Resource Allocation for Uplink Non-Orthogonal Multiple Access in Virtualized Wireless Networks

Daniel Tweed

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

November 2017

A thesis submitted to McGill University in partial fulfillment
of the requirements for the degree of Master of Engineering.

2017/11/17

Abstract

Wireless networks are strained by an exponential growth in mobile network traffic and new applications, such as the internet-of-things (IoT) paradigm and smart cities, are amplifying the problem as the density of networks increases. At the same time, network providers are faced with increasing infrastructure and service deployment costs which are not being offset by increased revenues. Multi-carrier non-orthogonal multiple access (NOMA) and virtualized wireless networks (VWN) are being positioned as promising techniques to jointly meet the needs of future network users and service providers by promoting the mutualization of network hardware and sharing of spectrum resources. With NOMA, sub-carriers can be shared by several users concurrently, with resulting reduction in spectrum requirements via increased spectral efficiency and re-use, increased power efficiency, and network density. Under VWN, hardware and radio resources are shared by several service providers with groups of users isolated from one another by minimum quality of service guarantees. The use of NOMA in VWNs has not been extensively studied and, due to the nature of wireless channels and user mobility, careful dynamic resource allocation is required to maintain system and user performance.

The purpose of this work is to study NOMA-based VWNs and propose efficient resource allocation algorithms to leverage the available gains for users and service and infrastructure providers. Specifically, the use of NOMA for uplink transmissions is examined due to the many proposed use-cases, such as distributed sensor networks, for which uplink traffic far outweighs downlink and the greater capability of base stations in processing concurrent user signals. Initially, performance of NOMA VWN in single-input single-output channels with perfect processing of received signals is examined. With the goal of minimizing required transmit power for battery-dependent devices, an efficient iterative algorithm is presented. The proposed algorithm is then extended to multiple-input multiple-output systems and a sensitivity analysis to increased interference from imperfect processing of received signals is presented. Since many of the proposed use-cases support critical applications such as health and public safety monitoring, we then examine the use of NOMA VWN subject to reliability constraints. The resource allocation problem is mapped to its robust counterpart and the techniques of chance-constrained robust optimization are used to develop an efficient iterative algorithm which minimizes required transmit power subject to user rate and outage constraints. In each of these scenarios, simulation results are presented demonstrating the performance of the proposed algorithms and the improvement compared to traditional orthogonal multiple access is evaluated.

Sommaire

Les réseaux sans fil subissent une croissance exponentielle du trafic des réseaux mobiles et les nouvelles applications, tels que le paradigme Internet-des-choses (IoT) et les villes intelligentes, intensifiant le problème à mesure que la densité des réseaux augmente. En plus, les fournisseurs de réseaux sont confrontés à des coûts d'infrastructure et de déploiement accrus qui ne sont pas compensés par les revenus. L'accès multiple non-orthogonal à plusieurs porteuses (NOMA) et la virtualisation des réseaux sans fils (VWN) ont été proposés pour satisfaire les besoins des utilisateurs et des fournisseurs de service en promouvant le partage des outils du réseau et des ressources spectrales. Avec la NOMA, les sous-porteuses peuvent être partagées par plusieurs utilisateurs d'une manière concurrentielle, ce qui réduit les besoins en spectre en augmentant l'efficacité spectrale et son réutilisation, augmente l'efficacité énergétique et la densité du réseau. Sous la VWN, les ressources matérielles et radio sont partagés par divers fournisseurs de service avec des groupes d'utilisateurs isolés l'un de l'autre en garantissant une qualité de service minimale. L'utilisation de NOMA dans les VWNs n'est pas largement étudiée et, à cause de la nature des canaux sans fil et la mobilité des utilisateurs, une allocation dynamique des ressources est nécessaire pour maintenir les performances du système.

L'objectif de ce travail est d'étudier les VWNs se basant sur la NOMA et de proposer des algorithmes d'allocation de ressource afin de tirer le maximum de profil pour les utilisateurs et les fournisseurs de service. Particulièrement, l'utilisation de la NOMA pour la transmission en lien montant est examinée pour couvrir plusieurs cas de figure tel que les réseaux de capteurs distribués, dans lequel le trafic en lien montant dépasse largement le trafic en lien descendant et la grande capacité des stations de bases à traiter plusieurs utilisateurs. Premièrement, nous examinant la performance de la NOMA VWN dans les canaux à une seule entrée à une seule sortie avec un traitement parfait des signaux reçus. Dans le but de minimiser la puissance de transmission pour les appareils dépendant de batteries, un algorithme itératif est proposé. L'algorithme proposé est ensuite étendu aux systèmes à plusieurs entrés à plusieurs sorties et une analyse de la sensibilité à l'interférence causé par le traitement imparfait des signaux reçus est présentée. Étant donné que de nombreux cas d'utilisation proposés prennent en charge des applications critiques telles que la surveillance de la santé et de la sécurité publique, nous examinons ensuite l'utilisation de la NOMA VWN sous réserve de contraintes de fiabilité. Le problème de l'allocation des ressources est mappé à sa contrepartie robuste et les techniques d'optimisation robustes contraintes au hasard sont utilisées pour développer un algorithme itératif efficace qui

minimise la puissance d'émission requise, soumise aux contraintes de débit des utilisateurs et de contrainte d'interruption. Dans chacun de ces scenarios, les résultats de la simulation démontrent l'amélioration en termes de performance des algorithmes proposés par rapport aux autres techniques classiques d'accès orthogonal.

Acknowledgments

My sincerest thanks to my academic supervisor, Prof. Tho Le-Ngoc, for affording me the opportunity to pursue the Master of Engineering degree at McGill University under his tutelage. His invaluable guidance and expansive knowledge and research experience has provided the foundation and impetus to complete meaningful and impactful research.

I wish to express my heartfelt gratitude to Drs. Quang-Dung Ho, Mahsa Derakhshani, and Saeedeh Parsaeefard. I am privileged to have been afforded the opportunity to collaborate with each of them, and in no small part I am a better researcher and engineer as a result.

I would also like to thank my colleagues and friends at McGill University, especially current and past members of the Broadband Communication Research Laboratory: Atoosa, Ruikai, Rajesh, Vikas, Gowdemy, Alfred, Quang, and Behnam, for their insightful and beneficial discussions about my work, their willingness to expand my horizons through discussions about their work, and for their friendship and openness in making me feel welcome from the moment I joined the group.

Finally, and most importantly, I would like to gratefully acknowledge and dedicate this thesis to my family, without whose love, support, and patience—especially patience—I could not have achieved this nor any of the important things in life.

Contents

1 Introduction **1**
1.1 Background and Motivation 1
1.2 Thesis Contributions and Organization 3

2 Current Trends in Wireless Networks and Key Enabling Technologies **6**
2.1 4G Networks: An Overview 6
2.2 5G Networks: Targets and Key Enabling Technologies 8
2.3 Virtualized Wireless Networks 9
2.4 Non-Orthogonal Multiple Access 10
2.5 VWN and NOMA: A Survey 12

3 Power and Sub-carrier Allocation for Uplink NOMA VWN **15**
3.1 Introduction 15
3.2 System Model and Problem Formulation 17
3.3 Proposed Algorithm 19
3.3.1 Sub-carrier Allocation 19
3.3.2 Power Allocation 21
3.3.3 Complexity Analysis 22
3.4 Numerical Results and Discussions 23
3.5 Extension to Multi-antenna Systems 27
3.5.1 Extended System Model 29
3.5.2 Simulation and Numerical Results 30
3.6 Chapter Summary 32

4 Resource Allocation for Uplink MIMO NOMA VWN with Imperfect SIC 33
4.1 Introduction . . . 33
4.2 System Model . . . 34
4.2.1 Signal Model . . . 34
4.2.2 Effective Power Gain . . . 36
4.3 Problem Formulation . . . 37
4.4 Proposed Algorithm . . . 38
4.4.1 Sub-carrier Allocation . . . 39
4.4.2 Power Allocation Sub-problem . . . 40
4.4.3 Complexity Analysis . . . 42
4.5 Numerical Results and Discussions . . . 43
4.6 Chapter Summary . . . 46

5 Outage Constrained Resource Allocation for Uplink NOMA VWN 48
5.1 Introduction . . . 48
5.2 System Model . . . 50
5.3 Robust Formulation with Outage Probability Analysis . . . 52
5.4 Proposed Algorithm . . . 56
5.4.1 Power Allocation . . . 56
5.4.2 Sub-carrier Allocation . . . 60
5.4.3 Complexity Analysis . . . 61
5.5 Numerical Results and Discussions . . . 62
5.6 Chapter Summary . . . 69

6 Conclusion 70
6.1 Summary . . . 70
6.2 Potential Extensions and Future Work . . . 71

Appendix : Overview of Complementary Geometric Programming 73
A.1 Preliminaries . . . 73
A.2 Geometric Programming . . . 74
A.3 Complementary Geometric Programming . . . 75

References 76

List of Figures

2.1 Power control and decoding in (a) DL and (b) UL NOMA systems 11

3.1 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$ 24
3.2 Slice average allocated sub-carriers versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$ 25
3.3 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N 26
3.4 Number of required iterations for each Algorithm versus (a) users, K, and (b) sub-carriers, N 27
3.5 Illustration of base station coverage area for short- and long-range users 28
3.6 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$, for MIMO NOMA 31

4.1 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$ 44
4.2 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N 46
4.3 Number of required iterations for each Algorithm versus (a) users, K, and (b) sub-carriers, N 47

5.1 Average UE transmit power versus reserved rate, $R^{\mathbf{rsv}}$, and Pr(outage), ϵ 63
5.2 Average UE transmit power versus reserved rate, $R^{\mathbf{rsv}}$ 64
5.3 Average UE transmit power versus Pr(outage), ϵ 65
5.4 Average user outage versus Pr(outage), ϵ 66
5.5 Average user outage versus reserved rate, $R^{\mathbf{rsv}}$ 67
5.6 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N 68

5.7 Number of required iterations each sub-problem versus (a) users, K, and (b) sub-carriers, N 69

List of Tables

2.1 ITU-A Requirements for 4G vs. LTE and LTE-A Achievements 7
2.2 ITU-A Requirements for 4G vs. ITU-2020 Requirements for 5G 8

List of Acronyms

1-5G	1st to 5th Generation cellular networks
3GPP	3rd Generation Partnership Project
AGMA	Arithmetic-Geometric Mean Approximation
AWGN	Additive White Gaussian Noise
bps	bits per second
BS	Base station
C-RAN	Cloud RAN
CDMA	Code-Division Multiple Access
CGP	Complimentary Geometric Programming
CoMP	Coordinated Multi-point
CSI	Channel State Information
dB	Decibel
dBm	Decibel-milliwatt
DC	Difference of Convex functions
DL	Downlink
eNB	Evolved NodeB
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FDD	Frequency-Division Duplexing
FDMA	Frequency-Division Multiple Access
GP	Geometric Programming
Hz	Hertz
i.i.d.	Independent and Identically Distributed
IoT	Internet-of-Things

