm, please refer to Appendix C.2.

Table 6.1 Dinkelbach's Algorithm

Algorithm 1 Dinkelbach's algorithm

1. **Initialization:** Set $\epsilon > 0, n = 0, \lambda_n = 0,$
2. **repeat**
3. $\mathbf{x}_n^* = \arg\max_{\mathbf{x}} \{f(\mathbf{x}_n) - \lambda_n g(\mathbf{x}_n)\},$
4. $F(\lambda_n) = f(\mathbf{x}_n^*) - \lambda_n g(\mathbf{x}_n^*),$
5. $\lambda_{n+1} = \dfrac{f(\mathbf{x}_n^*)}{g(\mathbf{x}_n^*)},$
6. $n = n + 1,$
7. **until** $F(\lambda_n) < \epsilon.$

To employ the Dinkelbach's algorithm, the problem (6.12) should be reformulated in the form of a concave-convex fractional problem (CCFP). Hence, to deal with non-convexity of constrain (6.12c), new variable $\nu_{l,k}$ is defined and the corresponding constraint is represented by the following inequalities:

$$\gamma_k \nu_{l,k} \leq p_k f_l^k, \ \ \forall k, \ l = k, k+1, \ldots, K, \tag{6.13a}$$

$$\sum_{m=1}^{k-1} p_m \, \bar{g}_l^m + \sum_{m=k+1}^{K} p_m \, g_l^m + \sigma_l^2 \leq \nu_{l,k}, \ \ \forall k, \ l = k, k+1, \ldots, K. \tag{6.13b}$$

Now, to deal with the product of two variables on the left hand-side of (6.13a), the following expression is utilized:

$$\gamma_k \nu_{l,k} = \frac{1}{4}\big[(\gamma_k + \nu_{l,k})^2 - (\gamma_k - \nu_{l,k})^2\big]. \tag{6.14}$$

Then, the second quadratic term in (6.14) can be approximated by a first order Taylor series around $\gamma_k^{(t)}$ and $\nu_{l,k}^{(t)}$. As such, the product of two variables can be transformed into a convex term as

$$\gamma_k \nu_{l,k} \approx \frac{1}{4}(\gamma_k + \nu_{l,k})^2 - \frac{1}{4}[(\gamma_k^{(t)} - \nu_{l,k}^{(t)})^2 + 2(\gamma_k^{(t)} - \nu_{l,k}^{(t)})(\gamma_k - \gamma_k^{(t)} - \nu_{l,k} + \nu_{l,k}^{(t)})]. \tag{6.15}$$

By recalling the above approximation, the following optimization problem should be solved in the t^{th} iteration:

$$\max_{\gamma_k, p_k, \nu_{l,k},\ \forall k,l} \quad \frac{\sum_{k=1}^{K} \log_2\left(1+\gamma_k\right)}{\sum_{k=1}^{K} p_k + P_{\mathrm{c}}}, \tag{6.16a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} p_k \ \leq \ P^{\max}, \tag{6.16b}$$

$$\frac{1}{4}\left((\gamma_k+\nu_{l,k})^2 - \left[(\gamma_k^{(t)}-\nu_{l,k}^{(t)})^2+2(\gamma_k^{(t)}-\nu_{l,k}^{(t)})(\gamma_k-\gamma_k^{(t)}-\nu_{l,k}+\nu_{l,k}^{(t)})\right]\right) \leq p_k f_l^k, \ \ \forall k,\ l=k,k+1,\ldots,K, \tag{6.16c}$$

$$\sum_{m=1}^{k-1} p_m\, \bar{g}_l^m + \sum_{m=k+1}^{K} p_m\, g_l^m + \sigma_l^2 \leq \nu_{l,k},\ \forall k,\ l=k,k+1,\ldots,K, \tag{6.16d}$$

$$\gamma_k \ \geq \ \gamma^{\min}, \ \ p_k \geq 0, \quad \forall k. \tag{6.16e}$$

In each iteration, the Dinkelbach's algorithm is employed to solve the fractional programming problem in (6.16) and the solution of problem is considered as $\gamma_k^{(t)}$ and $\nu_{l,k}^{(t)}$ in the next iteration. This procedure is repeated until it converges.

Subproblem 2

In this subproblem, the variables p_k and $\mathbf{v}_k$ needs to be updated for a given γ_k which is obtained by solving subproblem 1. By fixing γ_k, the numerator of the objective function remains constant and the denominator should be minimized. Hence, the following optimization problem for a given γ_k can be solved:

$$\min_{p_k, \mathbf{v}_k,\ \forall k} \quad \sum_{k=1}^{K} p_k \|\mathbf{v}_k\|_2^2, \tag{6.17a}$$

$$\text{subject to} \ \ \gamma_k \ \leq \ \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \mathrm{SINR}_l^{(k)}, \ \ \forall k,\ l=k,k+1,\ldots,K, \tag{6.17b}$$

$$p_k \geq 0,\ \|\mathbf{v}_k\|_2^2 = 1, \quad \forall k, \tag{6.17c}$$

$$\sum_{k=1}^{K} p_k \|\mathbf{v}_k\|_2^2 \ \leq \ P^{\max}. \tag{6.17d}$$

To solve this problem, first, the variables p_k and $\mathbf{v}_k$ are assigned back to $\mathbf{w}_k$ and the problem is reformulated as

$$\min_{\mathbf{w}_k,\ \forall k} \ \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{6.18a}$$

$$\text{subject to } \ \gamma_k \ \leq \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \text{SINR}_l^{(k)}, \ \ \forall k,\ l = k, k+1, \ldots, K, \tag{6.18b}$$

$$\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 \ \leq \ P^{\max}. \tag{6.18c}$$

where $\text{SINR}_l^{(k)}$ is defined in (5.5). By exploiting S-procedure provided in Appendix (B.1), the non-convex constraint (6.18b) is converted into LMI form and then SDR technique is applied through following Lemma:

Lemma 6: By introducing new variables $\mathbf{W}_k = \mathbf{w}_k\mathbf{w}_k^H$ such that $\text{Rank}(\mathbf{W}_k) = 1$, and applying S-procedure, the inequality constraints of problem are replaced with LMIs and the solution to the problem (6.18) can be determined by solving following relaxed problem:

$$\min_{\mathbf{W}_k,\mu_{kl},\ \forall k,l} \ \sum_{k=1}^{K} Tr(\mathbf{W}_k), \tag{6.19a}$$

$$\text{subject to } \begin{bmatrix} \mu_{kl}\mathbf{I} + \phi_k - \sum_{m=1}^{k-1}\mathbf{W}_m & \phi_k\hat{\mathbf{h}}_l \\ \hat{\mathbf{h}}_l^H\phi_k & \hat{\mathbf{h}}_l^H\phi_k\hat{\mathbf{h}}_l - \sigma_l^2 - \mu_{kl}\varepsilon_l^2 \end{bmatrix} \succeq \mathbf{0}, \quad \forall k,\ l = k, k+1, \ldots, K, \tag{6.19b}$$

$$\mu_{kl} \geq 0, \quad \forall k,\ l = k, k+1, \ldots, K, \tag{6.19c}$$

$$\mathbf{W}_k \succeq \mathbf{0}, \ \ \forall k, \tag{6.19d}$$

$$\sum_{k=1}^{K} Tr(\mathbf{W}_k) \leq P^{\max}, \tag{6.19e}$$

where $\phi_k = \dfrac{1}{\gamma_k}\mathbf{W}_k - \sum_{m=k+1}^{K}\mathbf{W}_m$.

Proof: Please refer to Appendix C.3. ■

The problem in (6.19) is a standard SDP and can be efficiently solved through convex optimization techniques. Note that the rank-one constraints on $\mathbf{W}_k$ in prob-

Table 6.2 Suboptimal algorithm for the worst-case problem (6.3)

Initialization:	Set iteration index $t = 0,\ \mathrm{EE}_0 = 0$, $\epsilon > 0$ and $\varepsilon_l > 0$, and initialize $\mathbf{v}_k^{(0)}$ and $p_k^{(0)}$ with a feasible set of points for problem (6.6).
Step 1	Solve Subproblem 1 with fix $\mathbf{v}_k = \mathbf{v}_k^{(t)}$: (1.1) Calculate $\gamma_k^{(t)}$ and $\nu_{l,k}^{(t)}$ as follows: $\gamma_k^{(t)} = \min_{l\in\{k,k+1,\ldots,K\}} \left(\frac{p_k^{(t)}\ f_l^k}{\sum_{m=1}^{k-1}\ p_m^{(t)}\ \bar{g}_l^m + \sum_{m=k+1}^{K}\ p_m^{(t)}\ g_l^m + \sigma_l^2} \right)$, $\nu_{l,k}^{(t)} = \sum_{m=1}^{k-1}\ p_m^{(t)}\ \bar{g}_l^m + \sum_{m=k+1}^{K}\ p_m^{(t)}\ g_l^m + \sigma_l^2$, where $f_l^k = \left\| \left(\left\| \hat{\mathbf{h}}_l^H \mathbf{v}_k \right\| - \varepsilon_l \|\mathbf{v}_k\| \right)^+ \right\|^2$, $g_l^m = \left\| \left\| \hat{\mathbf{h}}_l^H \mathbf{v}_m \right\| + \varepsilon_l \|\mathbf{v}_m\| \right\|^2$ and $\bar{g}_l^m = \varepsilon_l^2 \|\mathbf{v}_m\|^2$. (1.2) Solve problem (6.16) and obtain the solution $\{\gamma_k^*,\ p_k^*\}$. (1.3) if $\max_k \lvert\gamma_k^{(t)} - \gamma_k^*\rvert > \epsilon$, set $\gamma_k^{(t)} = \gamma_k^*, \forall k$ and repeat step (1.2). Otherwise, set $p_k^t = p_k^*$ and $\gamma_k^t = \gamma_k^*$ and go to Step 2.
Step 2	Stop If $\lvert\mathrm{EE}(\gamma_k^t,\ p_k^t) - \mathrm{EE}_0\rvert < \epsilon$ and set the suboptimal solution by $p_k^{(t)}, \gamma_k^{(t)}, \mathbf{v}_k^{(t)}$. Otherwise, set $\mathrm{EE}_0 = \mathrm{EE}(\gamma_k^t,\ p_k^t)$ and go to Step 3.
Step 3	Solve Subproblem 2: By fixing $\gamma_k = \gamma_k^{(t)}$, solve problem (6.19) and obtain $\mathbf{W}_k^*$. Obtain $\mathbf{w}_k^* = \sqrt{\lambda_k}\upsilon_k$, where λ_k is the dominant eigenvalue of $\mathbf{W}_k^*$ and υ_k is eigenvector that corresponds to λ_k.
Step 4	Update $p_k^{(t+1)} = \|\mathbf{w}_k^*\|_2^2$ and $\mathbf{v}_k^{(t+1)} = \frac{\mathbf{w}_k^*}{\|\mathbf{w}_k^*\|_2}$.
Step 5	Update t = t + 1 and go to Step 1.

lem (6.19) have been relaxed. In general, if the solution of the relaxed problem in (6.19) is a set of rank-one matrices $\mathbf{W}_k$, then it will be also the optimal solution to the original problem. Otherwise, different randomization techniques can be used to generate a set of rank-one solutions [115]. The beamforming vector $\mathbf{w}_k$ can be obtained from a rank-one $\mathbf{W}_k$ solution, as $\mathbf{w}_k = \sqrt{\lambda_k}\upsilon_k$ where λ_k and υ_k are the maximum eigenvalue and the corresponding eigenvector of $\mathbf{W}_k$, respectively. After solving the problem (6.19), p_k and $\mathbf{v}_k$ can be determined for problem (6.17) by using the relationship provided in (6.4). The proposed suboptimal algorithm for the worst-case problem (6.3) is summarized in Table 6.2. This algorithm will be terminated in Step 3 when the difference between the achieved EE values in two successive iterations is less than a predefined threshold. Note that problem (6.16) is solved by using Dinkelbach's algorithm and problem (6.19) is a standard convex SDP, hence, both problems can be solved efficiently by using interior point methods.

Note that before solving the problem in (6.3), it is important to check the feasibility of the original problem. It might be possible that the minimum SINR constraints in

(6.2c) might not be attainable in all users if the available total power at the BS is not sufficient. Hence, there exists a minimum required transmit power $P^{\min}$ which is sufficient to deliver the minimum data rate for each user and makes the problem in (6.3) feasible only under the condition $P^{\max} \geq P^{\min}$. Thus, it is important to determine a feasible range of $P^{\max}$ that should be able to meet the data rate requirements at each user. To obtain the feasible range of $P^{\max}$, an auxiliary optimization problem is formulated as

$$P^{\min} = \min_{\mathbf{w}_k,\ \forall k} \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2, \tag{6.20a}$$

$$\text{subject to} \min_{\|\Delta\hat{\mathbf{h}}_l\|_2 \leq \varepsilon_l} \text{SINR}_l^{(k)} \geq \gamma^{\min},\ \ \forall k,\ l = k, k+1, \ldots, K. \tag{6.20b}$$

This optimization problem can be solved by invoking the same technique in Lemma 6.

6.3 Robust Energy Efficient MISO Transmission for Clustering NOMA Scheme

One of the implementation issues of employing SIC at the receivers is the complexity introduced by scaling the number of users. A practical approach to reduce the complexity is grouping the users into different clusters with a small number of users. The users in each cluster are supported by NOMA scheme to share the same time-frequency block. Hence, in this section, the robust EE beamforming design is investigated for the clustering NOMA systems.

6.3.1 System Model

Consider a MISO NOMA downlink transmission where a BS, equipped with N transmit antennas, intends to communicate with $K = 2\acute{K}$ single antenna users as shown in Fig. 6.1. All users are grouped into $\acute{K}$ clusters $(\acute{K} \leq N)$ with two users per cluster by employing a clustering algorithm proposed in [77]. Note that the number of users in a cluster can be more than two, however, it is assumed that there are only two users in each cluster for the sake of brevity. The l^{th} user in the k^{th} cluster is denoted by $U_{l,k}$, for all $k \in \{1, \ldots, \acute{K}\}$ and $l = 1, 2$. For user pairing, the proposed clustering algorithm in [77] is applied which is based on the channel correlation, $\frac{|\mathbf{h}_i^T \mathbf{h}_j|}{\|\mathbf{h}_i\|\,\|\mathbf{h}_j\|}$, and gain difference,

$\big|\, \|\mathbf{h}_i\| - \|\mathbf{h}_j\| \,\big|$, between two users i and j. This clustering algorithm selects two users that have a high correlation and a large channel gain difference in each cluster.