ITU	International Telecommunications Union
LAA-LTE	License-Assisted Access LTE
LTE	Long Term Evolution
LTE-A	LTE-Advanced
M2M	Machine-to-Machine
NP	Non-deterministic, Polynomial time
M2M	Machine-to-Machine
MIMO	Multiple-Input Multiple-Output
MUST	Multi-User Superposition Transmission
NFV	Network Function Virtualization
NOMA	Non-orthogonal Multiple Access
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal FDMA
OMA	Orthogonal Multiple Access
Pr(x)	Probability of event x occuring
QoS	Quality-of-Service
RAN	Radio Access Network
RAT	Radio Access Technology
SAE	System Architecture Evolution
SCA	Successive Convex Approximation
SCALE	SCA for Low-complexity
SC-FDMA	Single-carrier FDMA
SDL	Supplemental DL
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single-Input Single-Output
SVD	Singular Value Decomposition
TDD	Time-division Duplexing
TDMA	Time-division Multiple Access
UE	User Equipment
UL	Uplink
VWN	Virtualized Wireless Network

List of Symbols

$\mathcal{S}, S$	The set and total number of slices in the VWN
$\mathcal{K}, K, K'$	The set of users, total number of users in the system, and the total number of independent data streams
$\mathcal{K}_s, \mathcal{K}'_s$	The sets of users and data streams, respectively, in slice $s \in \mathcal{S}$
$\mathcal{N}, N$	Set and total number of sub-carriers in the system
λ	Path loss exponent
M_r	Total number of receive antennas available at the BS
M_{t,k_s}	Total number of transmit antennas available to user k_s
M_t^*	Maximum number of transmit antenna amongst all users
$\alpha_{k_s,n}, \boldsymbol{\alpha}$	Sub-carrier assignment indicators: Variable for sub-carrier n to user k_s in slice s and vector for the whole VWN
$\beta_{k_s,n,m_t}, \beta_{k_s,n}, \tilde{\beta}_{k_s,n}\boldsymbol{\beta}_{k_s}, \boldsymbol{\beta}$	Power allocation indicators: Variable representing power allocated to user k_s for antenna m_t on sub-carrier n, over all transmit antenna, for data stream k_s on sub-carrier n, and vectors for user k_s, and the whole VWN, respectively
d_{k_s}	Distance of user k_s from the BS
$\mathcal{R}$	Radius of BS coverage area
r'	Radius of the boundary between short-range and long-range regions of the BS coverage area
ε	A constant with small value
$\boldsymbol{\epsilon}$	Probability of outage
$\gamma_{k_s,n}$	SINR of user k_s and on sub-carrier n
$\hat{\gamma}_{k_s,n}$	Defined as 1 + $\gamma_{k_s,n}$ to simplify notation in the outage probability constraint derivation

$h_{k_s,n}, \tilde{h}_{k_s,n}$	Channel gain of user k_s on sub-carrier n in SISO channels and the power gain of the virtual channel gain for data stream k_s on sub-carrier n in MIMO NOMA systems
$\mathbf{G}_{k_s,n}$	The channel amplitude response matrix for user k_s on sub-carrier n
$I^r_{k_s,n}$	Residual interference to the user k_s at sub-carrier n from user signals not removed by SIC
$I^e_{k_s,n}$	Interference to the user k_s at sub-carrier n from SIC residual cancellation errors
$P_{\mathbf{max}}$	Maximum uplink transmit power limit
$K_n^{\mathbf{max}}$	Maximum number of users per sub-carrier n
$R_s^{\mathbf{rsv}}$	Minimum reserved rate for slice s
$R_{k_s}^{\mathbf{rsv}}$	Minimum reserved rate for user k_s in slice s
R_s	Achieved rate of slice s
$R_{k_s,n}$	Rate of user k_s in slice s on sub-carrier n
R_{k_s}	Rate of user k_s in slice s over all sub-carriers $n \in \mathcal{N}$
$N_s^{\mathbf{rsv}}$	Minimum reserved sub-carriers for slice s
σ^2	Noise power of all users over all sub-carriers
$\sigma^2_{k_s,n}$	Noise power of user k_s on sub-carrier n
$\mathbf{z}_{k_s,n}$	Vector of AWGN over all transmit-receive antenna pairs for user k_s on sub-carrier n
σ_e^2	SIC residual cancellation error magnitude
$\sigma_{m_t,k_s,n}$	The $m_t{}^{\text{th}}$ singular value of user k_s on sub-carrier n
$\mathbf{v}_{k_s,n}$	Detection vector for user k_s on sub-carrier n

Chapter 1

Introduction

1.1 Background and Motivation

New and evolving applications are straining the capacity of current wireless networks with an exponential growth rate in data traffic and increasing traffic density demands [1]. Smart mobile devices, and the data intensive applications they support, are becoming increasingly ubiquitous and the trend is towards an Internet-of-Things (IoT), in which devices ranging from toasters and thermostats to cars and medical equipment require network connectivity. In addition to human users, by 2020 there are forecast to be 3.2 billion connections for direct machine-to-machine (M2M) communication for devices supporting vehicle navigation systems, sensor networks for industrial, health, public safety, and asset tracking applications. Many of these devices will be low-power sensors supporting mission critical applications which require strict guarantees on quality of service (QoS). Without new strategies, existing mobile networks will be overwhelmed and user experience and the ability to support these emerging use-cases will be impaired. At the same time, cellular network operators are being faced with increasing spectrum and infrastructure costs leading to a revenue gap where strictly upgrading the radio access and back-haul networks will neither provide the necessary performance gains nor the desired returns on investment.

Supporting new data-intensive user applications as well as large-scale M2M and sensor deployments requires significant improvements in spectral efficiency, network capacity, and traffic density. To meet the needs of network operators, these improvements must balance the required performance gains against economic realities by improving network utility, user and network power efficiency, and reducing the time and cost of service and infrastructure deployment, maintenance, and upgrades. As standards organizations work towards meeting the targets set out by

the International Telecommunications Union (ITU) for fifth generation (5G) cellular networks, many promising approaches to meet these competing goals are being investigated with the primary themes being resource sharing and re-use. The principle of collaborative consumption from the so-called "Sharing Economy" model, in which access to goods and services is monetized rather than promoting ownership of the underlying goods or infrastructure, has been widely applied to transportation, travel, technology, media, and access to expert and professional services. Under the sharing paradigm of collaborative consumption, idle assets are unused value and therefore wasted value. In the context of wireless networks, the sharing economy and collaborative consumption promotes the mutualization of network and base station (BS) hardware, radio spectrum, and computing power to minimize idleness and underutilization of resources and maximize value—in this case, in terms of network utility and return on investment, traffic density, and spectral and power efficiency.

Virtualized wireless networks (VWNs) and network function virtualization (NFV) are being positioned as key enabling technologies to meet the targets set for 5G networks [2–4]. Virtualized solutions such as cloud-based radio access networks (C-RAN) and end-to-end NFV of back-haul networks can reduce deployment scale and cost for service providers by supporting multiple network operators on the same underlying hardware. Further, the mutualization of wireless resources through software-defined cognitive radio, dynamic spectrum access, and super BSs which can support multiple service providers and radio access technologies (RATs), allows infrastructure providers to dynamically support several service providers and RATs concurrently with on-demand set-up and tear-down of software-defined BS services and back-haul channels. This type of as-needed deployment of functions and services on shared infrastructure maximizes utility of hardware and network resources, which translates into increased efficiencies and reduced operating costs. Under the virtualization paradigm, resources are no longer allocated to network operators on a fixed basis regardless of traffic load or network conditions. Instead, virtual service providers, called slices in the terminology of VWN, negotiate quality of service (QoS) through service level agreements to ensure that the needs of their users can be met. Sharing allows maximum use and utility as BS resources, such as antennas and radio spectrum, and network resources, such as front- and back-haul capacity, are allocated to network operators based on real-time requirements. To enable VWN, one requirement is that slices are isolated from one another so that the QoS received by each slice is not affected by the activity of other slices. The stochastic nature of wireless channels and user mobility makes maintaining QoS particularly challenging and efficient dynamic resource allocation algorithms are required to maintain this isolation.

Traditionally, wireless networks have leveraged orthogonal resources for multiple access, i.e., separating users in time, frequency, space, or through the use of orthogonal spreading codes. In all cases, the limited orthogonal resources enforce an upper limit on the number of users which can be supported in a given area, time interval, or spectrum allocation. In a further shift towards a sharing model for future networks, researchers and standards organizations are looking to multi-carrier non-orthogonal multiple access (NOMA). In NOMA, sub-carriers can be shared by multiple users to concurrently transmit in time and frequency with techniques such as successive interference cancellation (SIC) used to resolve individual signals [5]. By grouping users on sub-carriers and allocating transmission power intelligently, NOMA is able to outperform traditional orthogonal multiple access (OMA) in both achievable system rates and transmit power efficiency. However, optimal user grouping and power minimization in NOMA systems is an NP-hard problem and performing SIC requires significant processing power which adds additional restrictions when considering battery-dependent devices [6, 7]. So far, only a limited two-user downlink (DL) form of NOMA has been included in the 3GPP standards for fourth generation (4G) networks and new efficient dynamic resource allocation algorithms are required to expand this to more users as well as to uplink (UL) transmissions [8]. Through the application of NOMA in VWN, spectrum resources can be shared by users within and across slices and RATs, with the goal of leveraging resource sharing to improve overall network utility and performance with global application of the resource sharing paradigm.

The principles of collaborative consumption embraced by VWN and NOMA will allow future networks to support the diverse new and evolving use-cases proposed for 5G networks and dynamically adapt to changing requirements as new uses are developed. Further, minimizing idleness and underutilization serves to meet the technical requirements for future networks through increased spectral efficiency and traffic density, but also increases value for service and infrastructure providers with increased power efficiency and lower cost infrastructure and service setup, deployment, and management.

1.2 Thesis Contributions and Organization

The application of the principles of collaborative consumption to wireless networks is underway with the limited inclusion of NOMA in the 3GPP standard and new network architectures such as C-RAN and super BSs expected to play a large role in 5G networks. Practical realizations which can extract the maximum benefit from such systems require addressing the various

technical challenges associated with shared hardware and spectrum resources. In this thesis, the resource management problems associated with two technologies which, by definition, rely on the principles of collaborative consumption to jointly meet the needs of users and service and infrastructure providers, are investigated. Each of NOMA and VWN, and their combined use, presents unique challenges and this thesis examines the problem from a collaborative framework and proposes efficient resource management algorithms and solutions for several scenarios.

For perspective on the problems addressed in this thesis, Chapter 2 presents relevant background on the networks and technologies discussed. First, we present an overview of existing 4G networks as well as the requirements set for 5G networks, as formalized by the ITU. Then, detailed descriptions of NOMA and VWN are presented along with a survey of current research and developments.

As discussed, VWN is seen as the preferred architecture for 5G networks and NOMA represents the low-level sharing of time and frequency among users and slices. Chapter 3 examines the resource allocation problem for NOMA systems in supporting VWN comprised of low-power devices for which UL traffic will significantly outweigh DL, e.g., distributed sensor networks. Considering the requirements of such a system, we limit our investigations to the UL case and consider both rate- and resource-based QoS with the goal of minimizing required transmit power for battery-dependent devices. An efficient iterative algorithm for sub-carrier and power allocation is presented for single-input single-output (SISO) channels. An extension of the algorithm to multi-antenna systems is also presented and simulation results for the single and multi-antenna systems compare performance against traditional orthogonal multiple access (OMA).

NOMA relies on SIC to resolve individual user signals from the superposed received signal and errors in performing SIC can significantly impact system performance. In Chapter 4, we consider the problem of increased interference from residual cancellation errors due to imperfect SIC. We extend the previously derived joint resource allocation algorithm to address the issue of imperfect SIC in UL multiple-input multiple-output (MIMO) NOMA systems and consider worst-case residual interference. Via simulation results, we present a sensitivity analysis of such systems to residual cancellation errors and the performance of the proposed algorithm for MIMO NOMA is compared to MIMO OMA.

Proposed use-cases such as health and public safety monitoring are characterized by densely deployed high-priority low-power devices which rely heavily on UL traffic and for which outage can be equivalent to system failure. To support such mission-critical applications, the additional requirement of reliability must be addressed. In Chapter 5, we apply chance-constrained robust

optimization to tackle the uncertainty from imperfect SIC in NOMA VWNs. In particular, we derive an expression for the user outage probability as a function of SIC error variance and use this result to formulate a robust joint resource allocation problem and develop a computationally tractable two-step iterative algorithm. The iterative algorithm is then evaluated via simulation against both OMA and non-robust NOMA with both perfect and imperfect SIC.

Finally, Chapter 6 presents concluding remarks and potential extensions of this research.

Chapter 2

Current Trends in Wireless Networks and Key Enabling Technologies

The work presented in this thesis addresses the requirements of future networks to share hardware and spectrum resources in order to meet the competing requirements of increased network performance and reduced cost and deployment time for service and infrastructure providers. To support the changing network landscape, use-cases, user requirements, and device capabilities, new network architectures and multiple access strategies are needed. In this chapter, we will briefly present an overview of 4G cellular networks and the requirements set for 5G networks. We then present details and a brief literature review of two key technologies which are emblematic of the sharing paradigm and are the focus of this thesis; namely, NOMA and VWN[1].

2.1 4G Networks: An Overview

Enhancements to the Long Term Evolution/System Architecture Evolution (LTE/SAE) to meet the requirements set out for 4G cellular networks are collectively known as LTE-Advanced (LTE-A), formalized by the 3rd Generation Partnership Project (3GPP) in LTE releases 10 through 13 [11, 12]. LTE itself was a logical evolution from previous generations in order to meet increasing demands for higher data rates and improved quality of service. LTE met these demands at the access level through increased spectral efficiency and improved mobility support and cell edge data rates with new multiple access techniques. The increased spectral efficiency was achieved

[1] Parts of this chapter have been presented in [9, 10]

by using orthogonal frequency-division multiple access (OFDMA) and single-carrier frequency-division multiple access (SC-FDMA) in the DL and UL, respectively. Improvements in mobility support and cell edge data rates were achieved through enhanced adaptive modulation, bandwidth selection, and MIMO support.

LTE transitioned to an all-IP packet switched core network with the introduction of the evolved packet core (EPC) and a flattened network architecture of enhanced BSs called evolved NodeB's (eNB) which are interconnected via high-speed data links. With direct connections between neighbouring cells, this architecture facilitates more effective multi-point transmission, coordination, and inter-cell interference and load management, independent of conditions in other areas of the network. Local functions can be handled amongst the affected eNBs in neighbouring cells with global functions and connections to external networks handled in the EPC.

Combined, these fundamental changes to the cellular network architecture allowed LTE networks to significantly increase user data rates and reduce control and user plane latency, connection set-up, and handover delays. However, they fell short of the requirements set out for 4G networks by the ITU, specifically in the case of peak data rates, spectral efficiency, and cell edge performance [13]. The continuing evolution which became LTE-A was finally able to achieve the necessary targets to meet the ITU requirements for 4G (ITU-A). Some important ITU requirements, and achieved performance levels for LTE and LTE-A, are highlighted in Table 2.1.

Table 2.1 ITU-A Requirements for 4G vs. LTE and LTE-A Achievements [13–16]

Description/Requirements	ITU-A	LTE	LTE-A
DL peak spectral efficiency (bps/Hz)	15	15	30
UL peak spectral efficiency (bps/Hz)	6.75	3.75	15
Min. cell edge spectral efficiency (bps/Hz)	0.04	0.024	0.04
DL Peak data rates (Mbps)	1000†	300	1000
UL Peak data rates (Mbps)	1000†	75	500
Scalable bandwidth up to (MHz)	40	20	100‡

† For low mobility with requirement of minimum 100 Mbps for speeds of up to 350 km/h.
‡ With carrier aggregation of up to five carrier components.

Among other innovations, LTE-A expanded MIMO/spatial multiplexing support up to 8x8 for DL and 4x4 for UL, added coordinated multi-point (CoMP) operation to increase spectral efficiency and cell edge data rates, and improved heterogeneous network planning with the enhancement of support for small cells and relay nodes to increase area coverage with reduced

power requirements. LTE-A extended bandwidth scalability in LTE by supporting carrier aggregation, both within and across frequency bands. Discontiguous aggregation is supported to ensure a higher bandwidth is available for providers who cannot support it in contiguous spectrum allotments, allowing the development of license-assisted access (LAA-LTE) into the unlicensed and television whitespace bands. Backwards compatibility is maintained by using bandwidths for each carrier component which match those used in LTE. With 3GPP Release 13, LAA-LTE was introduced for supplemental DL (SDL) transmissions using a random access mechanism based on the the European Telecommunications Standards Institute (ETSI) standard [17]. Until the inclusion of LAA, LTE-A was strictly used on and designed for dedicated licensed spectrum allocations with centralized scheduling and random access was only used by user equipment (UE) which had not yet associated to an eNB [12].

2.2 5G Networks: Targets and Key Enabling Technologies

With the enhancements discussed in the previous section, LTE-A brings cellular networks into the realm of 4G, as defined by [13]. As far as LTE-A has gone, it will not be enough for 5G networks which are expected to support existing and new applications ranging from smart cities and IoT devices, requiring massive machine type communication, to self-driving vehicles and industrial automation, requiring ultra-reliable and low latency communications, with high speed mobile broadband on the order of giga*bytes* per second [18]. Some of the specific requirements for 5G networks are outlined in Table 2.2. Beyond these specific targets set by the ITU, 5G

Table 2.2 ITU-A Requirements for 4G vs. ITU-2020 Requirements for 5G [18]

Description/Requirements	ITU-A	ITU-2020
Peak data rates (Gbps)	1	20
Average user data rates (Mbps)	1	100
Mobility support(km/h)	350	500
Connection density (devices/km^2)	10^5	10^6
Traffic capacity (Mbits/s/m^2)	0.1	10

networks must also achieve 10$\times$ reduced latency, 3$\times$ improved spectral efficiency and 100$\times$ network energy efficiency, compared to 4G networks.

In order to move towards 5G networks, 3GPP has numerous study items underway and planned for future releases to meet the ITU requirements in [18]. These include significant

enhancements to inter- and intra-band carrier aggregation and LAA-LTE for both UL and DL to ISM bands, television white space, and other under-utilized spectrum resources, as well as multi-carrier enhancements and improved CoMP and device-to-device communications [19].

It is expected that 5G networks will depend heavily on wired and wireless network virtualization and hardware resource sharing with cloud- or fog-based radio access networks (RAN), software defined radio, and end-to-end network slicing and orchestration [20]. The requirements of improved spectral and power efficiency, increased traffic density, and reduced latency, and the goals of service and infrastructure provider costs, and service and network deployment time, require approaches which maximize the utility of network resources through resource sharing.

2.3 Virtualized Wireless Networks

Virtualization, or the "X-as-a-Service" model, has existed in computing and networking for decades as virtual machines, virtual disks, virtual private networks, to name a few. In recent years this paradigm has been increasingly applied to wireless networks and VWN is specifically positioned as a key enabling technology in support of 5G targets [2, 4, 21, 22].

In the terminology of virtualization, each share of resources, or the service provider they serve, is called a slice. As in other instances of virtualization, VWN increases network utility and power efficiency by allowing different slices to each access a share of physical hardware infrastructure and resources, such as BSs, antennas, and sub-carriers. Further, infrastructure and the supported services can be decoupled so that slice instances of virtual network operator can be created at a given BS to serve users on an as needed basis; adding cost and power efficiencies for infrastructure and service providers. Such sharing and decoupling can also allow greater flexibility in service customization, finer QoS management, and more efficient infrastructure utilitization.

Different levels of virtualization are possible in wireless networks and encompasses infrastructure virtualization, as in wired networks, but can also include spectrum sharing, software-defined radio, and radio-air interface virtualization [2]. One important feature of VWN is that slices must be both logically and effectively isolated from one another so that the service received is not affected by the activities of other slices. Slice isolation is achieved through negotiated QoS guarantees, for example, minimum achieved slice or user rates, minimum dedicated system resources, such as antennas or sub-carriers, or maximum slice or user singal-to-interference plus noise ratio (SINR) or rate outage probability. Maintaining isolation over time-varying channels

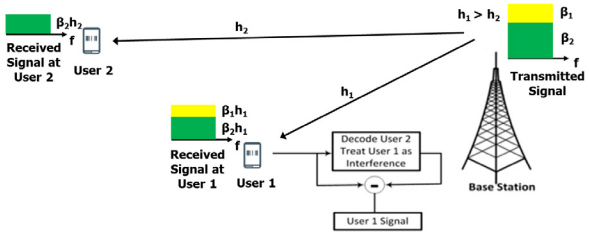
for mobile users is particularly difficult and dynamic resource allocation mechanisms are needed to achieve the desired gains in VWN.