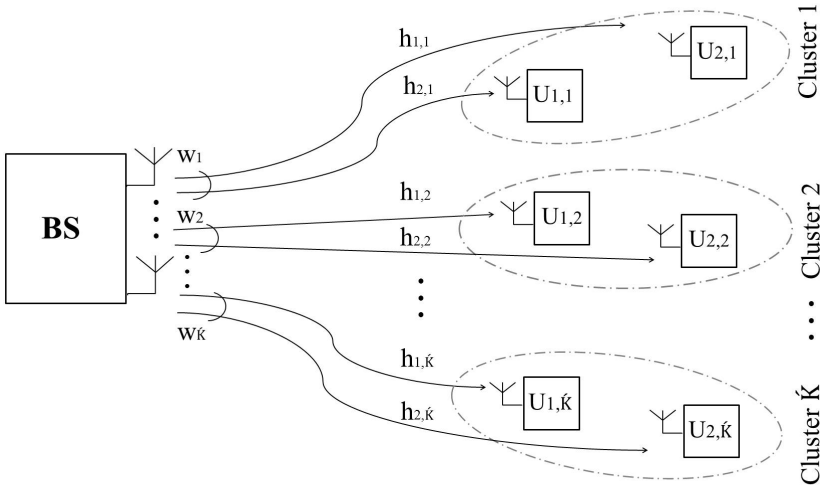


Fig. 6.1 A MISO NOMA system where imperfect CSI is available at the transmitter. One BS with N antennas serves $\acute{K}$ clusters with two users per cluster.

Let $\mathbf{h}_{l,k} \in \mathbb{C}^{N\times 1}$ represent the channel vector from the BS to $U_{l,k}$, which can be modeled as $\chi\sqrt{d_{l,k}^{-\beta}}$, where χ denotes the Rayleigh fading channel gain, $d_{l,k}$ is the distance between the BS and $U_{l,k}$, and β represents the path loss exponent. $U_{2,k}$ is considered as the strongest user which has a higher channel gain than that of $U_{1,k}$, such that $\|\mathbf{h}_{1,k}\| \leq \|\mathbf{h}_{2,k}\|,\ \forall k$ and $U_{1,k}$ is referred as the weakest user. The actual channel is modeled by the worst-case model in (5.1)-(5.2), and the norm-bounded channel uncertainties are incorporated in the analysis such that

$$\mathbf{h}_{l,k} = \hat{\mathbf{h}}_{l,k} + \Delta\hat{\mathbf{h}}_{l,k}, \tag{6.21}$$

where $\hat{\mathbf{h}}_{l,k}$ is the estimated channel of $U_{l,k}$, and $\Delta\hat{\mathbf{h}}_{l,k} \leq \varepsilon_{l,k}$ is the corresponding channel uncertainty. Let $\mathbf{w}_k$ and $p_{l,k}$ denote the beamforming vector steering towards the k^{th} cluster and the transmit power allocated to $U_{l,k}$, respectively. From the NOMA protocol, the BS broadcasts the superposition coded of users' signals as

$$\mathbf{x} = \sum_{k=1}^{\acute{K}} \mathbf{w}_k(\sqrt{p_{1,k}}\, s_{1,k} + \sqrt{p_{2,k}}\, s_{2,k}), \tag{6.22}$$

where $s_{1,k}$ and $s_{2,k}$ are the unit power information symbols for $U_{1,k}$ and $U_{2,k}$, respectively. Thus, the received signals at $U_{1,k}$ and $U_{2,k}$ are given by

$$y_{1,k} = \mathbf{h}_{1,k}^H \mathbf{x} + n_{1,k}, \tag{6.23}$$

$$y_{2,k} = \mathbf{h}_{2,k}^H \mathbf{x} + n_{2,k}, \tag{6.24}$$

where $n_{l,k} \sim \mathcal{CN}(0, \sigma_{l,k}^2)$ for $l = 1, 2$ is a zero-mean circularly symmetric AWGN with variance $\sigma_{l,k}^2$. By utilizing the SIC at the receivers, $U_{2,k}$ decodes and eliminates the data of $U_{1,k}$ from the aggregated received signal $y_{2,k}$, and then, decodes its own data.

To eliminate the interference between clusters, the ZF beamformer is utilized at the BS. To this end, the beamforming vector is designed based on the user's channel, $\mathbf{h}_{l,k}$, and fulfills the following condition:

$$\mathbf{h}_{l,m}^H \mathbf{w}_k = 0, \quad \forall m \neq k. \tag{6.25}$$

Note that when there are $\acute{K} \leq N < 2\acute{K} - 1$ antennas at the BS, it is not possible to simultaneously satisfy (6.25) for both channel vectors $\mathbf{h}_{i,m}$, and $\mathbf{h}_{2,m}$. Therefore, if it is assumed that the channel $\mathbf{h}_{i,m}$ is aligned with one of these users' channels, the other one will suffer from the interference caused by the other clusters. Consequently, this inter-cluster interference can severely degrade the performance of the SIC at the strong user to decode the weaker user's signal [77]. Therefore, to efficiently implement SIC, beamforming vectors are generated based on the channel of the strong users i.e., $\mathbf{h}_{i,m} = \mathbf{h}_{2,m}$ to satisfy the condition in (6.25) such that

$$\mathbf{h}_{2,m}^H \mathbf{w}_k = 0, \quad \forall m \neq k. \tag{6.26}$$

However, note that $\mathbf{h}_{1,m}^H \mathbf{w}_k \neq 0$, for any $m \neq k$, which is the source of inter-cluster interference. Since there is inter-cluster interference for $U_{1,k}$, this scheme is referred as a hybrid-ZF scheme. By defining the following matrix

$$\mathbf{H} = [\mathbf{h}_{2,1} \ \cdots \ \mathbf{h}_{2,\acute{K}}], \tag{6.27}$$

the beamforming vector can be obtained as

$$\mathbf{W} = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_{\acute{K}}] = \mathbf{H}^\dagger = \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1}, \tag{6.28}$$

where $\mathbf{H}^{\dagger}$ denotes the pseudo-inverse of the matrix $\mathbf{H}$ in (6.27), and $\mathbf{w}_k$ is the beamforming vector for the k^{th} cluster. Therefore, the received signal at $U_{2,k}$ can be written as

$$y_{2,k} = \mathbf{h}_{2,k}^{H}\mathbf{w}_k(\sqrt{p_{1,k}}\, s_{1,k} + \sqrt{p_{2,k}}\, s_{2,k}) + \Delta\hat{\mathbf{h}}_{2,k}^{H}\sum_{j\neq k}\mathbf{w}_j(\sqrt{p_{1,j}}\, s_{1,j} + \sqrt{p_{2,j}}\, s_{2,j}) + n_{2,k}, \tag{6.29}$$

where the second term in (6.29) refers to the residual inter-cluster interference which cannot be completely removed during the ZF process due to imperfect CSI. Overall, the SINR at the strong user to decode the weak user's signal is given by

$$\text{SINR}_{2,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{2,k}^{H}\mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{2,k}^{H}\mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j\neq k}|\Delta\hat{\mathbf{h}}_{2,k}^{H}\mathbf{w}_j|^2(p_{1,j}+p_{2,j})}_{\text{inter-cluster interference due to imperfect CSI}} + \sigma_{2,k}^2}, \tag{6.30}$$

and after removing the weak user's signal via SIC technique, the strong user achieves the following SINR to decode its own signal:

$$\text{SINR}_{2,k}^{(2)} = \frac{p_{2,k}|\mathbf{h}_{2,k}^{H}\mathbf{w}_k|^2}{\underbrace{p_{1,k}|\Delta\hat{\mathbf{h}}_{2,k}^{H}\mathbf{w}_k|^2}_{\text{intra-cluster interference due to imperfect CSI}} + \underbrace{\sum_{j\neq k}|\Delta\hat{\mathbf{h}}_{2,k}^{H}\mathbf{w}_j|^2(p_{1,j}+p_{2,j})}_{\text{inter-cluster interference due to imperfect CSI}} + \sigma_{2,k}^2}. \tag{6.31}$$

The first term of the denominator in (6.31) is considered due to the fact that the strong user cannot completely remove the weaker user's signal during the SIC process due to imperfect CSI. At the other end, the SINR of $U_{1,k}$ to decode its own signal is given by

$$\text{SINR}_{1,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{1,k}^{H}\mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{1,k}^{H}\mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j\neq k}|\mathbf{h}_{1,k}^{H}\mathbf{w}_j|^2(p_{1,j}+p_{2,j})}_{\text{intra-cluster interference}} + \sigma_{1,k}^2}. \tag{6.32}$$

Thus, the achievable rate at $U_{1,k}$ and $U_{2,k}$ can be respectively defined as follows:

$$R_{1,k} = \log_2 \left(1 + \min\{ \min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)}\}\right), \quad \forall k \tag{6.33}$$

$$R_{2,k} = \log_2(1 + \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}), \quad \forall k. \tag{6.34}$$

6.3.2 Robust EE Problem Formulation

To design an energy-efficient system, the worst-case global EE can be mathematically expressed as

$$\text{EE} = \frac{\sum_{k=1}^{\acute{K}} (R_{1,k} + R_{2,k})}{\sum_{k=1}^{\acute{K}} (p_{1,k} + p_{2,k}) + P_{\text{c}}}, \tag{6.35}$$

Furthermore, the optimization problem can be formulated to determine the transmit power allocation for maximizing the worst-case EE under limited power budget and the QoS constraint for each user as

$$\max_{p_{1,k}, p_{2,k}, \forall k} \text{EE}, \tag{6.36a}$$

$$\text{subject to} \quad \sum_{k=1}^{\acute{K}} (p_{1,k} + p_{2,k}) \ \leq \ P^{\max}, \tag{6.36b}$$

$$\min\{ \min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)}\} \ \geq \ \gamma^{\min}, \ \forall k, \tag{6.36c}$$

$$\min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)} \ \geq \ \gamma^{\min}, \ \forall k, \tag{6.36d}$$

This optimization problem is a non-convex in terms of power allocation variables and it is a non-linear fractional programming problem. To solve this EE maximization problem, an iterative approach is presented by using the Dinkelbach's algorithm to optimize an approximated convex problem.

The Proposed Solution

In this subsection, a power allocation scheme is proposed that maximizes the worst-case EE through an iterative algorithm. First, variables $\{\gamma_{1,k}, \gamma_{2,k}\} \in \mathbb{R}_+$ are intro-

duced to further simplify the optimization problem in (6.36) as follows:

$$\max_{\gamma_{1,k},\gamma_{2,k},p_{1,k},p_{2,k},\ \forall k} \quad \frac{\sum_{k=1}^{\acute{K}} \left(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\right)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k})+P_{\mathrm{c}}}, \tag{6.37a}$$

$$\text{subject to} \quad \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \leq P^{\max}, \tag{6.37b}$$

$$\gamma^{\min} \leq \gamma_{1,k} \leq \min\left\{\min_{\Delta\hat{\mathbf{h}}_{1,k}} \mathrm{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \mathrm{SINR}_{2,k}^{(1)}\right\},\ \forall k, \tag{6.37c}$$

$$\gamma^{\min} \leq \gamma_{2,k} \leq \min_{\Delta\hat{\mathbf{h}}_{2,k}} \mathrm{SINR}_{2,k}^{(2)},\ \forall k. \tag{6.37d}$$

The equivalent problem in (6.37) is still non-convex and NP-hard to determine a feasible solution. As there is a common parameter $\Delta\hat{\mathbf{h}}_{l,k}$ in the both numerator and the denominator of the SINR expression, the constraints in (6.37c) and (6.37d) are intractable. To circumvent this issue, the lower bound of SINR obtained in Lemma 5 is exploited. By applying the lower bound in (6.8) to the problem (6.37), the following optimization problem can be defined:

$$\max_{\gamma_{1,k},\gamma_{2,k},p_{1,k},p_{2,k}\ \forall k} \quad \frac{\sum_{k=1}^{\acute{K}} \left(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\right)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k})+P_{\mathrm{c}}}, \tag{6.38a}$$

$$\text{subject to} \quad \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \leq P^{\max}, \tag{6.38b}$$

$$\gamma_{1,k} \leq \frac{p_{1,k}\, f_{1,k}^{k}}{p_{2,k} g_{1,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) g_{1,k}^{m} + \sigma^2}, \quad \forall k, \tag{6.38c}$$

$$\gamma_{1,k} \leq \frac{p_{1,k}\, f_{2,k}^{k}}{p_{2,k} g_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^{m} + \sigma^2}, \quad \forall k, \tag{6.38d}$$

$$\gamma_{2,k} \leq \frac{p_{2,k}\, f_{2,k}^{k}}{p_{1,k} \overline{g}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^{m} + \sigma^2}, \quad \forall k, \tag{6.38e}$$

$$\gamma^{\min} \leq \gamma_{1,k},\ \ \gamma^{\min} \leq \gamma_{2,k}, \quad \forall k. \tag{6.38f}$$

To solve this fractional programming problem, the Dinkelbach's algorithm in Table 6.1 is employed to convert a nonlinear fractional optimization problem to an equivalent and a tractable problem. According to the requirement of Dinkelbach's algorithm, the problem in (6.38) should be reformulated in a CCFP form to apply this algorithm. To deal with the non-convex nature of constraints in (6.38c)-(6.38e), new variables

$\vartheta_{1,k}, \vartheta_{2,k}$ and ϑ_k are introduced and the corresponding constraints are redefined in the following inequalities:

$$(6.38c) \Rightarrow \begin{cases} \gamma_{1,k}\ \vartheta_{1,k} \leq p_{1,k}\ f_{1,k}^k, \\ p_{2,k} g_{1,k}^k + \sum_{m \neq k} (p_{1,m} + p_{2,m}) g_{1,k}^m + \sigma^2 \leq \vartheta_{1,k}, \end{cases} \quad \forall k, \qquad (6.39)$$

$$(6.38d) \Rightarrow \begin{cases} \gamma_{1,k}\ \vartheta_{2,k} \leq p_{1,k}\ f_{2,k}^k, \\ p_{2,k} g_{2,k}^k + \sum_{m \neq k} (p_{1,m} + p_{2,m}) \overline{g}_{2,k}^m + \sigma^2 \leq \vartheta_{2,k}, \end{cases} \quad \forall k, \qquad (6.40)$$

$$(6.38e) \Rightarrow \begin{cases} \gamma_{2,k}\ \vartheta_{k} \leq p_{2,k}\ f_{2,k}^k, \\ p_{1,k} \overline{g}_{2,k}^k + \sum_{m \neq k} (p_{1,m} + p_{2,m}) \overline{g}_{2,k}^m + \sigma^2 \leq \vartheta_{k}, \end{cases} \quad \forall k. \qquad (6.41)$$