2.4 Non-Orthogonal Multiple Access

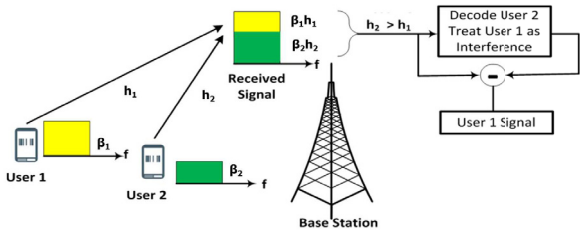
Current cellular networks rely on a variety of OMA schemes to serve users. One limitation of OMA, such as frequency-division multiple access (FDMA), code division multiple access (CDMA), or orthogonal frequency division multiple access (OFDMA), which are used in 2G, 3G, and 4G, networks, respectively, is the limited availability of orthogonal resources. In the case of CDMA, the are only a finite number of orthogonal codes of a given length, and the code length must be limited so signals can be processed efficiently. In the case of FDMA and OFDMA, the frequency spectrum is both limited and costly. Although unlicensed alternatives such as LAA-LTE have been proposed, reliability can suffer in the unlicensed bands where random access techniques are required. Recently, NOMA schemes have been positioned as a way to increase spectral efficiency and traffic density without necessarily expanding the available spectrum, either licensed or unlicensed [5, 23–25].

In contrast to OMA techniques, under NOMA several users can concurrently share available sub-carriers and user signals are resolved via SIC applied at the receiver. Broadly, existing approaches to NOMA can be divided into those that use the code-domain to separate user signals and those that use the power-domain. In contrast to CDMA, code-domain NOMA uses sparse or non-orthogonal spreading codes with low cross-correlation to control interference. In power-domain NOMA, the focus of this thesis, users are multiplexed in the power domain via power control to maintain distinctness of superposed signals and allow each transmitted signal to be resolved by the receiver, though this is handled differently for DL and UL transmissions. The basic system model for DL and UL NOMA and how power control is used in each is depicted in Fig. 2.1.

When first proposed for DL in [23], NOMA demonstrated gains over OMA in both spectral and power efficiency but also the ability to maintain those gains as the number of UE in the cell was increased. [6] demonstrated that DL NOMA can achieve both better outage performance and ergodic sum rate than OMA, subject to careful sub-carrier and power allocation, however, the SIC used in NOMA introduces significant complexity. This complexity of performing SIC at the receiver has so far limited its application to two-user DL systems, such as multi-user superposition transmission (MUST) which uses superposition coding to multiplex two users for SDL [8].



(a) Illustration of a single sub-carrier in a DL NOMA system with two users. Since User 1 experiences a strong channel gain, power is allocated so that User 2 is able to decode their own signal without significant interference and User 1 can decode and subtract the signal of User 2 before decoding their own.



(b) Illustration of a single sub-carrier in an UL NOMA system with two users. Power is allocated to maintain distinctness of individual signals in the superposed received signal. Since User 2 is the stronger received component, the BS decodes User 2 then subtracts the result from the received signal to resolve User 1.

Fig. 2.1 Power control and decoding in (a) DL and (b) UL NOMA systems

As shown in Fig. 2.1(a), in DL NOMA, and MUST specifically, greater power is allocated to the signal of the user with a low channel gain and the stronger user performs SIC to subtract the weaker user's signal before decoding their own.

In contrast to DL NOMA, in UL NOMA UE sharing the same channel can transmit at up to their maximum power levels and the BS performs SIC to resolve each signal component of the superposed received signal [5, 26]. Fig. 2.1(b) shows the basic system model for UL NOMA. Power levels are assigned by the BS based on the available channel state information (CSI) so that each component in the received superposed signal is distinct and resolvable for SIC to be successfully performed. Users are decoded in order from strongest to weakest channel gains, corresponding to strongest to weakest received power at the BS. With SIC performed by the BS, UL NOMA can potentially support many more users in each sub-carrier than the DL case, but the BS receives signals from all users and the intra-cell interference that is experienced by each user is a function of the CSI of all users in the cell. Further, issues of synchronization are presented by UL NOMA due to the different propagation delays and multi-path aspect of distributed mobile users. To address these challenges and achieve the gains available under UL NOMA efficient dynamic resource allocation algorithms are required.

2.5 VWN and NOMA: A Survey

With the ever-growing demands on wireless network infrastructure and operators, both VWN and NOMA have generated significant interest due to the gains over traditional networks and OMA techniques. In this section, we present a literature review of the current research developments and various considerations for each of these techniques, as well as their combined use. Both VWN and NOMA are sensitive to the stochastic nature of wireless channels and user mobility and dynamic efficient resource allocation is needed to achieve to maintain slice isolation and achieve performance gains over traditional techniques.

The complexity of performing SIC, which can overburden power and processing limited UE, has so far limited the application of NOMA to two-user DL systems and the DL case has been extensively studied with many resource allocation techniques proposed. In [7], the authors first provide a proof of the NP-hardness of the NOMA power minimization problem and subsequently develop an algorithm to solve a relaxed form which is convex and can be solved by standard optimization tools. In [27], an algorithm to derive jointly optimal power and sub-carrier allocation maximizing system throughput is presented based on monotonic optimization. Further, the authors develop a suboptimal solution based on successive convex approximation (SCA) and variable relaxation which achieves close to optimal performance. A two-tier approach to resource allocation maximizing sum-rate in DL NOMA systems was proposed in [28]. To solve the NP-hard problem, the authors decompose it into separate sub-carrier and power allocation problems which are repeatedly solved until the power allocation converges. Optimal sub-carrier allocation is derived using an integer linear programming approach. For power allocation, the authors present several algorithms, based on successive approximation for low-complexity (SCALE), arithmetic-geometric mean approximations (AGMA) and difference of convex functions (DC), and prove the convergence of each to a local optimal solution.

UL NOMA has been less extensively studied, but the work in [29] presents a dynamic power allocation schemed for both UL and DL NOMA two-user systems with the resource allocation structured so that individual user rates achieved by OMA are improved under NOMA. Addressing the issue of user synchronization in UL channels for dynamic mobile users, the authors of [30] present a low-complexity asynchronous SIC which maintains the NOMA performance gain over OFDMA in terms of capacity and bit error rate. An UL power control scheme for NOMA is proposed in [26] based on a discrete power allocation with users ranked in the channel according to experienced channel gains and power allocated in discrete steps based on this rank. Simulation

results for UL NOMA systems are presented in [31] with users modelled as independent Poisson point processes to derive user rate coverage probability and mean rate coverage probability of all users. Further, a closed form for the Laplace transform of inter-/intra-cluster interference is derived for both perfect and imperfect SIC. [32] presents a performance analysis of power minimization in UL NOMA systems supporting large deployments of low power machine type communication devices.

The application of NOMA to MIMO systems is examined in [33]. Considering the DL MIMO case and fixed power allocations, new precoding and detection matrices are proposed and simulation results demonstrate better outage performance for NOMA, even for users suffering strong co-channel interference. Further, the authors examine user pairing and cognitive radio inspired power allocations which further enhance the performance gains achieved under MIMO NOMA systems and derive closed-form expressions of rate gap between NOMA and OMA in MIMO systems. In [34], proof that there exists a power allocation under which the rate gap between MIMO NOMA and MIMO OMA exists and that NOMA can achieve a higher sum ergodic capacity in MIMO systems, is presented. [35] proposes a linear beamforming technique for DL MIMO NOMA with user grouping based on spatial proximity.

Uncertainty in the system and user statistics necessary to the NOMA resource allocation has also been studied. [36] examines user grouping, power allocation, and decoding order with statistical CSI to consider the user outage balancing problem. Power allocation and decoding order to balance outage probabilities experienced by users is solved using minimum weighted success probability maximization. The derived algorithm was validated against an exhaustive search for the optimal solution and performance gains against time-division multiple access (TDMA) were presented for both continuous and discrete resource allocation scenarios. User grouping was studied by evaluating the power allocation and decoding order algorithm against random and optimal user groupings. [37] examines NOMA for the case of perfect CSI being available to all users but with only 1-bit feedback to the BS, as has been used in broadcast fading channels, used to determine the power allocation policy. A closed form expression for the probability of any user being in outage is derived and a dynamic power allocation to minimize this probability is derived considering available feedback states. [38] studies power minimization for rate maximization under outage constraints in DL NOMA systems where only average CSI is available at the BS. [39] considers worst-case CSI uncertainty in DL NOMA supporting elastic and guaranteed minimum rate users to propose a robust joint resource allocation algorithm based on SCALE for power allocation and integer non-linear programming for sub-carrier allocation.

The application of VWN to cellular networks has been of increasing interest and, though most works consider applications to LTE/LTE-A, i.e., OFDMA-based networks, VWN and NFV represent the preferred end-to-end network architecture paradigm for 5G networks [4,22]. Maintaining slice isolation over time-varying wireless channels, with mobile users in dynamic environments, is particularly challenging and many efficient dynamic resource allocation algorithm have been proposed. For example, in [40], a framework for VWN in OFDMA-based systems is proposed wherein slices bid for access to resource blocks in a Vickey-Clarke-Groves auction. Since the calculation of valuations for the auction is of high complexity, a heuristic algorithm is presented where resources are assigned to the slice which reports the highest valuation for a given resource block, until all slice demands are met or all resource blocks have been allocated. In [41], a user association and power allocation algorithm for software defined heterogeneous cellular networks based on quantum-behaved particle swarm optimization is presented. An asymptotically optimal solution to bandwidth provisioning in OFDMA-based VWN systems is derived in [42]. In [3], VWN is presented as the preferred architecture paradigm for 5G networks. An outline of a potential architecture model is presented which relies on a cognitive plane which implements functions to collect system and resource state and user requirements to assist in the coordination of resource allocation. Simulation results presented demonstrate significant improvement in spectrum efficiency and reduction of packet loss rates, delay and jitter. The architecture is refined and further results presented in [4], which also introduces the trial environment developed to test the architecture, demonstrating a doubling of spectrum efficiency and reduction in packet loss of 95% when virtualization control is used. [43] proposes a joint user association and resource allocation in multi-cell VWN based on SCA and complementary geometric programming (CGP) to solve the non-convex and NP-hard resource allocation. The application of DL NOMA to support VWNs is investigated in [44] with the assumption of perfect SIC and no restriction on the number of users per sub-carrier.

Chapter 3

Power and Sub-carrier Allocation for Uplink NOMA VWN

In this chapter, we investigate the UL resource allocation problem for VWN supported by NOMA. The proposed algorithm for power and sub-carrier allocation is derived from the non-convex optimization minimizing power subject to rate and sub-carrier reservations, for which an optimal solution is NP-hard. To develop an efficient algorithm, we decompose the optimization into separate power and sub-carrier allocation problems and propose an iterative algorithm based on SCA and CGP. Initially, we consider SISO channels and in the sequel we present an extension of the resource allocation algorithm to multi-antenna systems. The proposed algorithm is evaluated via simulation and a comparisons with traditional OMA is presented.[1].

3.1 Introduction

Sub-carrier sharing and virtualization of wireless networks, as discussed earlier, represent two promising resource sharing approaches to improve spectral and power efficiency, data rates, traffic density, and network utility while reducing infrastructure and service roll-out time and cost [2, 5]. In NOMA, shared sub-carriers can be used concurrently by multiple users leveraging the power domain to resolve individual user signals, resulting in multiplexing gains and user power savings. In VWN all wireless resources, e.g., power, sub-carriers, and BS hardware resources, can be shared by multiple slices with negotiated QoS to increase network utility and

[1] Parts of this chapter have been presented in [10].

improve spectral and power efficiencies. In both NOMA and VWN, dynamic resource allocation mechanisms are needed to ensure user and slice performance for mobile users in time-varying channels.