Next, to deal with the product of optimization variables in (6.39)-(6.41), the expression in (6.14) is exploited, and then, the first order Taylor series approximation is applied to approximate the second quadratic term. As such, the product of two variables can be transformed into a linear convex term as

$$\begin{aligned} \gamma_{i,k}\ \vartheta_{j,k} &\approx \frac{1}{4} \left((\gamma_{i,k} + \vartheta_{j,k})^2 - \left[(\gamma_{i,k}^{(t)} - \vartheta_{j,k}^{(t)})^2 + 2(\gamma_{i,k}^{(t)} - \vartheta_{j,k}^{(t)})(\gamma_{i,k} - \gamma_{i,k}^{(t)} - \vartheta_{j,k} + \vartheta_{j,k}^{(t)}) \right] \right) \\ &\triangleq G(\gamma_{i,k}\ \vartheta_{j,k},\ \gamma_{i,k}^{(t)}\ \vartheta_{j,k}^{(t)}). \end{aligned} \qquad (6.42)$$

By recalling the above approximation, the following optimization problem in the t^{th} iteration should be solved:

Table 6.3 Energy Efficiency Maximization

1. Initialize $\mathbf{\Lambda^{(0)}}$ to a feasible value of (6.38), and set $t = 0$,
2. **repeat**
   Solve (6.43) by using Dinkelbach's algorithm,
   Set $\mathbf{\Lambda^{(t+1)}} = \mathbf{A}^*$,
   Update $t = t + 1$,
3. **until** required accuracy or maximum number of iterations.

$$\max_{p_{1,k},p_{2,k},\mathbf{A},\,\forall k} \frac{\sum_{k=1}^{\acute{K}} \big(\log_2(1+\gamma_{1,k}) + \log_2(1+\gamma_{2,k})\big)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) + P_{\mathrm{c}}}, \tag{6.43a}$$

$$\text{subject to} \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \le P^{\max}, \tag{6.43b}$$

$$G(\gamma_{1,k}\,\vartheta_{1,k},\,\gamma_{1,k}^{(t)}\,\vartheta_{1,k}^{(t)}) \le p_{1,k} f_{1,k}^{k},\quad \forall k, \tag{6.43c}$$

$$p_{2,k} g_{1,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) g_{1,k}^{m} + \sigma^2 \le \vartheta_{1,k},\quad \forall k, \tag{6.43d}$$

$$G(\gamma_{1,k}\,\vartheta_{2,k},\,\gamma_{1,k}^{(t)}\,\vartheta_{2,k}^{(t)}) \le p_{1,k} f_{2,k}^{k},\quad \forall k, \tag{6.43e}$$

$$p_{2,k} g_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^{m} + \sigma^2 \le \vartheta_{2,k},\quad \forall k, \tag{6.43f}$$

$$G(\gamma_{2,k}\,\vartheta_{k},\,\gamma_{2,k}^{(t)}\,\vartheta_{k}^{(t)}) \le p_{2,k} f_{2,k}^{k},\quad \forall k, \tag{6.43g}$$

$$p_{1,k} \overline{g}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^{m} + \sigma^2 \le \vartheta_{k},\quad \forall k, \tag{6.43h}$$

$$\gamma^{\min} \le \gamma_{1,k},\quad \gamma^{\min} \le \gamma_{2,k},\qquad \forall k, \tag{6.43i}$$

where $\mathbf{A} \triangleq \{\gamma_{1,k}, \gamma_{2,k}, \vartheta_{1,k}, \vartheta_{2,k}, \vartheta_{k}\}$. For notational simplicity, all parameters that have been used to linearize the quadratic terms in the the t^{th} iteration are defined as

$$\mathbf{\Lambda^{(t)}} \triangleq \{\gamma_{1,k}^{(t)},\ \gamma_{2,k}^{(t)},\ \vartheta_{1,k}^{(t)},\ \vartheta_{2,k}^{(t)},\ \vartheta_{k}^{(t)}\}. \tag{6.44}$$

The approximated problem in (6.43) should be iteratively solved for different values of $\mathbf{\Lambda}^{(t)}$ to obtain the best solution. Towards this end, if the solution of problem (6.43) in the t^{th} iteration is $\mathbf{A}^* \triangleq \{\gamma_{1,k}^*, \gamma_{2,k}^*, \vartheta_{1,k}^*, \vartheta_{2,k}^*, \vartheta_{k}^*\}$, it is considered as the initial point of the next iteration, i.e., $\mathbf{\Lambda}^{(t+1)}$, until it converges. The pseudo-code of the proposed iterative algorithm is summarized in Table (6.3). Furthermore, the minimum threshold to terminate the algorithm is chosen as the difference between two

successive values of achieved EE or the maximum number of iterations is reached to a predefined value.

Note that before solving the original problem in (6.36), it is important to obtain the feasible range of $P^{\max}$. To this end, an auxiliary optimization problem is formulated as

$$P^{\min} = \min_{p_{1,k},p_{2,k}\ \forall k} \sum_{k=1}^{\acute{K}} (p_{1,k} + p_{2,k}), \tag{6.45a}$$

$$\text{subject to} \quad \min \left\{ \min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)} \right\} \ \geq \ \gamma^{\min}, \ \forall k, \tag{6.45b}$$

$$\min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)} \ \geq \ \gamma^{\min}, \ \forall k. \tag{6.45c}$$

This optimization problem can be converted into a LP problem by invoking the same technique that used to solve main problem in (6.37).

6.3.3 Full-ZF beamforming scheme

In this section, full-ZF beamforming scheme is presented to completely mitigate the interference between clusters. In particular, it is assumed that the number of antennas deployed at the BS is $N \geq 2\acute{K} - 1$, which provides sufficient degrees of freedom for the ZF beamformer to completely remove the inter-cluster interference [156]:

$$\mathbf{h}_{i,m}^H \mathbf{w}_k = 0, \ \ \forall m \neq k, \ i = 1, 2. \tag{6.46}$$

To design the beamforming vector to satisfy condition in (6.46), the following matrix is first defined by including the other channel matrices:

$$\mathbf{H}_k = [\hat{\mathbf{H}}_1 \cdots \hat{\mathbf{H}}_{k-1} \ \hat{\mathbf{H}}_{k+1} \cdots \hat{\mathbf{H}}_{\acute{K}}], \tag{6.47}$$

where $\hat{\mathbf{H}}_k = [\mathbf{h}_{1,k} \ \mathbf{h}_{2,k}]$. Then, the null space of the matrix $\mathbf{H}_k$ in (6.47) can be utilized for the beamforming vector $\mathbf{w}_k$ which results in $\mathbf{H}_k^H \mathbf{w}_k = \mathbf{0}$. By exploiting this condition so-called full-ZF beamformer, the aggregated received signal at $U_{l,k}$ is

given by

$$\begin{aligned} y_{l,k} =& \mathbf{h}_{l,k}^H \mathbf{w}_k (\sqrt{p_{1,k}}\, s_{1,k} + \sqrt{p_{2,k}}\, s_{2,k}) \\ &+ \Delta\hat{\mathbf{h}}_{l,k}^H \sum_{j\neq k} \mathbf{w}_j (\sqrt{p_{1,j}}\, s_{1,j} + \sqrt{p_{2,j}}\, s_{2,j}) + n_{l,k}, \quad l = 1, 2, \end{aligned} \tag{6.48}$$

where the second term in (6.48) is considered due to imperfect CSI. Hence, the SINR at the weak user to decode its own signal can be defined as

$$\text{SINR}_{1,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j\neq k} |\Delta\hat{\mathbf{h}}_{l,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{inter-cluster interference due to imperfect CSI}} + \sigma_{1,k}^2}. \tag{6.49}$$

Similarly, the SINR at $U_{2,k}$ to decode weaker user's signal is given by

$$\text{SINR}_{2,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j\neq k} |\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{inter-cluster interference due to imperfect CSI}} + \sigma_{2,k}^2}, \tag{6.50}$$

and $U_{2,k}$ achieves the following SINR to decode its own message after performing SIC:

$$\text{SINR}_{2,k}^{(2)} = \frac{p_{2,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{1,k}|\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference due to imperfect CSI}} + \underbrace{\sum_{j\neq k} |\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{inter-cluster interference due to imperfect CSI}} + \sigma_{2,k}^2}. \tag{6.51}$$

Based on the definitions of SINRs, the worst-case EE can be expressed as

$$\text{EE} = \frac{\sum_{k=1}^{\acute{K}} \left(\log_2 \left(1 + \min\{ \min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)} \} \right) + \log_2 (1 + \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}) \right)}{\sum_{k=1}^{\acute{K}} (p_{1,k} + p_{2,k}) + P_{\text{c}}}. \tag{6.52}$$

Accordingly, the following optimization problem is solved to determine the best power allocation for maximizing the worst-case EE:

$$\max_{p_{1,k},p_{2,k}\ \forall k} \text{EE}, \tag{6.53a}$$

$$\text{subject to} \ \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \ \leq \ P^{\max}, \tag{6.53b}$$

$$\log_2\left(1+\min\left\{\min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)}\right\}\right) \geq R^{\min}, \ \forall k, \tag{6.53c}$$

$$\log_2\left(1+\min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}\right) \geq R^{\min}, \ \forall k. \tag{6.53d}$$

To solve the fractional programming problem in (6.53), the same procedure in the previous section is utilized. Towards this end, the problem in (6.53) is equivalently reformulated by introducing variables $\gamma_{1,k}$ and $\gamma_{2,k}$ as follows:

$$\max_{\gamma_{1,k},\gamma_{2,k},p_{1,k},p_{2,k}\ \forall k} \frac{\sum_{k=1}^{\acute{K}}\left(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\right)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k})+P_{\text{c}}}, \tag{6.54a}$$

$$\text{subject to} \ \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \leq P^{\max}, \tag{6.54b}$$

$$\gamma^{\min} \leq \gamma_{1,k} \leq \min\left\{\min_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)}\right\}, \ \forall k, \tag{6.54c}$$

$$\gamma^{\min} \leq \gamma_{2,k} \leq \min_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}, \ \forall k. \tag{6.54d}$$

where $\gamma^{\min}=2^{R^{\min}}-1$ is minimum required SINR for each user. By invoking Lemma 5, the following optimization problem can be defined

$$\max_{\gamma_{1,k},\gamma_{2,k},p_{1,k},p_{2,k}\ \forall k} \frac{\sum_{k=1}^{\acute{K}} \big(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\big)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k})+P_{\mathrm{c}}}, \tag{6.55a}$$

$$\text{subject to } \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \le P^{\max}, \tag{6.55b}$$

$$\gamma_{1,k} \le \frac{p_{1,k}\, f_{l,k}^{k}}{p_{2,k} g_{l,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{l,k}^{m} + \sigma_{l,k}^{2}}, \quad \forall k,\ l=1,2, \tag{6.55c}$$

$$\gamma_{2,k} \le \frac{p_{2,k}\, f_{2,k}^{k}}{p_{1,k} \overline{g}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^{m} + \sigma_{2,k}^{2}}, \quad \forall k, \tag{6.55d}$$

$$\gamma^{\min} \le \gamma_{1,k},\ \ \gamma^{\min} \le \gamma_{2,k}, \qquad \forall k. \tag{6.55e}$$

where

$$f_{l,k}^{k} = \left| \Big(\big| \hat{\mathbf{h}}_{l,k}^{H} \mathbf{w}_k \big| - \varepsilon_{l,k} \|\mathbf{w}_k\| \Big)^{+} \right|^2, \tag{6.56}$$

$$g_{l,k}^{k} = \left| \big| \hat{\mathbf{h}}_{l,k}^{H} \mathbf{w}_k \big| + \varepsilon_{l,k} \|\mathbf{w}_k\| \right|^2, \tag{6.57}$$

$$\overline{g}_{l,k}^{m} = \big(\varepsilon_{l,k} \|\mathbf{w}_m\|\big)^2. \tag{6.58}$$

Finally, the fractional programming problem in (6.55) can be solved by leveraging Dinkelbach's algorithm. According to the requirement of the Dinkelbach's algorithm, the problem should be reformulated in a CCFP form to apply this algorithm. To deal with the non-convex nature of constraints in (6.55c) and (6.55d), new variables $\vartheta_{1,k}, \vartheta_{2,k}$ and ϑ_k are introduced and the corresponding constraints are redefined in the following inequalities:

$$\text{(6.55c)} \Rightarrow \begin{cases} \gamma_{1,k}\, \vartheta_{l,k} \le p_{1,k}\, f_{l,k}^{k}, \\ p_{2,k} g_{l,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{l,k}^{m} + \sigma_{l,k}^{2} \le \vartheta_{l,k}, \end{cases} \quad \forall k,\ l=1,2, \tag{6.59}$$

and

$$\text{(6.55d)} \Rightarrow \begin{cases} \gamma_{2,k}\, \vartheta_{k} \le p_{2,k}\, f_{2,k}^{k}, \\ p_{1,k} \overline{g}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^{m} + \sigma_{2,k}^{2} \le \vartheta_{k}, \end{cases} \quad \forall k. \tag{6.60}$$

In order to deal with the product of optimization variables in (6.59) and (6.60), the expression in (6.14) and the first order Taylor series approximation around $\gamma_{i,k}^{(t)}$ and $\vartheta_{j,k}^{(t)}$ are used to transform them into convex linear terms. As such, the product of two variables can be transformed into a convex linear term as

$$\gamma_{i,k}\,\vartheta_{j,k} \approx \frac{1}{4}\left((\gamma_{i,k}+\vartheta_{j,k})^2 - \left[(\gamma_{i,k}^{(t)}-\vartheta_{j,k}^{(t)})^2 + 2(\gamma_{i,k}^{(t)}-\vartheta_{j,k}^{(t)})(\gamma_{i,k}-\gamma_{i,k}^{(t)}-\vartheta_{j,k}+\vartheta_{j,k}^{(t)})\right]\right)$$
$$\triangleq G(\gamma_{i,k}\,\vartheta_{j,k},\ \gamma_{i,k}^{(t)}\,\vartheta_{j,k}^{(t)}). \tag{6.61}$$

By recalling the above approximation, the following optimization problem in the t^{th} iteration is solved

$$\max_{p_{1,k},p_{2,k},\mathbf{A}\ \forall k} \quad \frac{\sum_{k=1}^{\acute{K}}\left(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\right)}{\sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k})+P_{\mathrm{c}}}, \tag{6.62a}$$

$$\text{subject to } \sum_{k=1}^{\acute{K}}(p_{1,k}+p_{2,k}) \le P^{\max}, \tag{6.62b}$$

$$G(\gamma_{1,k}\,\vartheta_{l,k},\ \gamma_{1,k}^{(t)}\,\vartheta_{l,k}^{(t)}) \le p_{1,k} f_{l,k}^{k},\ \ \forall k,\ l=1,2, \tag{6.62c}$$

$$p_{2,k} g_{l,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{l,k}^{m} + \sigma^2 \le \vartheta_{l,k},\ \ \forall k,\ l=1,2, \tag{6.62d}$$

$$G(\gamma_{2,k}\,\vartheta_{k},\ \gamma_{2,k}^{(t)}\,\vartheta_{k}^{(t)}) \le p_{2,k} f_{2,k}^{k},\ \ \forall k, \tag{6.62e}$$

$$p_{1,k}\overline{g}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^{m} + \sigma^2 \le \vartheta_{k},\ \ \forall k, \tag{6.62f}$$

$$\gamma^{\min} \le \gamma_{1,k},\ \ \gamma^{\min} \le \gamma_{2,k}, \qquad \forall k, \tag{6.62g}$$

where $\mathbf{A} \triangleq \{\gamma_{1,k}, \gamma_{2,k}, \vartheta_{1,k}, \vartheta_{2,k}, \vartheta_{k}\}$. For notational simplicity, $\mathbf{\Lambda^{(t)}}$ represents the points at which the quadratic terms have been linearized in t^{th} iteration as

$$\mathbf{\Lambda}^{(t)} \triangleq \{\gamma_{1,k}^{(t)},\ \gamma_{2,k}^{(t)},\ \vartheta_{1,k}^{(t)},\ \vartheta_{2,k}^{(t)},\ \vartheta_{k}^{(t)}\}. \tag{6.63}$$

Finally, the approximated problem in (6.62) is iteratively solved for different values of $\mathbf{\Lambda}^{(t)}$ and update the approximations to obtain the best local solution similar to the proposed iterative algorithm in Table (6.3).

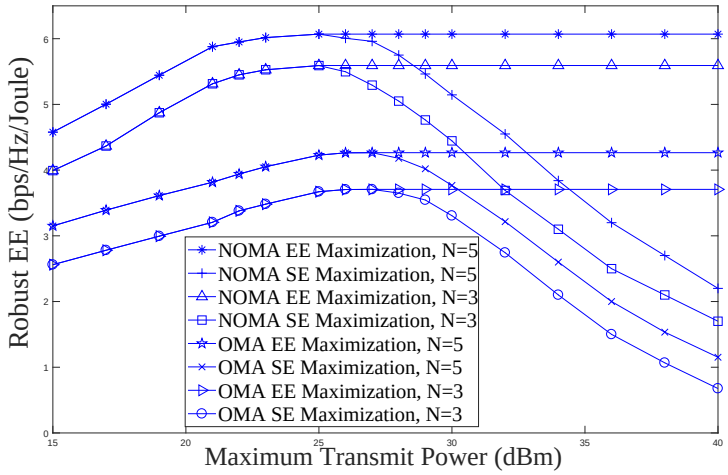


Fig. 6.2 Robust EE performance versus the maximum available power at the BS for $K = 4$ users by applying non-clustering approach. The error bound is set as $\varepsilon = 0.001$.

6.4 Simulation Results

In this section, the performance of the proposed robust EE design for both clustering and non-clustering MISO NOMA systems is evaluated by generating 1000 Monte-Carlo realizations of the flat fading channels. A downlink transmission is considered in a single cell with one BS equipped with N antennas. In the clustering approach, the users are clustered into $\acute{K}$ clusters with two single-antenna users per cluster. The small-scale fading of the channels is assumed to be Rayleigh fading which represents an isotropic scattering environment. The large-scale fading effect is modelled by ${d_{lk}}^{-\beta}$ to incorporate the path-loss effects where d_{lk} is the distance between $U_{l,k}$ and the BS, measured in meters and β is the path-loss exponent. Hence, the channel coefficients between the BS and user $U_{l,k}$ are generated using $\mathbf{h}_{l,k} = \chi\sqrt{{d_{lk}}^{-\beta}}$ where $\chi \sim \mathcal{CN}(0, \mathbf{I})$ and $\beta = 3.8$ [37]. Users are uniformly distributed within the circle with a radius of 50 meters around the BS, but no closer than 1 meter. Throughout the simulations, it is assumed that the users' locations are fixed and the average is taken over the small-scale fading of the propagation channels. In addition, it is assumed that the noise power is $\sigma_{l,k}^2 = \sigma^2 = 0.01$ at each receiver, and the minimum QoS requirement for all users is the same. For the simulations the Yalmip package [150] is used.

6.4.1 Performance Study of Non-Clustering Robust Design

In this subsection, the performance of the robust EE design is evaluated for non-clustering schemes. In Fig.6.2, the achievable robust EE is presented against maximum

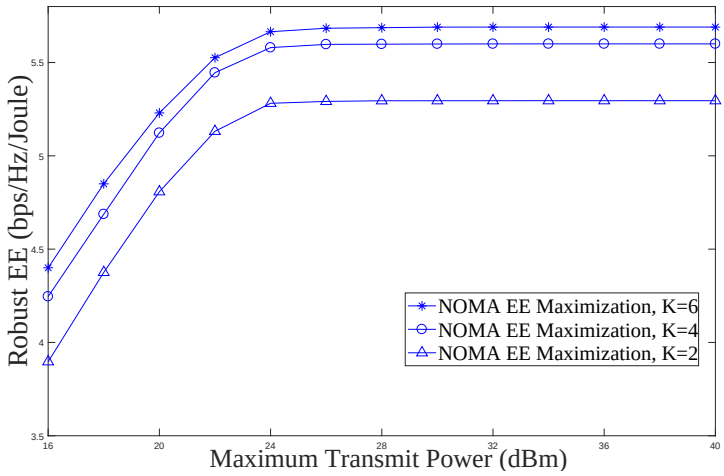


Fig. 6.3 Robust EE performance versus the maximum available power for different number of users with $N = 3$ antennas at the BS. The error bound is set as $\varepsilon = 0.001$.

available transmit power at the BS for both NOMA and conventional OMA schemes. The SE maximization refers to the achievable EE that obtained by maximizing the sum rate of the system. In other words, the sum rate maximization problem is solved and then the EE is determined based on the power allocation for the defined SE maximization problem. As shown in Fig. 6.2, the achievable EE reaches a maximum value with a certain available power (referred as green power in the literature) and then it remains constant for any available power which is more than the green power. Hence, one can conclude that just a portion of the power budget contributes to achieve the maximum EE, and using more power will deteriorate the performance of the system in terms of the EE which is the case in the SE maximization based design. As it is shown, more EE can be achieved by increasing the number of antennas. From Fig. 6.2, it is observed that the performance of NOMA system is much better than that of the conventional OMA scheme. In fact, NOMA can provide higher data rate by simultaneously allocating the radio resources to more than one user.

In Fig. 6.3, the performance of EE is evaluated versus the maximum available power at BS for different number of users K. In the simulation, the error bound is set as $\varepsilon = 0.001$. It is shown that the EE increases when the number of the users increases. As the number of users increases, the EE continues to increase, but the rate of growth becomes slower, as expected from the rate formulation in calculating the EE. In fact, by increasing the number of users, the interference from stronger users is enhanced, hence, the rate increment is less than power increment which degrades the performance of the overall system in terms of achievable EE.

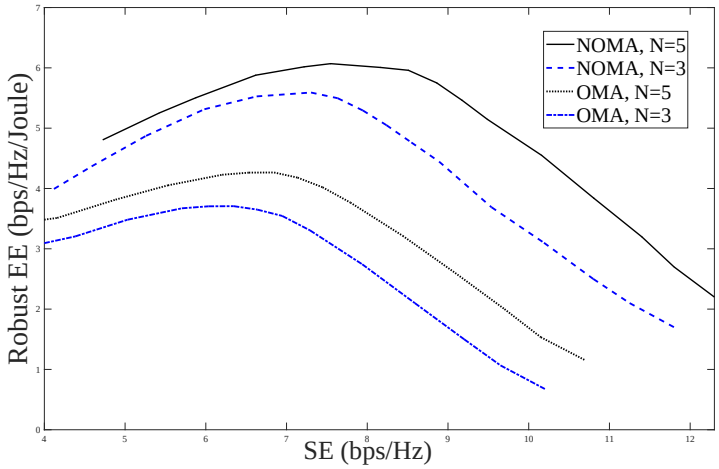


Fig. 6.4 The EE-SE tradeoff for $K = 4$ users with different number of antennas at BS. The error bound is set as $\varepsilon = 0.001$.

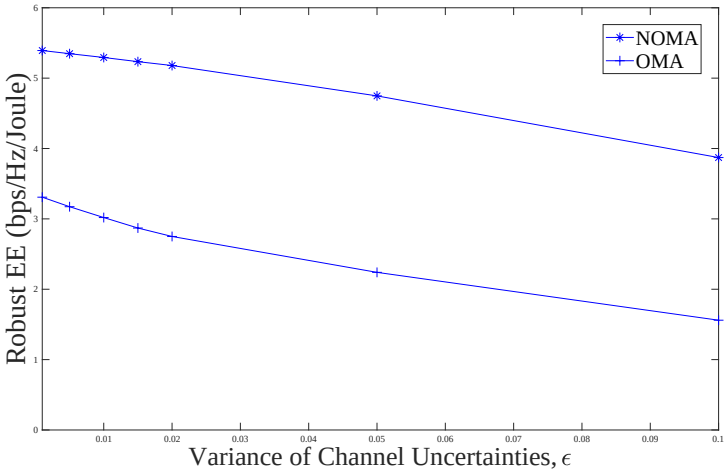


Fig. 6.5 Robust EE performance with different variance of channel uncertainty in NOMA and OMA schemes. System parameters are $K = 4$ users, $N = 3$ antennas.

Next, the EE-SE tradeoff of the proposed NOMA scheme and the traditional OMA scheme is presented in Fig. 6.4. As shown, the NOMA scheme significantly outperforms the OMA scheme in terms of both SE and EE. In Fig. 6.5, the impact of different channel uncertainties on the achieved EE is represented. It can be observed from Fig. 6.5 that the EE decreases for both schemes as the variances of the channel uncertainty in the CSI increase which requires more power consumption to satisfy the worst-case SINR at all users.

The impact of the proposed robust design on the achievable EE and rate is demonstrated by comparing with the performance of the non-robust scheme. The achieved EE of the robust and the non-robust designs are depicted in Fig. 6.6 for different available

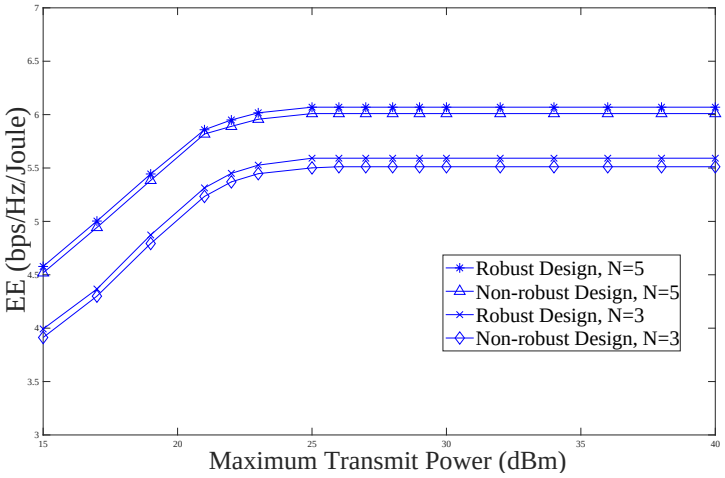


Fig. 6.6 EE performance versus maximum available power for the robust and the non-robust schemes with $K = 4$ users and channel estimation error bound $\varepsilon = 0.001$.

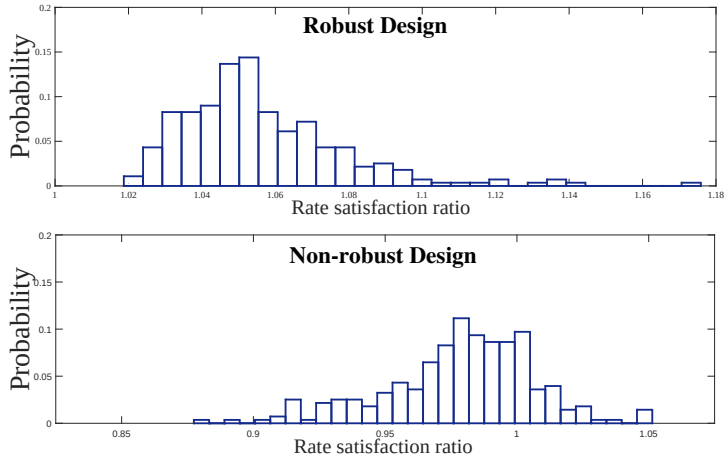


Fig. 6.7 Histogram for rate satisfaction ratio in the robust and non-robust NOMA scheme with channel estimation error bound $\varepsilon = 0.001$.

transmit power at the BS. As shown, the results of the robust and non-robust schemes are almost identical for $\varepsilon = 0.001$. To have a better comparison, the performance of the robust and the non-robust schemes is compared in term of rate satisfaction ratio, which is defined as the ratio between the achieved rate and the target rate at each user. Hence, rate satisfaction ratio greater than 1 indicates that the rate requirement is satisfied at each user. Fig. 6.7 depicts the histogram of the rate satisfaction ratio for the robust and the non-robust schemes. The simulation result implies that the rate constraint in the robust design is satisfied all the times regardless of the channel uncertainties. However, the non-robust design cannot satisfy the target rate requirement for more cases since it does not take the channel uncertainties into account.

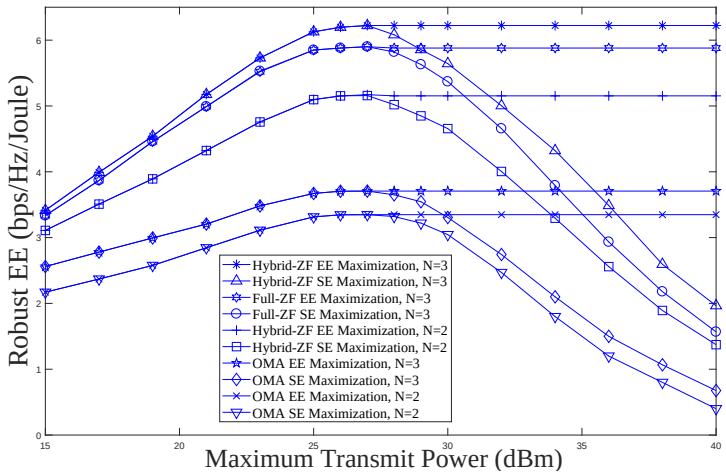


Fig. 6.8 Robust EE performance versus the maximum available power at the BS for $\acute{K} = 2$ clusters in full-ZF and hybrid-ZF schemes as well as OMA scheme. The error bound is set as $\varepsilon = 0.001$.

6.4.2 Performance Study of Clustering Robust Design

In this subsection, the performance of the robust EE design is evaluated for clustering scheme. The achievable robust EE against maximum available transmit power at the BS is presented in Fig. 6.8 for both full-ZF and hybrid-ZF schemes as well as the conventional OMA scheme. In this figure, the EE maximization represents the solution to the original optimization problem in (6.43) and (6.62). As shown in Fig. 6.8, both methods i.e., EE maximization and SE maximization, have a same performance in terms of the EE in the low transmit power and the EE increases with the transmit power. However, as the transmit power reaches a certain point, further increase in the transmit power does not yield a higher EE. After that point, the EE remains constant in the EE maximization method while the EE decreases in the SE maximization method. . For a given transmit power and with minimum required transmit antennas in each scheme (i.e. 2 antennas in hybrid-ZF scheme and 3 antennas in full-ZF scheme), the full-ZF can achieve more EE than that of the hybrid-ZF scheme. In fact, full-ZF scheme can provide higher data rate by completely removing other clusters interference at the cost of more required transmit antennas at the BS.