In this chapter, we examine the performance of UL NOMA in supporting slice isolation in VWN with both rate- and resource-based QoS. As was discussed in Section 2.4, NOMA was first proposed for DL transmissions and presented significant gains over OMA systems [23]. DL NOMA has been studied extensively and many resource allocation schemes have been proposed, however, these works are primarily limited to two-user systems to limit complexity of performing SIC, as in MUST for SDL in 3GPP Release 13 [8]. Recently, UL NOMA has received more attention as higher order multiplexing gains are possible by leveraging the increased capability of the BS to perform SIC over more users. We formalize the problem as minimizing required UE transmit power for battery dependent devices subject to slice reservations, UE power limitations, and maximum number of user per sub-carrier. For comparison, we also consider the OMA case by restricting to one user per sub-carrier. Initially, we consider resource allocation over SISO channels to simplify the derivation and in the sequel we generalize the resource allocation algorithm to multi-antenna systems. The NOMA power minimization, including sub-carrier allocation across groups of users, is computationally intractable and provably NP-hard [7]. To develop an efficient algorithm, we first decompose the problem into simpler sub-problems and iteratively solve for optimal allocations of each of the two resources, i.e., transmission power and sub-carrier assignment. Each sub-problem remains complex and in order to develop an efficient solution variable relaxation and CGP are utilized with AGMA applied to approximate non-convex constraints, as described in Appendix A.3. Via these techniques, each sub-problem is converted to the equivalent geometric programming (GP) form which can be solved with available convex optimization software packages. The sub-problems are repeatedly solved until an overall solution converges in both sub-carrier and transmit power allocations.

Simulation results are presented which demonstrate that compared to OMA, NOMA is able to provide better UE power minimization performance and support more users over a limited set of resources for both SISO and MIMO systems. Specifically, as rate or sub-carrier reservations or the ratio of users to sub-carriers in the system is increased, required UE transmit power also increases in both types of systems, however, without the flexibility under NOMA of grouping users on stronger sub-carriers, users will be pushed toward their maximum allowable transmit power under OMA and the feasibility of continuing to meet all reservations can be impaired.

The remainder of this chapter is organized as follows: Section 3.2 presents the SISO system

model and problem formulation. The derivation and definition of the proposed algorithm is provided in Section 3.3 and Section 3.4 presents simulation results and analysis. In Section 3.5, the generalization of the proposed algorithm to multi-antenna systems is presented and evaluated, followed by concluding remarks in Section 3.6

3.2 System Model and Problem Formulation

Consider a BS equipped with a single receive antenna serving a set of slices, $\mathcal{S}$, in which each slice $s \in \mathcal{S}$ has its own set of single-antenna users, $\mathcal{K}_s = \{1, \ldots, K_s\}$, comprising classes of devices with distinct traffic patterns, priorities, and service requirements. To meet these needs, each slice has negotiated QoS as a minimum reserved slice rate, $R_s^{\mathbf{rsv}}$, and a minimum reserved number of sub-carriers, $N_s^{\mathbf{rsv}}$ allocated to the slice. The total number of users in the system is $K = \sum_{s\in\mathcal{S}} K_s$ and each of the available sub-carriers, $n \in \mathcal{N} = \{1, \ldots, N\}$, can be shared by at most $K_n^{\mathbf{max}}$ users. A Rayleigh fading model is assumed, and the channel gain power between user k_s and the BS on sub-carrier n is given by

$$h_{k_s,n} = \chi_{k_s,n} d_{k_s}^{-\lambda}, \tag{3.1}$$

where $\chi_{k_s,n}$ is the random channel power coefficient which follows an exponential distribution with parameter 1; d_{k_s} is the distance between user k_s and the BS; and λ is the path-loss exponent.

In each sub-carrier n, for UL NOMA, users are ranked according to channel power gain such that $h_{1,n} > h_{2,n} > \cdots > h_{K,n}$ [5]. Then, to decode user ranked i, the BS performs SIC to remove the signals from all users whose indices are lower than i. The remaining users, whose indices are greater than i, are treated as unresolvable interference. Thus, assuming user i transmits with power $\beta_{i,n}$, the signal-to-interference-plus-noise ratio (SINR) after SIC at the BS is

$$\gamma_{i,n} = \frac{\beta_{i,n} h_{i,n}}{\sigma_{i,n}^2 + I_{i,n}^r}, \tag{3.2}$$

where $\sigma_{i,n}^2$ is the noise power and

$$I_{i,n}^r = \sum\nolimits_{j=i+1}^{K} \beta_{j,n} h_{j,n}, \tag{3.3}$$

is the residual interference from users $i < j \leq K$.

The rate for user ranked i on sub-carrier n is then

$$R_{i,n} = \log_2\Big(1 + \gamma_{i,n}\Big). \tag{3.4}$$

Defining $\alpha_{k_s,n} \in \{0, 1\}$ to be the sub-carrier allocation indicator, where $\alpha_{k_s,n} = 1$ means that sub-carrier n is allocated to user k_s, the sum rate achieved by each slice s can be expressed as

$$R_s = \sum\nolimits_{k_s \in \mathcal{K}_s} \sum\nolimits_{n \in \mathcal{N}} \alpha_{k_s,n} R_{i,n}. \tag{3.5}$$

Thus, the minimum UE transmit power to meet slice and system constraints can be formalized in the following optimization problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \beta_{k_s,n}.$$

Subject to: (3.6)

C1: $R_s \geq R_s^{\mathbf{rsv}}$ $, \forall s \in \mathcal{S},$

C2: $\sum\nolimits_{n \in \mathcal{N}} \sum\nolimits_{k_s \in \mathcal{K}_s} \alpha_{k_s,n} \geq N_s^{\mathbf{rsv}}$ $, \forall s \in \mathcal{S},$

C3: $\sum\nolimits_{n \in \mathcal{N}} \beta_{k_s,n} \leq P_{\mathbf{rsv}}$ $, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s,$

C4: $\sum\nolimits_{s \in \mathcal{S}} \sum\nolimits_{k_s \in \mathcal{K}_s} \alpha_{k_s,n} \leq K_n^{\mathbf{max}}$ $, \forall n \in \mathcal{N},$

C5: $\beta_{k_s,n} - \alpha_{k_s,n} \times P_{\mathbf{max}} \leq 0$ $, \forall n \in \mathcal{N}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s.$

To maximize UE battery life, in this formulation, we are minimizing the maximum UE transmit power over sub-carrier allocation, $\boldsymbol{\alpha} = [\alpha_{k_s,n}]_{K \times N}$ and power, $\boldsymbol{\beta} = [\beta_{k_s,n}]_{K \times N}$ while meeting slice QoS reservations. Slice isolation is enforced by the reserved rate and sub-carriers in C1 and C2. Transmit power is maintained below the maximum allowable UE power level by C3 and, to limit the complexity of performing SIC, C4 restricts the number of users per sub-carrier. Finally, to ensure that power is not allocated to sub-carriers which the UE is not using, C5 forces the power allocation to zero when the sub-carrier allocation indicator is zero.

Algorithm 3.1 Iterative Power and Sub-carrier Allocation

Initialize: Set $t = 0$, $\boldsymbol{\alpha}^*(0) = [1]_{K\times N}$ and $\boldsymbol{\beta}^*(0) = [{}^{P_{\max}}/_{N}]_{K\times N}$
repeat
$t = t + 1$
Step 1: Derive sub-carrier allocation $\boldsymbol{\alpha}^*(t)$ according to Algorithm 3.1.1 with input $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$
Step 2: Derive power allocation $\boldsymbol{\beta}^*(t)$ according to Algorithm 3.1.2 with input $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$
until $\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}^*(t-1)\| \leq \varepsilon_1$ and $\|\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)\| \leq \varepsilon_2$, $0 < \varepsilon_1, \varepsilon_2 \ll 1$

3.3 Proposed Algorithm

The constraint given by C1 is non-convex due to the interference terms in Eq. 3.4, which relies on both power and sub-carrier allocation. Further, C2 and C5 are also non-convex and the binary variable adds significant complexity to the problem. In fact, both the NOMA power minimization problem and multi-user sub-carrier allocation in a general OFDM system are NP-hard combinatorial problems [7, 45]. In order to efficiently solve the problem in (3.6) the iterative algorithm shown in Algorithm 3.1 is proposed.

Initially, power is evenly distributed across all sub-carriers and in Step 1, an optimal sub-carrier allocation is derived. Then in Step 2, an optimal power allocation is derived for the sub-carrier allocation found in Step 1. These steps are repeated until the solution to each sub-problem is not much different than the previous iteration. While the sub-problems are simpler than (3.6), they each remain challenging and an iterative approach is again taken. We apply variable relaxation to reduce the complexity of the problem and then use the techniques of CGP discussed in Appendix A.3 to approximate non-convex constraints applying AGMA at each iteration until the solution converges.

3.3.1 Sub-carrier Allocation

Given a fixed power allocation, we have the following residual optimization problem

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s\in\mathcal{S} \\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n} \tag{3.7}$$

Subject to: C1, C2, C4.

Algorithm 3.1.1 Sub-carrier Allocation

Require: Power allocation $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$
Initialize: Set $t_1 = 1$, $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}^*(t-1)$
repeat
Step 1: Update $\eta_{k_s,n}(t_1)$ and $\kappa_{k_s,n}(t_1)$ from (3.9), (3.11)
Step 2: Derive $\boldsymbol{\alpha}(t_1)$, according to (3.12) using CVX [46]
until $\|\boldsymbol{\alpha}(t_1) - \boldsymbol{\alpha}(t_1-1)\| \leq \epsilon_1$, otherwise set $t_1 = t_1 + 1$
return $\boldsymbol{\alpha}^*(t) = \boldsymbol{\alpha}(t_1)$

Each of C1 and C2 is non-convex and, due to the binary variable $\boldsymbol{\alpha}$, this problem suffers from high computational complexity. First, we reduce the complexity by relaxing the binary variable $\alpha_{k_s,n}$ to be continuous on the interval $[0, 1]$. Then, for each iteration, t_1, we can then approximate the non-convex constraints as convex functions by applying AGMA. When Alg. 3.1 is converged, $\boldsymbol{\alpha}$ can be recovered to binary via integer rounding, i.e., in each sub-carrier we choose the $K_n^{\mathbf{max}}$ largest $\alpha_{k_s,n} > 0.5$ and set them to 1 while the remaining elements are set to 0.