To draw a fair comparison, it is assumed that the equal number of transmit antennas is employed for both schemes. As seen in Fig. 6.9, the hybrid-ZF scheme outperforms the full-ZF in terms of achieved EE when there are a few clusters. This is due to the fact that the full-ZF requires more transmit power to completely remove the inter-cluster interference, while this type of interference has less impact in the systems with

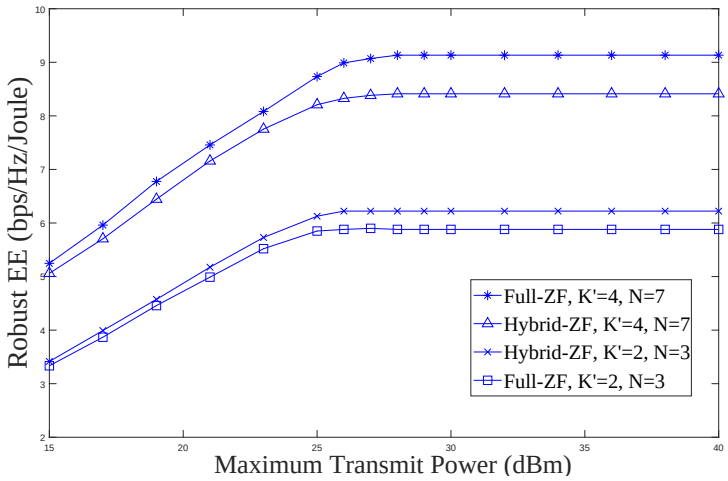


Fig. 6.9 Robust EE performance versus the maximum available power in full-ZF and hybrid-ZF schemes with the same number of transmit antennas at the BS. The error bound is set as $\varepsilon = 0.001$.

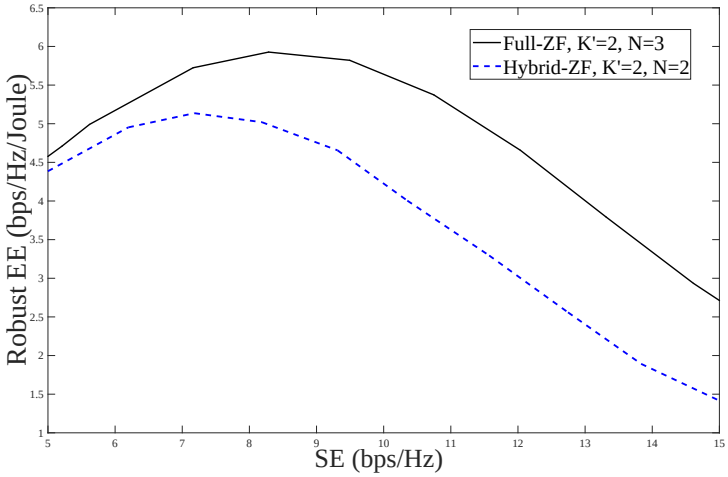


Fig. 6.10 The EE-SE tradeoff for full-ZF and hybrid-ZF schemes. System parameters are $\acute{K} = 2$ clusters, error bound $\varepsilon = 0.001$.

a few clusters. In other words, the rate improvement in full-ZF is not as much as the required power which degrades the system performance in terms of EE. However, by increasing the number of clusters, the full-ZF scheme outperforms the hybrid-ZF scheme because the inter-cluster interference increases and has a significant impact on the over all performance of the system.

Next the tradeoff between the SE and EE of the proposed schemes is evaluated. Fig. 6.10 depicts the EE-SE tradeoff of the both full ZF and hybrid ZF schemes. As shown in Fig. 6.10, both the SE and EE increase up to a maximum level which is known as the best tradeoff point, and then EE decreases while increasing SE. Beyond

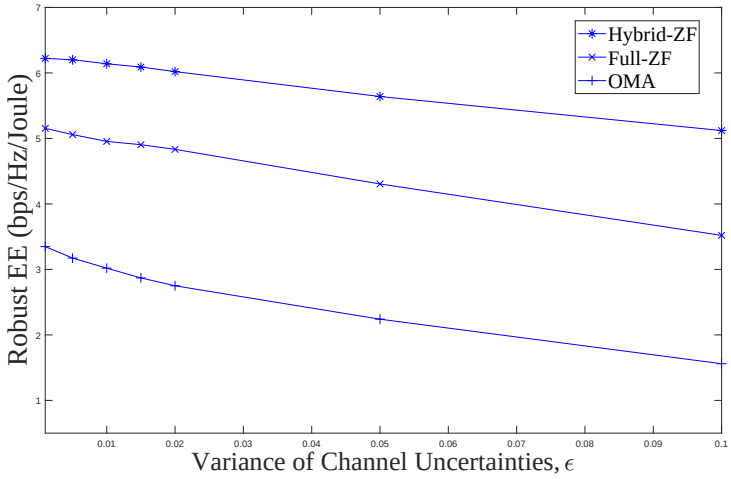


Fig. 6.11 Robust EE performance with different variance of channel uncertainty in full-ZF and hybrid-ZF schemes. System parameters are $\acute{K} = 2$ clusters, $N = 3$ antennas.

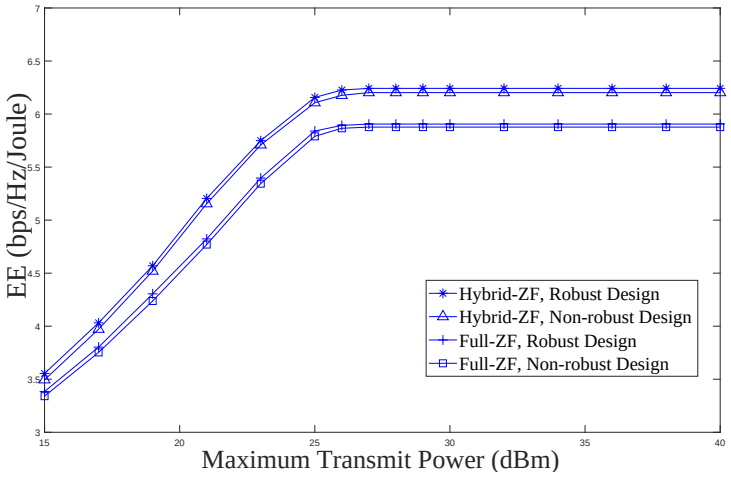


Fig. 6.12 EE performance versus maximum available power for the robust and the non-robust schemes with channel estimation error bound $\varepsilon = 0.001$. System parameters are $\acute{K} = 2$ clusters, $N = 3$ antennas.

this best tradeoff point, the EE should be sacrificed to achieve higher SE for which the BS requires more transmit power. On the other hand, the impact of different channel uncertainties on the achieved EE is represented in Fig. 6.11. It can be observed from Fig. 6.11 that the EE decreases for both schemes as the variances of the channel uncertainty in the CSI increase.

Next, the impact of the proposed robust design on the achievable EE and rate is demonstrated by comparing with the performance of the non-robust scheme. The achieved EE of the robust and the non-robust designs are depicted in Fig. 6.12 for different available transmit power at the BS. As seen in Fig. 6.12, both schemes have

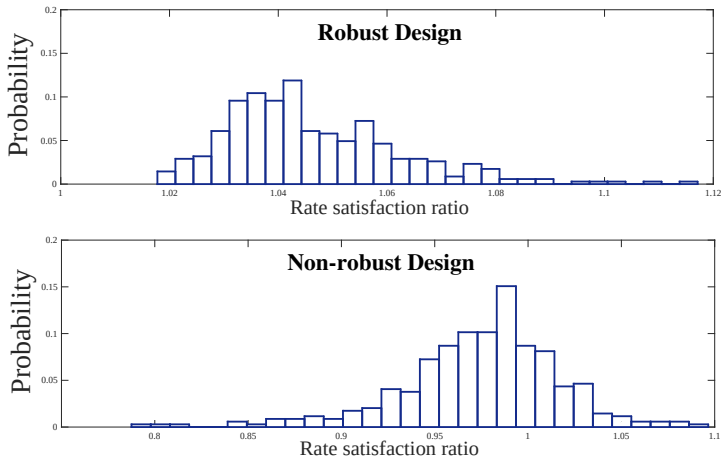


Fig. 6.13 Histogram for rate satisfaction ratio in the robust and non-robust NOMA scheme with channel estimation error bound $\varepsilon = 0.001$.

almost the identical performance in the terms of the achieved EE. For a better comparison, the histogram of the rate satisfaction ratio is depicted in Fig. 6.13 for the robust and the non-robust schemes. As evidenced by the results, the robust scheme outperforms the non-robust scheme and satisfies the rate requirement all the time regardless of the channel uncertainties whereas the non-robust design fails to satisfy the rate requirement.

6.5 Summary

With the immense increase of the traffic data and mobile devices, the energy consumption increased and become an important issue in the green cellular network. Hence, EE has become a key performance metric in the development of 5G and beyond wireless networks. To this end, an EE design is investigated in this chapter for NOMA systems where only imperfect CSI is available at the BS. In general, there are two schemes available in the literature for NOMA, non-clustering scheme and clustering scheme. In the first scheme, there is no cluster and NOMA is employed to share the radio resources between all users. In this scheme, each user is supported by its own beamforming vector. However, in the clustering scheme the users in a cell are grouped into different clusters to decrease the complexity of SIC at receivers as the users in the same cluster are supported by the NOMA scheme. For beamforming design in clustering scheme, the ZF is employed to mitigate the inter-cluster interference. In particular, two different ZF schemes are proposed, namely, hybrid-ZF and full-ZF based on the

number of users and available transmit antennas at the BS. The objective function that defines the EE of the system is a non-convex and a nonlinear function which formulates the original problem into a fractional programming. The Dinkelbach's algorithm is employed to convert the non-linear fractional programming problem into a simple subtractive form. As simulation results demonstrate, the complexity of non-clustering scheme increases by the number of users and more transmit power is required to satisfy the minimum QoS at each user which decreases the EE. In addition, in clustering scheme, despite the full-ZF scheme can completely remove the interference between different clusters, it requires more transmit antennas than that of the hybrid-ZF scheme to serve the same number of users. However, by increasing the number of clusters, the inter-cluster interference increases, and consequently, the full-ZF approach shows a better performance in terms of EE. In addition, the results confirmed that the proposed robust approaches outperform the non-robust scheme in terms of the rate satisfaction ratio at each user.

Chapter 7

Conclusion and Future Work

NOMA is envisioned as one of the key enabling techniques for future wireless communications to overcome a major problem of OMA techniques which do not allow frequency reuse within one cell. In this thesis, the NOMA scheme is investigated, which is expected to increase system throughput and support massive connectivity in future wireless networks. As NOMA utilizes the power domain to share the same wireless resources between multiple users, the overall performance in NOMA is very dependent on how efficiently power allocated between the users. Hence, resource allocation plays a crucial role in designing the NOMA based systems. As such, this thesis explored different resource allocation techniques to design beamforing vectors for MISO-NOMA systems.

The beamforming vectors are designed based on power minimization and max-min fairness in terms of data-rate among users in Chapter 4 with the assumption of perfect CSI at the transmitter. To circumvent the non-convexity issues of the optimization problems and to solve it, CCP is exploited with the first order Taylor series approximation and SDR approach to determine solutions for the original optimization problems. The performance of proposed approaches are evaluated in terms of power consumption and computational complexity while comparing the performance with the conventional OMA scheme. Simulation results demonstrated that the NOMA scheme outperforms the conventional multiple access in terms of transmit power consumption. Moreover, two proposed schemes based on Taylor series approximation and SDR show a similar performance with a few users in the system. However, by increasing the number of users, the performance gap between these two schemes increases and SDR outperforms the Taylor series approximation scheme in terms of required transmit power. The reason for this performance difference is that SDR can provide the optimal solu-

tion given that the solution is rank one whereas the other scheme relies on the Taylor series approximation which might lead to a suboptimal solution.

Then, practical assumption is taken into account in Chapter 5 by introducing the concept of robust beamforming design. Two different approaches are considered to model channel uncertainties and the corresponding performance analyses are also presented. In the first approach, norm-bounded channel uncertainties are assumed to study the robust power minimization problem. This robust scheme is developed based on the worst-case performance optimization framework. In the second approach, an outage probabilistic based robust scheme is investigated by incorporating channel uncertainties where the total transmit power is minimized while satisfying the outage probabilities at each user. In terms of beamforming vectors, the original robust designs are non-convex and they are reformulated into a convex problem by exploiting the S-procedure and SDR approach. Simulation results demonstrated that NOMA can satisfy fairness requirements through appropriate power allocation. In addition, these results confirm that the proposed robust scheme offers a better performance than that of the non-robust approach by satisfying the target rate requirements at each user all the time regardless of associated channel uncertainties.

With the high demand for energy to meet the unprecedented requirements in future wireless networks, EE has aroused wide interest and has become a major concern in the green cellular network. Hence, Chapter 6 investigates robust EE design in MISO NOMA system by leveraging the norm-bounded channel uncertainty model. First, the EE beamformers are designed for general system model where there is no clustering and each user is supported by its own NOMA based beamforming vector. Then, to reduce the complexity of SIC at the receivers, clustering is applied to group the users into different clusters with a small number of users. In the clustering NOMA scheme, users in each cluster are supported by a NOMA based beamforming approach and beamforming vectors are designed to support each cluster through conventional multiuser beamforming designs. To remove the interference between different clusters, two different types of ZF technique, namely hybrid-ZF and full-ZF, are applied in clustering scheme. In the first approach, i.e., hybrid-ZF, the inter-cluster interference is partially canceled while the second scheme, i.e., full-ZF, completely removes the inter-cluster interference at the cost of more antennas at the transmitter. To transform the original robust problems into a tractable form, the lower bound is considered for the worst-case SINR and an iterative scheme is developed by exploiting Dinkelbach's algorithm to solve the problem. The numerical results confirm that the proposed robust schemes outperform the non-robust scheme in terms of the rate satisfaction ratio

at each user. In addition, the simulation results reveal that hybrid-ZF outperforms the full-ZF scheme with a few clusters, while the full-ZF shows a better performance with a higher number of clusters.

7.1 Future work

The research work presented in this thesis showed that NOMA is a promising candidate for multiple access technique in future wireless network. However, for the practical realization of the NOMA still much research effort is required to develop advanced NOMA signal processing techniques with reduced complexity. There are a number of interesting research directions in which the research of this thesis can be extended, as follows:

- **NOMA with massive MIMO:** Massive MIMO systems have the potential capabilities to significantly improve the data throughput. Hence, it will be very interesting to extend the work in this thesis to a massive MIMO system. However, the problem will become more complex. Hence, finding effective ways to combine NOMA and massive MIMO is an interesting direction which can be considered as a future research work.

- **Cooperative NOMA System:** The basic idea of cooperative NOMA transmission is that stronger users act as relays to help weaker users. As SIC is employed at receivers in NOMA systems, the signals to the weaker users have already been decoded by the stronger users. Hence, the stronger users can be considered as relays. A typical cooperative NOMA transmission scheme can be divided into two phases, the direct transmission phase and cooperative transmission phase. During the direct transmission phase, the BS broadcasts a combination of signals for weak and strong users. During the cooperative transmission phase, after applying SIC at the strong user for decoding weaker user's signal, the strong user acts as a relay to forward the decoded information to the weak user. Therefore, weak users receive two copies of the signals through different channels. As a result, the reception reliability of the weak users is significantly improved. As power allocation has been recognized to have a great impact on the performance of NOMA systems, investigating optimal power allocation to further improve the performance of cooperative NOMA systems is another interesting research direction. Hence, the beamforming design in the cooperative NOMA system would be an interesting possible extension.

- **Hybrid topology of NOMA:** In this thesis, employing NOMA as a multiple access technique for future wireless systems is investigated. However, it is envisioned that future wireless networks will be designed using more than one multiple access technique. Combining NOMA and OMA is an effective way which can significantly improve the capacity of 5G and beyond wireless networks. Therefore, it is important to study how to effectively combine NOMA with other types of multiple access schemes to achieve the best performance with available radio resources. In addition, a hybrid multiple access system is viewed as a promising solution to reduce the complexity of cooperative NOMA system. Hence, it would be very interesting to extend the system model used in this thesis by dividing users into multiple clusters to incorporate cooperative NOMA within each cluster, while implementing OMA between clusters.

Appendix A

Proofs for Chapter 4

A.1 Proof of Lemma 1

First the following complex derivatives are summarized as:

Generalized complex derivative

$$\frac{\partial f(z)}{\partial z} = \frac{1}{2}\left(\frac{\partial f(z)}{\partial \Re(z)} - i\frac{\partial f(z)}{\partial \Im(z)}\right) \tag{A.1}$$

Conjugate complex derivative

$$\frac{\partial f(z)}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f(z)}{\partial \Re(z)} + i\frac{\partial f(z)}{\partial \Im(z)}\right) \tag{A.2}$$

$$\Rightarrow \qquad \frac{\partial f(z)}{\partial \Re(z)} = 2\Re\left[\frac{\partial f(z)}{\partial z^*}\right], \qquad \frac{\partial f(z)}{\partial \Im(z)} = 2\Im\left[\frac{\partial f(z)}{\partial z^*}\right] \tag{A.3}$$

Next, the first order Taylor series approximation is presented for a function $g(A)$ around A_0 as follows:

$$\begin{aligned} g(A) \approx g(A, A_0) &= f(A_0) + \mathrm{Tr}\Big(\left[\frac{\partial f(A_0)}{\partial \Re(A)}\right]^T \Re(A - A_0)\Big) \\ &\quad + \mathrm{Tr}\Big(\left[\frac{\partial f(A_0)}{\partial \Im(A)}\right]^T \Im(A - A_0)\Big) \\ &= f(A_0) + \mathrm{Tr}\Big(2 \left[\frac{\partial f(A_0)}{\partial A^*}\right]^T \Re(A - A_0)\Big) \\ &\quad + \mathrm{Tr}\Big(2 \left[\frac{\partial f(A_0)}{\partial A^*}\right]^T \Im(A - A_0)\Big) \\ &= f(A_0) + 2\Re\Big(\mathrm{Tr} \left[\frac{\partial f(A_0)}{\partial A^*}\right]^H (A - A_0)\Big) \end{aligned} \tag{A.4}$$

Based on the above approximation, the following approximation can be obtained $f_l(\mathbf{w}_k) = \mathbf{w}_k^H \mathbf{h}_l \mathbf{h}_l^H \mathbf{w}_k$ as follows:

$$\begin{aligned} g_k(\mathbf{w}_k, \tilde{\mathbf{w}}_k) &= \tilde{\mathbf{w}}_k^H \mathbf{h}_l {\mathbf{h}_l}^H \tilde{\mathbf{w}}_k + \Big[\Big(\frac{\partial f_l(\mathbf{w}_k)}{\partial \Re(\mathbf{w}_k)}\Big|_{\mathbf{w}_k = \tilde{\mathbf{w}}_k}\Big)^T \Re(\mathbf{w}_k - \tilde{\mathbf{w}}_k)\Big] \\ &\quad + \Big[\Big(\frac{\partial f_l(\mathbf{w}_k)}{\partial \Im(\mathbf{w}_k)}\Big|_{\mathbf{w}_k = \tilde{\mathbf{w}}_k}\Big)^T \Im(\mathbf{w}_k - \tilde{\mathbf{w}}_k)\Big] \end{aligned} \tag{A.5}$$

and based on equations in (A.3), the following terms can be written

$$\frac{\partial f_l(\mathbf{w}_k)}{\partial \Re(\mathbf{w}_k)} = 2\Re(\mathbf{h}_l {\mathbf{h}_l}^H \mathbf{w}_k) \tag{A.6}$$

$$\frac{\partial f_l(\mathbf{w}_k)}{\partial \Im(\mathbf{w}_k)} = 2\Im(\mathbf{h}_l {\mathbf{h}_l}^H \mathbf{w}_k) \tag{A.7}$$

$$\begin{aligned} g_l(\mathbf{w}_k, \tilde{\mathbf{w}}_k) &= \tilde{\mathbf{w}}_k^H \mathbf{h}_l \mathbf{h}_l^H \tilde{\mathbf{w}}_k + 2\big(\Re(\tilde{\mathbf{w}}_k^H \mathbf{h}_l {\mathbf{h}_l}^H)\Re(\mathbf{w}_k - \tilde{\mathbf{w}}_k)\big) \\ &\quad + 2\big(\Im(\tilde{\mathbf{w}}_k^H \mathbf{h}_l {\mathbf{h}_l}^H)\Im(\mathbf{w}_k - \tilde{\mathbf{w}}_k)\big) \\ &= \tilde{\mathbf{w}}_k^H \mathbf{h}_l \mathbf{h}_l^H \tilde{\mathbf{w}}_k + 2\Re[\tilde{\mathbf{w}}_{\mathbf{k}}^H \mathbf{h}_1 {\mathbf{h}_1}^H (\mathbf{w}_k - \tilde{\mathbf{w}}_k)] \end{aligned} \tag{A.8}$$

This completes the proof of Lemma 1. ■

A.2 Proof of Lemma 2

A maximization optimization problem is quasi-concave when the objective function is quasi-concave and the constraints are convex. The constraint (4.17b) is convex, however the constraint (4.17c) can be converted to a convex one by using the methods provided in Section 4.3. Based on the quasi-convex definition a function f is called quasi-convex if its domain and all its sublevel sets $S_\alpha = \{\mathbf{x} \in \mathbf{dom} f | f(\mathbf{x}) \leq \alpha\}$, for $\alpha \in \mathbb{R}_+$, are convex. A function is quasi-concave if $-f$ is quasi-convex, i.e., every super level set $\{\mathbf{x} \in \mathbf{dom} f | f(\mathbf{x}) \geq \alpha\}$ is convex. Clearly, for the objective function in (4.17), $\min R_k(\mathbf{w})$, to be quasi-concave, all its super level sets must be convex, i.e., $S_\alpha = \{\min R_k(\mathbf{w}) \geq \alpha\}$, which represents the set $\mathbf{w} = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ that makes the objective function greater than a specific threshold, α. Since there is the $\min$ operator, it can be rewritten as $S_\alpha = \{R_k \geq \alpha, \forall k\}$ and the constraints $R_k \geq \alpha, \forall k$ can be obtained as

$$(2^\alpha - 1)(\Sigma_{m=k+1}^{K} |\mathbf{h}_l^H \mathbf{w}_m|^2 + \sigma_l^2) \leq |\mathbf{h}_l^H \mathbf{w}_k|^2, \quad \forall k,\ l = k, k+1, \ldots, K. \tag{A.9}$$

By employing the Taylor series approximation presented in Section 4.3.1, the constraint in (4.17c) and (A.9) can be transformed into a convex one, which completes the proof. ■

Appendix B

Proofs for Chapter 5

B.1 Proof of Lemma 3

To incorporate the channel uncertainties in the robust optimization problem in (5.8), let consider the following lemma [151]:

Lemma 3.1: (S-procedure): Let $\mathbf{A}_i \in \mathbb{C}^{N\times N}$ be symmetric matrices, $\mathbf{b}_i \in \mathbb{C}^{N\times 1}$ be vectors and c_i be real numbers. Assume that there is some $\mathbf{x}_0$ such that the following inequality holds.

$$\mathbf{x}_0^H \mathbf{A}_1 \mathbf{x}_0 + 2\mathrm{Re}\{\mathbf{b}_1^H \mathbf{x}_0\} + c_1 \leq 0,$$

Then, the implication

$$\begin{aligned} &\mathbf{x}^H \mathbf{A}_1 \mathbf{x} + 2\mathrm{Re}\{\mathbf{b}_1^H \mathbf{x}\} + c_1 \leq 0 \\ \Longrightarrow\ &\mathbf{x}^H \mathbf{A}_2 \mathbf{x} + 2\mathrm{Re}\{\mathbf{b}_2^H \mathbf{x}\} + c_2 \leq 0, \end{aligned}$$

holds if and only if there exists some nonnegative number $\mu \geq 0$ such that

$$\mu \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^H & c_1 \end{bmatrix} - \begin{bmatrix} \mathbf{A}_2 & \mathbf{b}_2 \\ \mathbf{b}_2^H & c_2 \end{bmatrix} \succeq 0,$$

is positive semi-definite.

For applying S-procedure, the constraint (5.8b) is derived as

$$\begin{aligned}
&\Delta\hat{\mathbf{h}}_l^H \mathbf{I} \Delta\hat{\mathbf{h}}_l - \epsilon^2 \leq 0 \\
\Longrightarrow\ &\Delta\hat{\mathbf{h}}_l^H \left(\gamma_k^{\min} \sum_{m \neq k} \mathbf{W}_m - \mathbf{W}_k \right) \Delta\hat{\mathbf{h}}_l \\
&+ 2\Re \left\{ \hat{\mathbf{h}}_l^H \left(\gamma_k^{\min} \sum_{m=k+1}^{K} \mathbf{W}_m - \mathbf{W}_k \right) \Delta\hat{\mathbf{h}}_l \right\} \\
&+ \hat{\mathbf{h}}_l^H \left(\gamma_k^{\min} \sum_{m=k+1}^{K} \mathbf{W}_m - \mathbf{W}_k \right) \hat{\mathbf{h}}_l + \gamma_k^{\min} \sigma_l^2 \leq 0,\ \forall\, k,\ \ l = k, \ldots, K.
\end{aligned}$$

Then, based on Lemma 3.1, the constraint (5.8b) can be reformulated with $\mu_{kl} \geq 0$ as

$$C_{kl} = \begin{bmatrix} \mu_{kl}\mathbf{I} + \varphi_k & \phi_k \hat{\mathbf{h}}_l \\ \hat{\mathbf{h}}_l^H \phi_k & \hat{\mathbf{h}}_l^H \phi_k \hat{\mathbf{h}}_l - \gamma_k^{\min} \sigma_k^2 - \mu_{kl} \epsilon^2 \end{bmatrix} \succeq 0, \tag{B.1}$$

where $\varphi_k = \mathbf{W}_k - \gamma_k^{\min} \sum_{m \neq k} \mathbf{W}_m$ and $\phi_k = \mathbf{W}_k - \gamma_k^{\min} \sum_{m=k+1}^{K} \mathbf{W}_m$.

This completes the proof of Lemma 3. ■

B.2 Proof of Lemma 4

To convert probabilistic constraints in (5.15b) into a tractable form the following lemma is required:

Lemma 4.1: Consider a hermitian random matrix $\mathbf{X} \in \mathbb{C}^{N\times N}$ with each element being independently characterized as $[\mathbf{X}]_{ij} \sim \mathcal{CN}(0, \sigma_{ij}^2)$. Then, for any hermitian matrix $\mathbf{Y} \in \mathbb{C}^{N\times N}$, it holds

$$\mathrm{Tr}(\mathbf{YX}) \sim \mathcal{CN}(0, \|\mathbf{Y} \odot \Sigma_{\mathrm{X}}\|_{\mathrm{F}}^2), \tag{B.2}$$

$$\mathrm{Tr}(\mathbf{YX}) = \|\mathbf{Y} \odot \Sigma_{\mathrm{X}}\|_{\mathrm{F}}\mathrm{U},\ \mathrm{U} \sim \mathcal{N}(0,1), \tag{B.3}$$

where $\odot$ indicates Hadamard product and Σ_X represents a real valued $N \times N$ matrix with each entry $[\Sigma_X]_{ij} = \sigma_{ij}$, [115].

By exploiting Lemma 4.1 and the CDF of a standard normal distribution as

$$\mathbf{U} \sim \mathcal{N}(0,1) \ \Rightarrow\ \Pr\left(U \le u\right) = \frac{1}{2}\left[1 + \mathrm{erf}(\frac{u}{\sqrt{2}})\right], \tag{B.4}$$

the constraint in (5.15b) can be represented as follows:

$$\begin{aligned}
\Pr\left(\mathrm{Tr}(-\acute{\mathbf{B}}_k\Delta_l) \le \mathrm{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma^2\right) &\overset{(B.3)}{=} \Pr\left(\| - \acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}\|_F U \le \mathrm{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma^2\right) \\
&= \Pr\left(U \le \frac{\mathrm{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma^2}{\| - \acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}\|_F}\right) \\
&\overset{(B.4)}{=} \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathrm{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma^2}{\sqrt{2}\| - \acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}\|_F}\right)\right] \\
&\ge (1-\rho_k), \quad \forall k,\ l = k, k+1, \ldots, K.
\end{aligned} \tag{B.5}$$

The inequalities in (B.5) can be written in the following forms:

$$\Phi_{kl} \ge \sqrt{2}\,\mathrm{erf}^{-1}(1-2\rho_k)\|\mathrm{vec}(-\acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l})\|, \tag{B.6}$$

where $\Phi_{kl} = \mathrm{Tr}(\mathbf{B}_k\hat{\mathbf{C}}_l) - \sigma^2$.