C1 can be written as $\frac{R_s^{\mathbf{rsv}}}{\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} R_{k_s,n}} \leq 1$, the left-hand side of which can be approximated by the following convex function

$$x_s(t_1) = R_s^{\mathbf{rsv}} \times \prod_{k_s\in\mathcal{K}_S}\prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1)R_{k_s,n}}{\eta_{k_s,n}(t_1)}\right)^{-\eta_{k_s,n}(t_1)}, \tag{3.8}$$

where

$$\eta_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1-1)R_{k_s,n}}{\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{k_s,n}(t_1-1)R_{k_s,n}}. \tag{3.9}$$

Similarly, we can transform C2 and define the convex function

$$y_s(t_1) = N_s^{\mathbf{rsv}} \times \prod_{k_s\in\mathcal{K}_S}\prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1)}{\kappa_{k_s,n}(t_1)}\right)^{-\kappa_{k_s,n}(t_1)}, \tag{3.10}$$

where

$$\kappa_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1-1)}{\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{k_s,n}(t_1-1)}. \tag{3.11}$$

Thus, at each iteration, t_1, we solve

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1)\beta_{k_s,n} \tag{3.12}$$

Subject to: C4

$$x_s(t_1) \leq 1 \qquad ,\forall s \in \mathcal{S}$$
$$y_s(t_1) \leq 1 \qquad ,\forall s \in \mathcal{S}.$$

as described in Algorithm 3.1.1.

3.3.2 Power Allocation

Given a sub-carrier allocation, we solve the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n} \\ &\text{Subject to: C1, C3, C5.} \end{aligned} \tag{3.13}$$

Here, C1 is non-convex due to the presence of the interference terms in the expression for $R_{k_s,n}$. If we rewrite the rate expression as

$$R_{i,n} = \log_2\Big(\frac{\sigma^2_{i,n} + I^r_{i,n} + \beta_{i,n}h_{i,n}}{\sigma^2_{i,n} + I^r_{i,n}}\Big), \tag{3.14}$$

C1 can be expressed as

$$\text{C1}: \prod_{k_s \in \mathcal{K}_S} \prod_{n \in \mathcal{N}} \Big(\frac{\sigma^2_{i,n} + I^r_{i,n}}{\sigma^2_{i,n} + I^r_{i,n} + \beta_{i,n}h_{i,n}}\Big)^{\alpha_{k_s,n}} \leq 2^{-R_s^{\mathbf{rsv}}} \qquad ,\forall s \in \mathcal{S}. \tag{3.15}$$

Applying AGMA, at iteration t_2 we can then approximate the product terms with the following convex function

$$\begin{aligned} x_{i,n}(t_2) = \big(\sigma^2_{i,n} + I^r_{i,n}(t_2)\big) \times \Big(\frac{\sigma^2}{\phi_{i,n}(t_2)}\Big)^{-\phi_{i,n}(t_2)} \quad &\times \prod_{j=i+1}^{K} \Big(\frac{\beta_{j,n}(t_2)h_{j,n}}{\mu_{j,n}(t_2)}\Big)^{-\mu_{j,n}(t_2)} \\ &\times \Big(\frac{\beta_{i,n}(t_2)h_{i,n}}{\rho_{i,n}(t_2)}\Big)^{-\rho_{i,n}(t_2)}, \end{aligned} \tag{3.16}$$

Algorithm 3.1.2 Power Allocation

Require: Sub-carrier allocation $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$
  Initialize: Set $t_2 = 1$, $\boldsymbol{\beta}(0) = \boldsymbol{\beta}^*(t-1)$
  repeat
    Step 1.1: Update $\phi_{i,n}(t_2)$, $\mu_{j,n}(t_2)$, $\rho_{i,n}(t_2)$, and $\zeta_{i,n}(t_2)$ from (3.17-3.20)
    Step 1.2: Derive $\boldsymbol{\beta}(t_2)$ according to (3.21) using CVX [46]
  until $\|\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_2-1)\| \leq \epsilon_2$, otherwise set $t_2 = t_2 + 1$
  return $\boldsymbol{\beta}^*(t) = \boldsymbol{\beta}(t_2)$

where

$$\phi_{i,n}(t_2) = {}^{\sigma^2}/_{\zeta_{i,n}(t_2)}, \tag{3.17}$$

$$\mu_{j,n}(t_2) = {}^{\beta_{j,n}(t_2-1)h_{j,n}}/_{\zeta_{i,n}(t_2)}, \tag{3.18}$$

$$\rho_{i,n}(t_2) = {}^{\beta_{i,n}(t_2-1)h_{i,n}}/_{\zeta_{i,n}(t_2)}, \tag{3.19}$$

and

$$\zeta_{i,n}(t_2) = \sigma^2 + I^r_{j,n}(t_2-1) + \beta_{i,n}(t_2-1)h_{i,n}. \tag{3.20}$$

Then at each iteration, t_2, we solve

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n}(t_2)$$

Subject to: C3, C5

$$\prod_{k_s \in \mathcal{K}_S} \prod_{n \in \mathcal{N}} [x_{i,n}(t_2)]^{\alpha_{k_s,n}} \leq 2^{-R^{\mathbf{rsv}}_s} \qquad ,\forall s \in \mathcal{S}. \tag{3.21}$$

as described in Algorithm 3.1.2.

3.3.3 Complexity Analysis

3.3.3.1 *Convergence*

Problems of the form of (3.12) and (3.21) are solved using an interior point method in CVX. According to [47], the required number of iterations (Newton steps) to solve by this method is $\log({}^{c}/_{t^0\delta})/\log(\xi)$, where c is the total number of constraints, t^0 is the initial point used by the

solver in CVX to apply the interior point method, $0 < \delta \ll 1$ is the stopping criterion, and ξ is used for updating the accuracy of the method.

For (3.12), the total number of constraints is $c_1 = 2S + N$ and for (3.21) we have $c_2 = S + K + KN$. Therefore, the required iterations for each of the sub-problems to converge is

$$\begin{cases} \frac{\log(c_1/(t_1^0\delta_1))}{\log(\xi_1)}, & \text{Algorithm 3.1.1,} \\ \frac{\log(c_2/(t_2^0\delta_2))}{\log(\xi_2)}, & \text{Algorithm 3.1.2.} \end{cases} \tag{3.22}$$

With a polynomial presentation, the number of iterations (Newton steps) grows as $\mathcal{O}(\sqrt{c})$ and we see that the complexity of Algorithm 3.1.2 is higher than Algorithm 3.1.1 and is more sensitive to K and N [47,48]. The number of iterations required to for each algorithm to converge is studied further in Section 3.4.

3.3.3.2 *Computational Complexity*

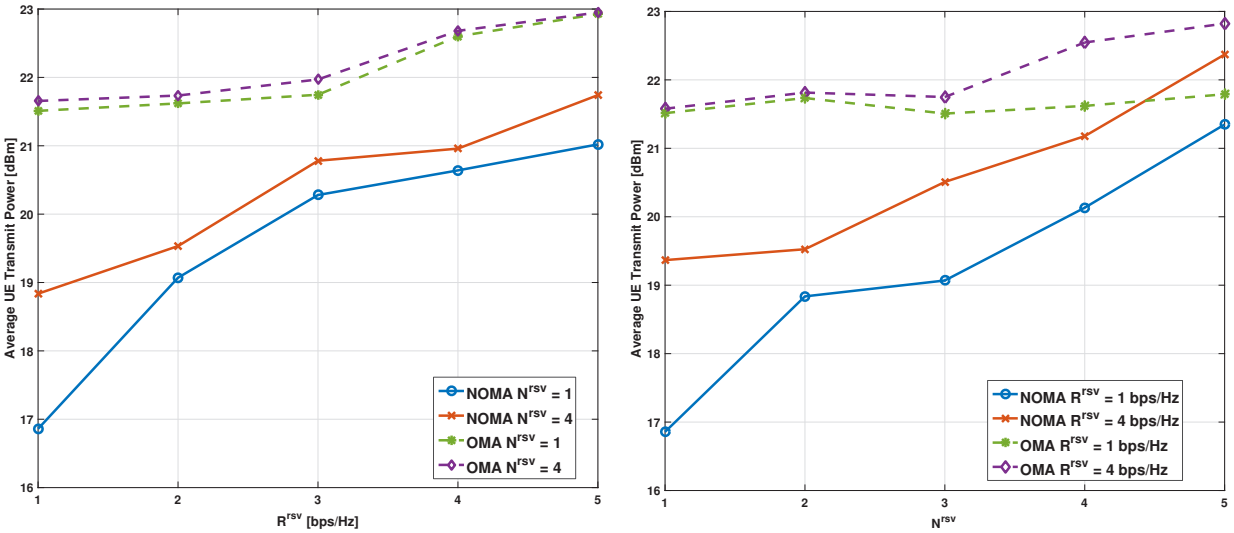
For each iteration, a Newton step of an interior-point method costs $\mathcal{O}(c\varrho^2)$ operations where c is the number of constraints and ϱ is the number of variables [48]. For both Algorithm 3.1.1 and Algorithm 3.1.2 we have $\varrho = KN$. Thus, the total computational complexity of solving the GP problem (i.e., number of Newton steps $\times$ required operations per step) is polynomial, namely,

$$C_{\text{GP}} = \begin{cases} \mathcal{O}(c_1^{1.5}\varrho^2) = \mathcal{O}(K^2N^{3.5}), & \text{Algorithm 3.1.1,} \\ \mathcal{O}(c_2^{1.5}\varrho^2) = \mathcal{O}(K^{3.5}N^{3.5}), & \text{Algorithm 3.1.2.} \end{cases} \tag{3.23}$$

In addition to the computations required to solve the GP problems, additional computational complexity is incurred by applying the AGMA approximations to convert the problem to the GP form. In Algorithms 3.1.1 and 3.1.2, the worst-case number of computations required is $i_1 = 4KN = \mathcal{O}(KN)$ and $i_2 = K^2N + 4KN = \mathcal{O}(K^2N)$, respectively, each of which is of lower order than C_{GP}. Thus, yielding an overall complexity per iteration of each of the Algorithms of $C_{\text{I}} = C_{\text{GP}}$.