After converting the constraint in (5.15b) to (B.6), the original robust problem can be transformed into a convex optimization framework through the following lemma:

Lemma 4.2: The following second order cone constraint on $\mathbf{x}$

$$\|\mathbf{A}\mathbf{x}+\mathbf{b}\| \leq \mathbf{e}^T\mathbf{x}+d, \tag{B.7}$$

can be represented with the following LMI [127]:

$$\begin{bmatrix} (\mathbf{e}^T\mathbf{x}+d)\mathbf{I} & \mathbf{A}\mathbf{x}+\mathbf{b} \\ (\mathbf{A}\mathbf{x}+\mathbf{b})^T & \mathbf{e}^T\mathbf{x}+d \end{bmatrix} \succeq 0. \tag{B.8}$$

By applying Lemma 4.2, the constraints in (B.6) can be reformulated as

$$\mathbf{C}_{kl} = \begin{bmatrix} \frac{\Phi_{kl}}{\sqrt{2}\mathrm{erf}^{-1}(1-2\gamma_k)}\mathbf{I}_{M^2} & \mathrm{vec}(-\acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}) \\ \mathrm{vec}^H(-\acute{\mathbf{B}}_k \odot \Sigma_{\Delta_l}) & \frac{\Phi_{kl}}{\sqrt{2}\mathrm{erf}^{-1}(1-2\gamma_k)} \end{bmatrix}, \ \forall k, \ l=k,k+1,\ldots,K. \tag{B.9}$$

This completes the proof of Lemma 4. ■

Appendix C

Proofs for Chapter 6

C.1 Proof of Lemma 5

Assume that the numerator and denominator of $\text{SINR}_l^{(k)}$ in (6.5) are independent and then their worst-case terms are derived, separatively. Based on this assumption, a function $\varphi_l^{(k)}$ is first introduced as a lower bound of $\min_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\ \text{SINR}_l^{(k)}$ as

$$\varphi_l^{(k)} = \frac{p_k\ \min_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_k\right|^2\right)}{\sum_{m=1}^{k-1}p_m\max\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|\Delta\hat{\mathbf{h}}_l^H\mathbf{v}_m\right|^2\right)+\sum\limits_{m=k+1}^{K}p_m\max\limits_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_m\right|^2\right)+\sigma^2}. \tag{C.1}$$

By employing triangle inequality and Cauchy-Schwarz and considering $\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon$, the following holds:

$$\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_k\right| \geq \left|\hat{\mathbf{h}}_l^H\mathbf{v}_k\right| - \left|\Delta\hat{\mathbf{h}}_l^H\mathbf{v}_k\right| \geq \left|\hat{\mathbf{h}}_l^H\mathbf{v}_k\right| - \varepsilon\|\mathbf{v}_k\|, \tag{C.2}$$

$$\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_m\right| \leq \left|\hat{\mathbf{h}}_l^H\mathbf{v}_m\right| + \left|\Delta\hat{\mathbf{h}}_l^H\mathbf{v}_m\right| \leq \left|\hat{\mathbf{h}}_l^H\mathbf{v}_m\right| + \varepsilon\|\mathbf{v}_m\|. \tag{C.3}$$

Then, (C.2) and (C.3) are applied to the numerator and denominator of (C.1) as follows:

$$\min_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_k\right|^2\right) \triangleq \left|\left(\left|\hat{\mathbf{h}}_l^H\mathbf{v}_k\right| - \varepsilon\|\mathbf{v}_k\|\right)^+\right|^2, \tag{C.4}$$

$$\max_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_l+\Delta\hat{\mathbf{h}}_l)^H\mathbf{v}_m\right|^2\right) \triangleq \left|\left|\hat{\mathbf{h}}_l^H\mathbf{v}_m\right| + \varepsilon\|\mathbf{v}_m\|\right|^2 \tag{C.5}$$

$$\max_{\|\Delta\hat{\mathbf{h}}_l\|_2\leq\varepsilon}\left(\left|\Delta\hat{\mathbf{h}}_l^H\mathbf{v}_m\right|^2\right) \triangleq \varepsilon^2\ \|\mathbf{v}_m\|^2. \tag{C.6}$$

This completes the proof of Lemma 5. ■

C.2 Dinkelbach's algorithm

Dinkelbach's algorithm is a well-known technique to solve the a CCFP

$$\max_{x} \quad \frac{f(\mathbf{x})}{g(\mathbf{x})}, \tag{C.7a}$$

$$\text{subject to} \quad c_i(\mathbf{x}) \leq 0, \ \forall\, i = 1, \ldots, I, \tag{C.7b}$$

$$h_j(\mathbf{x}) = 0, \ \forall\, j = 1, \ldots, J, \tag{C.7c}$$

where $f(\mathbf{x})$ is a non-negative differentiable concave function, $g(\mathbf{x})$ is a positive differentiable convex function, c_i is convex for all $i = 1, \ldots, I$, and h_j is an affine function for all $j = 1, \ldots, J$.

Dinkelbach's algorithm has been originally introduced in [157, 158]. Furthermore, it belongs to the class of parametric algorithms. The fundamental concepts of these algorithms is to obtain the solution of a CCFP by solving a sequence of simple subproblems which converges to the global optimal solution of the original CCFP. The fundamental result upon which Dinkelbach's algorithm is built is the relation between the CCFP in (C.7) and the function of real variable as

$$F(\lambda) = \max_{\mathbf{x}} \{f(\mathbf{x}) - \lambda g(\mathbf{x})\}. \tag{C.8}$$

Solving a fractional problem is equivalent to finding the unique zero of the auxiliary function $F(\lambda)$ where Dinkelbach's algorithm allows to accomplish this. The pseudocode of Dinkelbach's algorithm is provided in Table 6.1.

C.3 Proof of Lemma 6

By introducing new variables $\mathbf{W}_k$, the constraint (6.18b) can be rewritten as

$$\gamma_k \leq \frac{\mathbf{h}_l^H \mathbf{W}_k \mathbf{h}_l}{\sum_{m=1}^{k-1} \Delta\hat{\mathbf{h}}_l^H \mathbf{W}_m \Delta\hat{\mathbf{h}}_l + \sum_{m=k+1}^{K} \mathbf{h}_l^H \mathbf{W}_m \mathbf{h}_l + \sigma_l^2}, \quad \forall k, l = k, k+1, \ldots, K. \tag{C.9}$$

Then, by applying S-procedure in Appendix B.1, the constraint (C.9) can be derived as

$$\begin{aligned}
&\Delta\hat{\mathbf{h}}_l^H \mathbf{I} \Delta\hat{\mathbf{h}}_l - \varepsilon_l^2 \leq 0 \\
\Rightarrow &\Delta\hat{\mathbf{h}}_l^H \left(\sum_{m \neq k} \mathbf{W}_m - \mathbf{W}_k/\gamma_k \right) \Delta\hat{\mathbf{h}}_l + 2\Re\left\{ \hat{\mathbf{h}}_l^H \left(\sum_{m=k+1}^{K} \mathbf{W}_m - \mathbf{W}_k/\gamma_k \right) \Delta\hat{\mathbf{h}}_l \right\} \\
&+ \hat{\mathbf{h}}_l^H \left(\sum_{m=k+1}^{K} \mathbf{W}_m - \mathbf{W}_k/\gamma_k \right) \hat{\mathbf{h}}_l + \sigma_l^2 \leq 0.
\end{aligned} \tag{C.10}$$

Based on S-procedure, the inequality (C.10) is satisfied for all possible channel uncertainties if there exists $\mu_{kl} \geq 0$ and the following LMI holds:

$$\begin{bmatrix} \mu_{kl}\mathbf{I} + \phi_k - \sum_{m=1}^{k-1} \mathbf{W}_m & \phi_k \hat{\mathbf{h}}_l \\ \hat{\mathbf{h}}_l^H \phi_k & \hat{\mathbf{h}}_l^H \phi_k \hat{\mathbf{h}}_l - \sigma_l^2 - \mu_{kl}\varepsilon_l^2 \end{bmatrix} \succeq \mathbf{0}, \tag{C.11}$$

This completes the proof of Lemma 6. ■

Glossary

1G	First Generation
2G	Second Generation
3G	Third Generation
4G	Forth Generation
5G	Fifth Generation
A/D	Analog-to-Digital
AWGN	Additive White Gaussian Noise
BS	Base Station
CCP	Convex-Concave Procedure
CCFP	Concave-Convex Fractional Problem
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CSI	Channel State Information
D2D	Device-to-Device
EE	Energy Efficiency
GPRS	General Packet Radio Service
HD	High Definition
i.i.d	independently and identically distributed

IMT	International Mobile Telecommunications
IoT	Internet-of-Things
ITU-R	International Telecommunications Union-Radio
LDS	Low-Density Spreading
LMI	Linear Matrix Inequality
LP	Linear Program
LTE	Long Term Evolution
M2M	Machine-to-Machine
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
mm-Wave	Millimeter Wave
MUSA	Multi-User Shared Access
NLP	Non-Linear Program
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PDF	Probability Density Function
PDMA	Pattern Division Multiple Access
QCQP	Quadratically Constrained Quadratic Program
QoS	Quality of Service
QP	Quadratic Program
SC	Superposition Coding

SCMA	Sparse Code Multiple Access
SDMA	Spatial Division Multiple Access
SDP	Semidefinite Programming
SDR	Semidefinite Relaxation
SE	Spectral Efficiency
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SOC	Second-Order Cone
SOCP	Second-Order Cone Programming
TDMA	Time Division Multiple Access
ZF	Zero-Forcing

Notations

$\mathbb{C}^m$	Set of complex $m \times 1$ vectors
$\mathbb{R}^m$	Set of real $m \times 1$ vectors
$\mathbb{C}^{m \times n}$	Set of complex $m \times n$ matrices
$\mathbb{R}^{m \times n}$	Set of real $m \times n$ matrices
$\mathbb{R}_+^m$	Set of nonnegative real $m \times 1$ vectors
$\lvert x \rvert$	Modulus of complex number x
$\mathbf{x}$	Vector $\mathbf{x}$
$\mathbf{X}$	Matrix $\mathbf{X}$
$[\mathbf{X}]_{i,j}$	(i,j)-th element of matrix $\mathbf{X}$
$(\cdot)^T$	Transpose
$(\cdot)^H$	Transpose conjugate
$\mathbf{X}^{-1}$	Inverse of the matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	Pseudo-inverse of the matrix $\mathbf{X}$
$\mathbf{X} \succeq \mathbf{0}$	Positive semi-definite matrix
$\mathrm{Tr}(\mathbf{X})$	Trace of the matrix $\mathbf{X}$
$\mathrm{vec}(\mathbf{X})$	Vector of the matrix $\mathbf{X}$
$\Re(x)$	Real part of a complex number x
$\Im(x)$	Imaginary part of a complex number x
$\lvert\lvert \cdot \rvert\rvert_2$	Euclidian vector norm, i.e., $\lvert\lvert\mathbf{x}\rvert\rvert_2 = \sqrt{\mathbf{x}^H\mathbf{x}}$
$\odot$	Hadamard product
$\mathbf{I}_N$	$N \times N$ Identity matrix
$\mathbb{E}_{\mathbf{X}}[\cdot]$	Expectation of a random variable over $\mathbf{X}$
$\Pr(\cdot)$	Probability operator
$\mathcal{CN}(\mu, \sigma^2)$	Complex Gaussian random variable with mean μ and variance σ^2
$\mathcal{N}(\mu, \sigma^2)$	Real Gaussian random variable with mean μ and variance σ^2
dom f	Domain of a function f
$\min\{\}, \max\{\}$	Minimum and maximum function

References

[1] A. J. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[2] R. Vannithamby and S. Talwar, *Towards 5G: Applications, Requirements and Candidate Technologies*. Wiley-Blackwell.

[3] A. A. Huurdeman, *The Worldwide History of Telecommunications*. New Jersey: John Wiley and Sons Press, 2003.

[4] "ITU-R M.2134: Requirements related to technical performance for IMTadvanced radio interface(s)," ITU-R, Tech. Rep., 2008.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[6] Tech27Systems, "Smart world," *http://www.tech27.com/industries/Infrastructure*.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, February 2014.

[8] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart community: an internet of things application," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 68–75, November 2011.

[9] K. Cumanan, H. Xing, P. Xu, G. Zheng, X. Dai, A. Nallanathan, Z. Ding, and G. K. Karagiannidis, "Physical layer security jamming: Theoretical limits and practical designs in wireless networks," *IEEE Access*, vol. 5, pp. 3603–3611, 2017.

[10] Cisco, "Internet of things cloud: Architecture and implementation," *Cisco visual network index: Global mobile traffic forecast update*, 2012.

[11] V. Kostakos, T. Ojala, and T. Juntunen, "Traffic in the smart city: Exploring city-wide sensing for traffic control center augmentation," *IEEE Internet Computing*, vol. 17, no. 6, pp. 22–29, November 2013.

[12] J. A. Stankovic, "Research directions for the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, February 2014.

[13] B. Ahlgren, M. Hidell, and E. C. H. Ngai, "Internet of things for smart cities: Interoperability and open data," *IEEE Internet Computing*, vol. 20, no. 6, pp. 52–56, November 2016.

[14] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband.* Academic Press, 2011.

[15] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, February 2014.

[16] Y. Zhang and M.Chen, *Cloud Based 5G Wireless Networks*, ser. SpringerBriefs in Computer Science. Springer International Publishing, 2016. [Online]. Available: https://books.google.se/books?id=LNN5DQAAQBAJ

[17] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 34, pp. 185–396, 2015.

[18] B. Lannoo, "Overview of ICT energy consumption," *iMinds, Flanders' digital research center report FP7-288021*, February 2013.

[19] E. Gelenbe and Y. Caseau, "The impact of information technology on energy consumption and carbon emissions," *Ubiquity*, June. 2015.

[20] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless commu-

nications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[21] 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, February 2014.

[22] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, February 2014.

[23] F. Alavi, N. M. Yamchi, M. R. Javan, and K. Cumanan, "Limited feedback scheme for device-to-device communications in 5G cellular networks with reliability and cellular secrecy outage constraints," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8072–8085, September 2017.

[24] L. Dai, B. Wang, Y. Yuan, S. Han, C. L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, September 2015.

[25] J. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Trans. Commun.*, vol. 35, no. 11, pp. 1222–1230, Nov. 1987.

[26] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[27] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecom*, vol. 10, no. 6, pp. 585–595, 1999.

[28] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1445–1459, March 2018.

[29] C. S. Park, Y. P. E. Wang, G. Jongren, and D. Hammarwall, "Evolution of uplink MIMO for LTE-advanced," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 112–121, February 2011.

[30] M. Fozooni, M. Matthaiou, E. Bjornson, and T. Q. Duong, "Performance limits of MIMO systems with nonlinear power amplifiers," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2015, pp. 1–7.

[31] N. Khalid and O. B. Akan, "Experimental throughput analysis of low-THz MIMO communication channel in 5G wireless networks," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 616–619, December 2016.