3.4 Numerical Results and Discussions

To evaluate the proposed algorithm for UL NOMA, we simulate a single cell VWN with $N = 16$ sub-carriers, each of which can each be shared by at most $K_n^{\max} = 4$ users, i.e., $\forall n \in \mathcal{N}$, serving



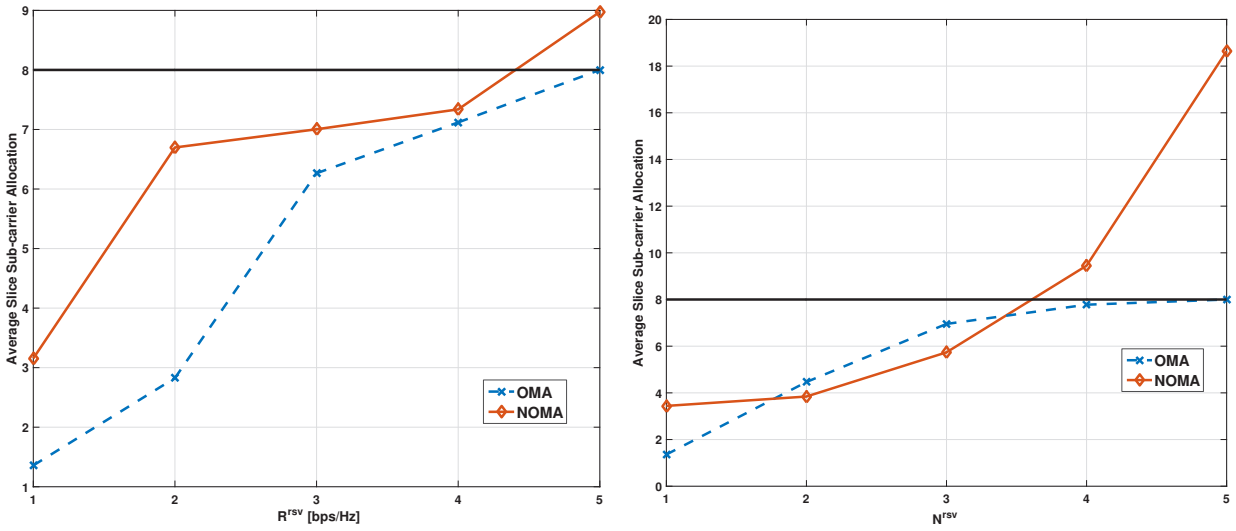
(a) (b)

Fig. 3.1 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$

two slices each with $K_s = 4$ users, for $K = 8$ total users in the system, except where otherwise noted. For comparison, results for OMA are presented where we have set $K_n^{\max} = 1, \forall n \in \mathcal{N}$ to enforce orthogonality between the sub-carriers. In all trials we have $P_{\max} = 23$ dBm, and $R_s^{\mathbf{rsv}} = R^{\mathbf{rsv}}$ and $N_s^{\mathbf{rsv}} = N^{\mathbf{rsv}}$ for all slices. The users are placed randomly within the BS coverage area following a uniform distribution and the channel power gains are derived according to the Rayleigh fading model with $\lambda = 3$ and distances normalized to the radius of the coverage area. The results shown are taken over the average of 100 channel realizations.

In Fig. 3.1 (a) the average transmit power per UE versus $R^{\mathbf{rsv}}$ for both OMA and NOMA are plotted. Power increases with $R^{\mathbf{rsv}}$ for both OMA and NOMA, because each users must transmit with higher power to satisfy the slice reservation. In both cases shown, for $N^{\mathbf{rsv}} = 1$ and 4, NOMA is more power efficient than OMA. For example, for $R^{\mathbf{rsv}} = 1$ bps/Hz, NOMA requires an average UE transmit power of 4.65 dB and 2.82 dB lower than OMA for $N^{\mathbf{rsv}} = 1$ and 4, respectively. For $R^{\mathbf{rsv}} = 3$ bps/Hz, NOMA requires an average UE transmit power of 1.47 dB and 1.19 dB lower than OMA for $N^{\mathbf{rsv}} = 1$ and 4, respectively.

For both OMA and NOMA, higher transmit power is required for higher sub-carrier reservations, $N^{\mathbf{rsv}}$, as shown in Fig. 3.1 (a), though the difference is not significant for the OMA case. To demonstrate this relationship more clearly, Fig. 3.1 (b) plots the average UE transmit

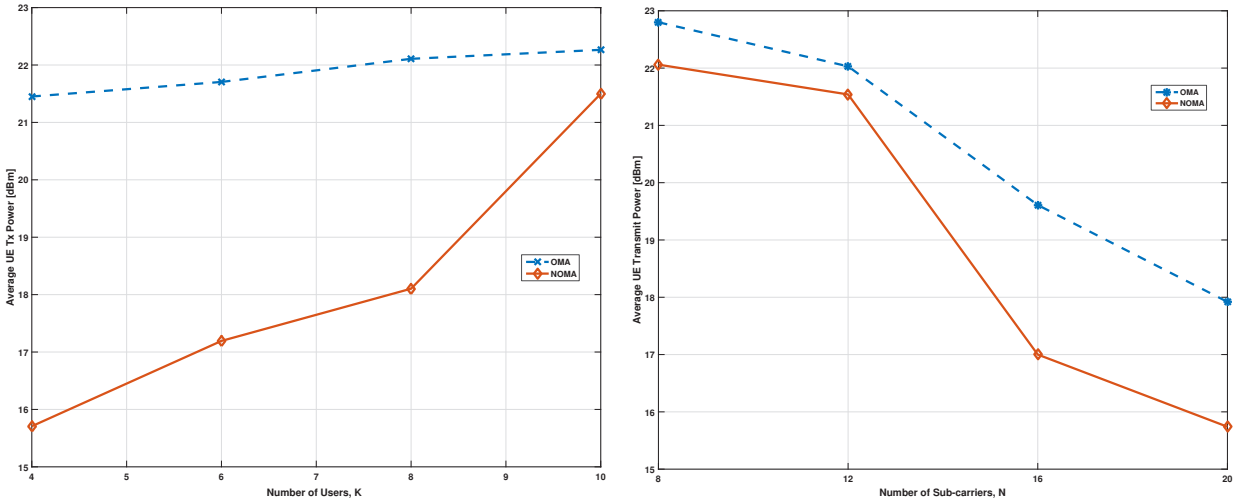


(a) (b)

Fig. 3.2 Slice average allocated sub-carriers versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$

power versus $N^{\mathbf{rsv}}$ for $R^{\mathbf{rsv}} = 1$ and 4 bps/Hz. As $N^{\mathbf{rsv}}$ increases, required average UE transmit power increases. For NOMA, lower average power can be achieved through sub-carrier sharing, however, inter-user interference from higher ranked users increases with higher sub-carrier reservations. The effects of $N^{\mathbf{rsv}}$ on OMA are not as significant as for NOMA, with only one user in each sub-carrier more sub-carriers are required by each slice to meet rate reservations, which necessitates the use of weaker sub-carriers and an already higher power than NOMA. In NOMA, power domain multiplexing allows more flexibility in sharing stronger channels and avoiding weaker ones but increasing $N^{\mathbf{rsv}}$ reduces this flexibility, especially for $N^{\mathbf{rsv}} \geq K_s$ since each slice serves K_s users. For $N^{\mathbf{rsv}} = 1$, NOMA requires an average UE transmit power of 4.65 dB and 2.21 dB lower than OMA, for $R^{\mathbf{rsv}} = 1$ and 4 bps/Hz, respectively. For $N^{\mathbf{rsv}}$ increased to 3, this gain drops to 2.44 dB and 1.24 dB, for $R^{\mathbf{rsv}} = 1$ and 4 bps/Hz, respectively.

Fig. 3.2 examines the relationships between slice reservations and the actual slice sub-carrier allocations, with varied $R^{\mathbf{rsv}}$ and $N^{\mathbf{rsv}}$. In all cases, each slice was found to have approximately equal allocations between them, so the average is shown to illustrate the trend. In each of the sub-figures, the black line, $y = 8$, shows the maximum sub-carrier allocation per slice possible under OMA, as no sharing can occur. The theoretical limit for NOMA with $K_n^{\mathbf{max}} = 4$ users to be assigned to all $N = 16$ sub-carriers is 64 in this example. In both cases, we see that for lower reservations fewer sub-carriers are used in the OMA case, but from the previous figures, we



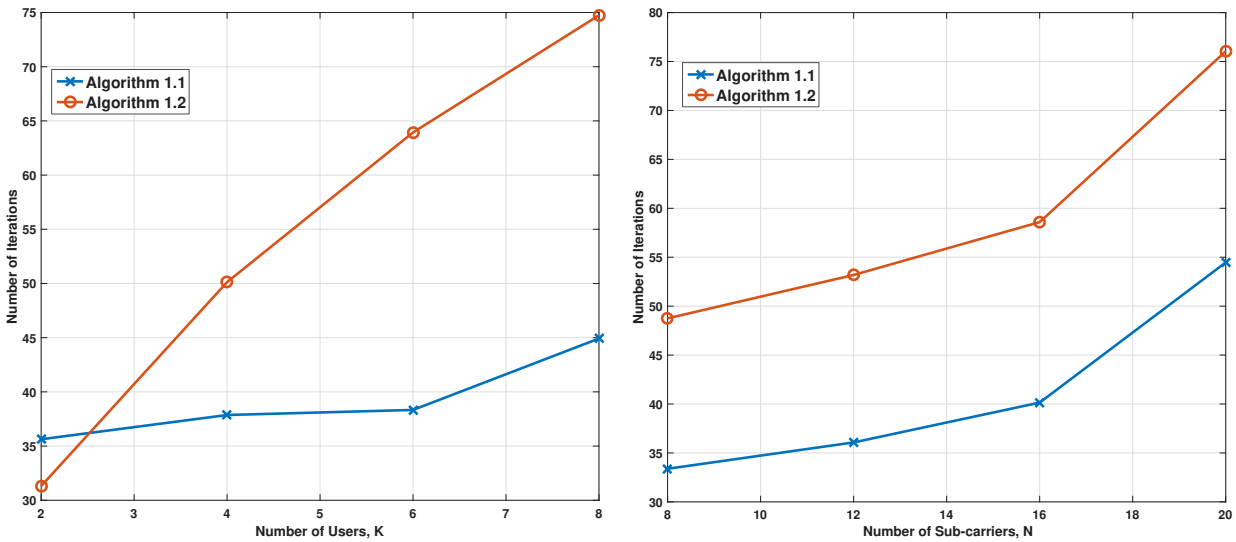
(a) (b)

Fig. 3.3 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N

note the higher power is required. Up to the point where OMA approaches its maximum, more sub-carriers may be used in NOMA to benefit from multiplexing on stronger channels. This is more clearly seen in Fig. 3.2(b), which shows allocation versus $N^{\mathbf{rsv}}$, because the sub-carrier reservation for each slice must be met which reduces the flexibility of the system.

In addition to the slice reservations, we also examined the performance impact of system density, i.e., the ratio of users to sub-carriers in the system. Fig. 3.3 (a) plots the average transmit power versus the number of users K for $N = 16$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz for OMA and NOMA. As the number of users increases, since the total number of sub-carriers is fixed, required power per user increases due to a greater reliance on shared sub-carriers. In order to resolve each user signal, higher power is needed to maintain the distinctness of components in the superposed signal received at the BS in order to perform SIC. Although the maximum number of users per sub-carrier is fixed at $K_n^{\mathbf{max}} = 4$ for all trials, as the number of users increases, more sub-carriers would be required to approach this maximum to leverage strong sub-carriers. The effect of varying the total number of sub-carriers, N, in the system for fixed number of users is depicted in Fig. 3.3 (b), for $K = 4$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. As before, we see that with decreasing system density, in this case for increasing number of sub-carriers, required average required power for both OMA and NOMA goes down. Required power under NOMA is lower than for OMA in all cases, and we noted that as the ratio of users to sub-carriers increases that users in OMA are quickly pushed towards the maximum allowable transmit power.