[32] 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, May 2014.

[33] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, October 2014.

[34] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, May 2014.

[35] K. Cumanan, Y. Rahulamathavan, S. Lambotharan, and Z. Ding, "MMSE-based beamforming techniques for relay broadcast channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4045–4051, October 2013.

[36] K. Cumanan, R. Zhang, and S. Lambotharan, "A new design paradigm for MIMO cognitive radio with primary user rate constraint," *IEEE Communications Letters*, vol. 16, no. 5, pp. 706–709, May 2012.

[37] M. Fozooni, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Massive MIMO relaying with hybrid processing," in *Proc. IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[38] M. Fozooni, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation," *IEEE Transactions on Communications*, pp. 1–1, 2018.

[39] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[40] ——, "Aspects of favorable propagation in massive MIMO," in *European Signal Processing Conference (EUSIPCO)*, September 2014, pp. 76–80.

[41] L. Jiang, H. Tian, Z. Xing, K. Wang, K. Zhang, S. Maharjan, S. Gjessing, and Y. Zhang, "Social-aware energy harvesting device-to-device communications in 5G networks," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 20–27, August 2016.

[42] O. N. C. Yilmaz, Z. Li, K. Valkealahti, M. A. Uusitalo, M. Moisio, P. Lunden, and C. Wijting, "Smart mobility management for D2D communications in 5G networks," in *Proc. IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2014, pp. 219–223.

[43] H. A. U. Mustafa, M. A. Imran, M. Z. Shakir, A. Imran, and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 419–445, 2016.

[44] L. Musavian and S. Aissa, "On the achievable sum-rate of correlated MIMO multiple access channel with imperfect channel estimation," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2549–2559, July 2008.

[45] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: key elements and technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, March 2014.

[46] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, November 2013, pp. 770–774.

[47] S. Sesia, I. Toufik, and M. Baker, *Orthogonal Frequency Division Multiple Access (OFDMA)*. Wiley Telecom, 2011, pp. 792–.

[48] N. Abu-Ali, A. E. M. Taha, M. Salah, and H. Hassanein, "Uplink scheduling in LTE and LTE-advanced: Tutorial, survey and evaluation framework," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1239–1265, 2014.

[49] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2734–2744, September 2009.

[50] H. Zhu, "Radio resource allocation for OFDMA systems in high speed environments," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 748–759, May 2012.

[51] F. Alavi, N. Mokari, and H. Saeedi, "Secure resource allocation in OFDMA-based cognitive radio networks with two-way relays," in *Proc. Iranian Conference on Electrical Engineering*, May 2015, pp. 171–176.

[52] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[53] Y. Endo, Y. Kishiyama, and K. Higuchi, "Uplink non-orthogonal access with MMSE-SIC in the presence of inter-cell interference," in *International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2012, pp. 261–265.

[54] R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Communications Surveys Tutorials*, vol. 13, no. 3, pp. 503–520, 2011.

[55] G. Wunder, M. Kasparick, S. T. Brink, F. Schaich, T. Wild, Y. Chen, I. Gaspar, N. Michailow, G. Fettweis, D. Ktenas, N. Cassiau, M. Dryjanski, K. Sorokosz, S. Pietrzyk, and B. Eged, "System-level interfaces and performance evaluation methodology for 5G physical layer based on non-orthogonal waveforms," in *Asilomar Conference on Signals, Systems and Computers*, November 2013, pp. 1659–163.

[56] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

[57] M. T. P. Le, G. C. Ferrante, T. Q. S. Quek, and M. G. D. Benedetto, "Fundamental limits of low-density spreading noma with fading," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2018.

[58] Y. Wu, C. Wang, Y. Chen, and A. Bayesteh, "Sparse code multiple access for 5G radio transmission," in *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, September 2017, pp. 1–6.

[59] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.

[60] J. Zeng, B. Li, X. Su, L. Rong, and R. Xing, "Pattern division multiple access (PDMA) for cellular future radio access," in *Proc. International Conference on Wireless Communications Signal Processing (WCSP)*, October 2015, pp. 1–5.

[61] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for OFDM/SDMA," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1204–1222, March 2014.

[62] WhitePaper, "Rethink mobile communications for 2020+," *Future Mobile Communication Forum 5G SIG*, November 2014.

[63] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 127–135, March 2013.

[64] H. Zhu, S. Karachontzitis, and D. Toumpakaris, "Low-complexity resource allocation and its application to distributed antenna systems [coordinated and distributed MIMO]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 44–50, June 2010.

[65] H. Zhu, "On frequency reuse in cooperative distributed antenna systems," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 85–89, April 2012.

[66] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 675–683, May 2012.

[67] H. Elsawy, E. Hossain, and D. I. Kim, "HetNets with cognitive small cells: user offloading and distributed channel access techniques," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 28–36, June 2013.

[68] M. Z. Shakir, K. A. Qaraqe, H. Tabassum, M. S. Alouini, E. Serpedin, and M. A. Imran, "Green heterogeneous small-cell networks: toward reducing the CO2 emissions of mobile communications industry using uplink power adaptation," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 52–61, June 2013.

[69] H. Wang, X. Zhou, and M. C. Reed, "Coverage and throughput analysis with a non-uniform small cell deployment," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2047–2059, April 2014.

[70] 3rd Generation Partnership Project 2, "Physical layer for ultra mobile broadband (UMB)," *IEEE Communications Magazine*.

[71] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[72] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, April 2014.

[73] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[74] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Beamforming techniques for non-orthogonal multiple access in 5G cellular networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2018.

[75] ——, "Robust beamforming techniques for non-orthogonal multiple access systems with bounded channel uncertainties," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2033–2036, September 2017.

[76] ——, "Outage constraint based robust beamforming design for non-orthogonal multiple access in 5G cellular networks," in *Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, October 2017, pp. 1–5.

[77] 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 *IEEE Military Communications Conference (MILCOM)*, November 2013, pp. 1278–1283.

[78] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, September 2013, pp. 611–615.

[79] 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, January 2016.

[80] S. Verdu, *Multiuser Detection*, 1st ed. New York, NY, USA: Cambridge University Press, 1998.

[81] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 312–335, 2013.

[82] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[83] I. Krikidis, "Analysis and optimization issues for superposition modulation in cooperative networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4837–4847, November 2009.

[84] A. Benjebbour, A. Li, Y. Kishiyama, H. Jiang, and T. Nakamura, "System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements," in *IEEE Globecom Workshops (GC Wkshps)*, December 2014, pp. 706–710.

[85] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE Vehicular Technology Conference (VTC Fall)*, September 2013, pp. 1–5.

[86] N. Nonaka, Y. Kishiyama, and K. Higuchi, "Non-orthogonal multiple access using intra-beam superposition coding and SIC in base station cooperative MIMO

cellular downlink," in *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, September 2014, pp. 1–5.

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

[88] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6174–6189, December 2016.

[89] 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, October 2017.

[90] Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, and X. You, "Signal processing for MIMO-NOMA: Present and future challenges," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 32–38, April 2018.

[91] Q. Sun, S. Han, C. L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, August 2015.

[92] X. Chen, Z. Zhang, C. Zhong, and D. W. K. Ng, "Exploiting multiple-antenna techniques for non-orthogonal multiple access," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2207–2220, October 2017.

[93] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4438–4454, June 2016.

[94] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 791–800, March 2015.

[95] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for non-orthogonal multiple-access systems in MISO channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10 231–10 236, December 2016.

[96] G. Parsaee and A. Yarali, "OFDMA for the 4th generation cellular networks," in *Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513)*, vol. 4, May 2004, pp. 2325–2330 Vol.4.

[97] T. Jiang, W. Xiang, H. H. Chen, and Q. Ni, "Multicast broadcast services support in OFDMA-based WiMAX systems [advances in mobile multimedia]," *IEEE Communications Magazine*, vol. 45, no. 8, pp. 78–86, August 2007.

[98] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y. D. Kim, E. Kim, and Y. C. Cheong, "An overview of radio resource management in relay-enhanced OFDMA-based networks," *IEEE Communications Surveys Tutorials*, vol. 12, no. 3, pp. 422–438, 2010.

[99] M. M. Wang, A. Agrawal, A. Khandekar, and S. Aedudodla, "Preamble design, system acquisition, and determination in modern OFDMA cellular communications: an overview," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 164–175, July 2011.

[100] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. L. I, 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, February 2017.

[101] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Processing Letters*, vol. 21, no. 12, pp. 1501–1505, December 2014.

[102] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. 11th International Symposium on Wireless Communications Systems (ISWCS)*, August 2014, pp. 781–785.

[103] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, August 2016.

[104] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, August 2015.

[105] C. Xue, Q. Zhang, Q. Li, and J. Qin, "Joint power allocation and relay beamforming in nonorthogonal multiple access amplify-and-forward relay networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7558–7562, August 2017.

[106] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to nonorthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, January 2016.

[107] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 629–633, May 2016.

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

[109] Z. Liu, L. Lei, N. Zhang, G. Kang, and S. Chatzinotas, "Joint beamforming and power optimization with iterative user clustering for MISO-NOMA systems," *IEEE Access*, 2017.

[110] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Secure beamforming in downlink MISO nonorthogonal multiple access systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7563–7567, August 2017.

[111] Q. Li and W. K. Ma, "Optimal and robust transmit designs for MISO channel secrecy by semidefinite programming," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3799–3812, August 2011.

[112] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, June 2012.

[113] S. K. Joshi, U. L. Wijewardhana, M. Codreanu, and M. Latva-aho, "Maximization of worst-case weighted sum-rate for MISO downlink systems with imperfect channel knowledge," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3671–3685, October 2015.

[114] K. Cumanan, Z. Ding, Y. Rahulamathavan, M. M. Molu, and H. H. Chen, "Robust MMSE beamforming for multiantenna relay networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3900–3912, May 2017.

[115] S. Nasseri and M. R. Nakhai, "Robust interference management via outage-constrained downlink beamforming in multicell networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2013, pp. 3470–3475.

[116] C. Shen, T. H. Chang, K. Y. Wang, Z. Qiu, and C. Y. Chi, "Chance-constrained robust beamforming for multi-cell coordinated downlink," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2012, pp. 4957–4962.

[117] X. He and Y. C. Wu, "Tight probabilistic SINR constrained beamforming under channel uncertainties," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3490–3505, July 2015.

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

[119] Z. Ding, Z. Zhao, M. Peng, and H. V. Poor, "On the spectral efficiency and security enhancements of NOMA assisted multicast-unicast streaming," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3151–3163, July 2017.

[120] 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, October 2017.

[121] Q. Sun, S. Han, C. L. I, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE International Conference on Communications (ICC)*, June 2015, pp. 2668–2673.

[122] 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, March 2017.

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

[124] ——, "Energy-efficient resource scheduling for NOMA systems with imperfect channel state information," in *Proc. IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–5.

[125] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2874–2885, December 2017.

[126] P. Wu, J. Zeng, X. Su, H. Gao, and T. Lv, "On energy efficiency optimization in downlink MIMO-NOMA," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017, pp. 399–404.

[127] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[128] 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, August 2006.

[129] F. Alavi and H. Saeedi, "Radio resource allocation to provide physical layer security in relay-assisted cognitive radio networks," *IET Communications*, vol. 9, no. 17, pp. 2124–2130, 2015.

[130] Y. Ye, *Intelior Point Algorithms: Theory and Analysis*. John Wiley and Sons, 1997.

[131] I. Polik and T. Terlaky, *Interior Point Methods for Nonlinear Optimization*. New York, NY, USA: Springer, 2010.

[132] K. Cumanan, Z. Ding, M. Xu, and H. V. Poor, "Secrecy rate optimization for secure multicast communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 8, pp. 1417–1432, December 2016.

[133] K. Cumanan, G. C. Alexandropoulos, Z. Ding, and G. K. Karagiannidis, "Secure communications with cooperative jamming: Optimal power allocation and secrecy outage analysis," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7495–7505, August 2017.

[134] M. R. Javan, N. Mokari, F. Alavi, and A. Rahmati, "Resource allocation in decode-and-forward cooperative communication networks with limited rate feedback channel," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 256–267, January 2017.

[135] K. Cumanan, R. Krishna, L. Musavian, and S. Lambotharan, "Joint beamforming and user maximization techniques for cognitive radio networks based on branch and bound method," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3082–3092, October 2010.

[136] P. Xu, K. Cumanan, Z. Ding, X. Dai, and K. K. Leung, "Group secret key generation in wireless networks: Algorithms and rate optimization," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1831–1846, August 2016.

[137] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl., Special Issue on Linear Algebra in Control, Signals and Image Processing*, pp. 193–228, November 1998.

[138] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

[139] B. Lee, D. Park, and H. Seo, "Wireless communications resource management," in *Wiley*, 2009.

[140] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. Allerton Conference on Communication, Control, and Computing*, September 1999, pp. 987–996.

[141] ——, *Optimal and suboptimal transmit beamforming*. Handbook of Antennas in Wireless Communications, CRC Press, 2001.

[142] E. Björnson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. Now Foundations and Trends, 2013, vol. 9, no. 2-3. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8187586

[143] N. Mokari, F. Alavi, S. Parsaeefard, and T. Le-Ngoc, "Limited-feedback resource allocation in heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2509–2521, April 2016.

[144] W. Dean, "Computational complexity theory," in *The Stanford Encyclopedia of Philosophy*, winter 2016 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2016.

[145] K. Cumanan, L. Musavian, S. Lambotharan, and A. B. Gershman, "SINR balancing technique for downlink beamforming in cognitive radio networks," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 133–136, February 2010.

[146] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, January 2004.

[147] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2064–2078, April 2012.

[148] 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, October 2015.

[149] P. Xu, K. Cumanan, and Z. Yang, "Optimal power allocation scheme for NOMA with adaptive rates and alpha-fairness," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2017, pp. 1–6.

[150] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[151] X. Zhang, *Matrix Analysis and Applications*. Tsinghua University Press, 2004.

[152] A. Helmy, L. Musavian, and T. Le-Ngoc, "Energy-efficient power adaptation over a frequency-selective fading channel with delay and power constraints," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4529–4541, September 2013.

[153] L. Musavian and T. Le-Ngoc, "Energy-efficient power allocation over nakagami-$m$fading channels under delay-outage constraints," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4081–4091, August 2014.

[154] Y. Wu, Y. Chen, J. Tang, D. K. C. So, Z. Xu, C. L. I., P. Ferrand, J. M. Gorce, C. H. Tang, P. R. Li, K. T. Feng, L. C. Wang, K. Borner, and L. Thiele, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 112–120, November 2014.

[155] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 686–705, 2016.

[156] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.

[157] R. Jagannathan, "On some pproperties of programming problems in parametric form pertaining to fractional programming," in *Management Science*, vol. 12, no. 7, 1966.

[158] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, March 1967.