To examine the convergence and complexity of the proposed algorithms, the number of iterations for convergence of Algorithms 3.1.1 and 3.1.2 are plotted in Fig. 3.4(a) and 3.4(b), respectively, for $N^{\text{rsv}} = 1$ and $R^{\text{rsv}} = 1$ bps/Hz. As was noted in Section 3.3.3, as K and N increase



(a) (b)

Fig. 3.4 Number of required iterations for each Algorithm versus (a) users, K, and (b) sub-carriers, N

the required number of iterations for each algorithm to converge increases. The complexity of Algorithm 3.1.2 was found to be higher than that of Algorithm 3.1.1, which is confirmed in the required number of iterations. As expected, both algorithms are similarly impacted by increasing the number of sub-carriers, N, and increasing the number of users, K, more significantly impacts Algorithm 3.1.2.

3.5 Extension to Multi-antenna Systems

In the preceding sections, we have been considering SISO channels in our analysis of NOMA-based VWN. In this section, we present an extension of the proposed iterative algorithm to general multi-antenna systems. As we have seen, in NOMA systems user grouping and decoding order are determined by ranking users according to expected received SINR via channel gains. This is straightforward for SISO channels with a single channel gain to consider for each user but for multi-antenna systems each transmit-receive antenna pair needs to be considered. In a general

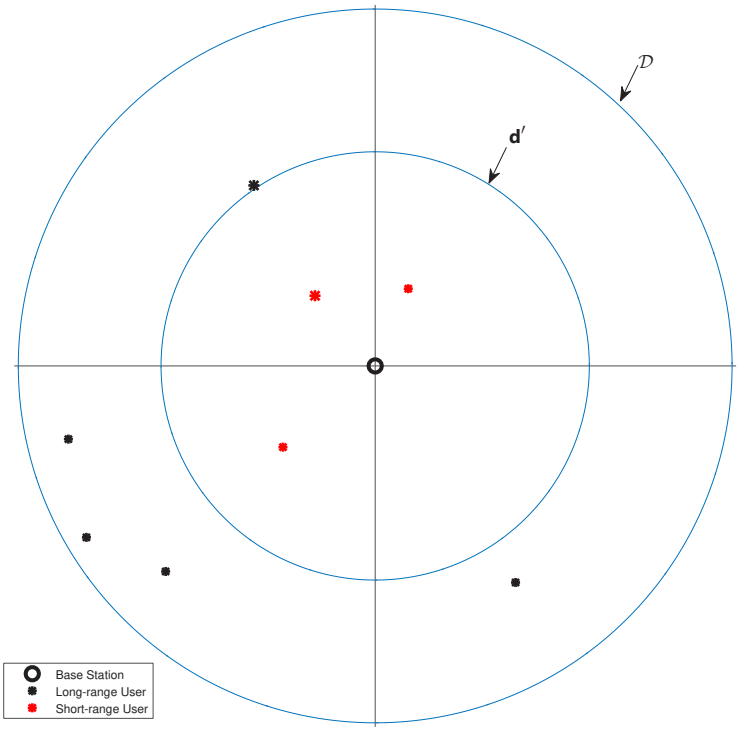


Fig. 3.5 Illustration of base station coverage area for short- and long-range users

multi-antenna system where the BS is equipped with M_r receive antennas and each user k_s is equipped with $M_{t,k_s} \geq 1$ transmit antennas we have $M_r \sum_{k_s \in K} M_{t,k_s}$ possible UL channels per sub-carrier. In general, user k_s with its $M_{t,k_s} \geq 1$ antennas can transmit $x \geq 1$ parallel independent data streams, with each data stream using n_y antennas where $\sum_{y=1}^{x} n_y \leq M_{t,k_s}$. Unfortunately, such a general approach will require impractically complex user grouping and decoding ordering operations. For simplicity, we consider the approach in which user ks will use its $M_{t,k_s} \geq 1$ antennas for one of 2 possible modes:

i Either to transmit just 1 data stream to achieve the maximum diversity gain when user k_s is far from the BS (i.e., in the cell-edge or long-range region), with relatively low average channel power gain,

ii or to transmit M_{t,k_s} independent parallel data streams to achieve the maximum multiplexing gain, when user k_s is close to the BS (i.e., in the cell-center or short-range region) with relatively higher average channel power gain.

An illustration of the long-range and short-range regions is depicted in Fig. 3.5. The multiple parallel independent data streams generated by short-range users must be treated independently in resource allocation, subject to the constraint that total power across all antennas is within the maximum power available to the user and that sub-carrier assignment groups these data streams appropriately.

3.5.1 Extended System Model

First, let $\mathbf{G}_{k_s,n} \in \mathbb{C}^{M_r \times M_{t,k_s}}$ be the MIMO channel response matrix for user k_s on sub-carrier n, each complex-valued element of which follows a Rayleigh fading model as discussed in Sec. 3.2. As noted, the channel conditions are crucial to the processing of received signals in NOMA, with grouping based on relative channel gains and SIC decoding order determined by received signal strength, both facilitated by channel statistics [5]. To simplify and facilitate user grouping and decoding, we consider the effective channel power gain after detection and combining at the receiver. With careful design of precoding and detection, the MIMO channel can be decomposed into independent parallel virtual SISO channels [33, 35, 49].

One method to design the required precoding and detection is to take the singular value decomposition (SVD) of the MIMO channel, from which we obtain

$$\mathbf{G}_{k_s,n} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{H}}, \tag{3.24}$$

where the $M_r \times M_r$ matrix $\mathbf{U}$ and the $M_{t,k_s} \times M_{t,k_s}$ matrix $\mathbf{V}$ are unitary[2] matrices used for receive separation and transmit precoding, respectively [50]. $\mathbf{\Sigma}$ is an $M_r \times M_{t,k_s}$ diagonal matrix of the singular values of $\mathbf{G}_{k_s,n}$, of which $R_{G_{k_s,n}} \leq \min(M_r, M_{t,k_s})$ are non-zero, $\sqrt{\sigma_{r,k_s,n}}$, $r = 1, 2, \ldots R_{G_{k_s,n}}$ and $\sigma_{r,k_s,n}$'s are the eigenvalues of the matrix $G^H_{k_s,n}G_{k_s,n}$ representing the power gains of the $R_{G_{k_s,n}}$ parallel virtual SISO channels after transmit precoding and receive separating, and $R_{G_{k_s,n}}$ is the rank of $G_{k_s,n}$. Therefore, user k_s with channel response matrix $G_{k_s,n}$ of rank $R_{G_{k_s,n}}$ has

- One virtual channel with power gain of $\tilde{h}_{k_s,n} = \sum_{r=1}^{R_{G_{k_s,n}}} \sigma_{r,k_s,n}$, if user k_s in the long-range region, or

[2]Unitary matrices are defined such that, for the $N \times N$ unitary matrix $\mathbf{U}$, $\mathbf{U}\mathbf{U}^{\mathbf{H}} = \mathbf{I}_N$, where $\mathbf{U}^{\mathbf{H}}$ denotes the Hermitian or conjugate transpose of $\mathbf{U}$ and $\mathbf{I}_N$ is the $N \times N$ identity matrix.

- $R_{G_{k_s,n}}$ virtual channels with power gains of $\tilde{h}_{k_s^{(r)},n} = \sigma_{r,k_s,n},\ r = 1, 2, .., R_{G_{k_s,n}}$, if user k_s in the short-range region.

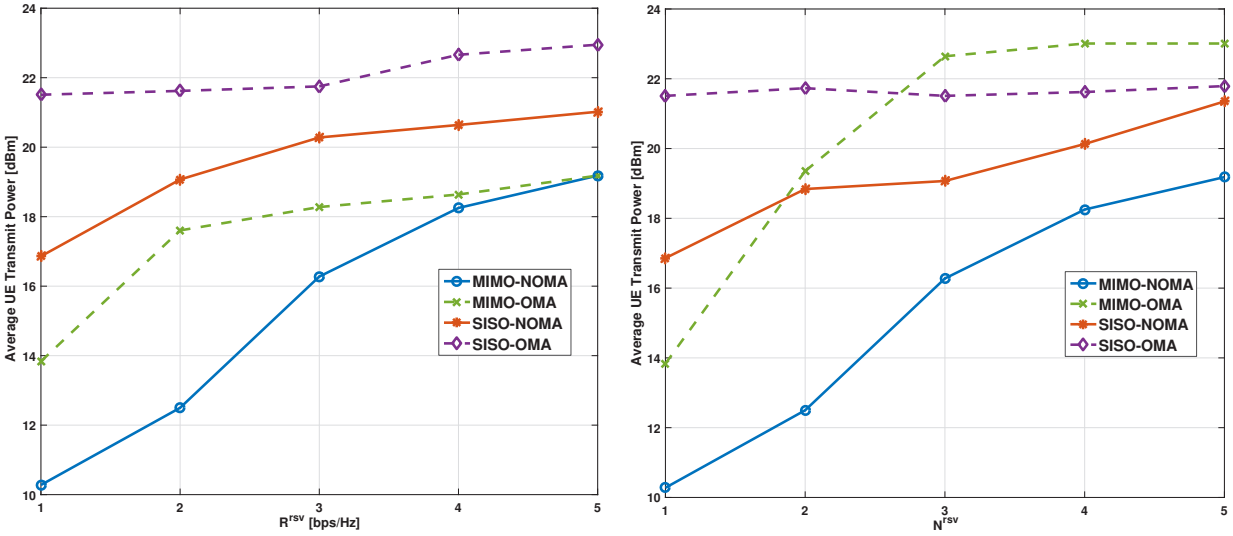
Let K' be the total number of virtual channels of all users in the system. Then, the K' virtual channels gains are sorted in the descending order as follows

$$\tilde{h}_{1,n} > \tilde{h}_{2,n} > \cdots > \tilde{h}_{K',n}.$$

3.5.2 Simulation and Numerical Results

To evaluate the proposed algorithm and extension to multi-antenna systems, we simulate a single cell VWN with $N = 8$ sub-carriers available to the BS equipped with $M_r = 4$ antennas to serve two slices each with $K_s = 2$, except where otherwise noted. Users are equipped with $M_{t,k_s} = 2$ antennas except for a randomly selected user for which $M_{t,k_s} = 1$. As before, users are randomly located in the BS coverage area following a uniform distribution and, with precoding and detection as defined in Section 3.5.1, the channel is modelled such that $||\mathbf{V}_{k_s,n}\mathbf{H}_{k_s,n}\mathbf{P}_{k_s}||^2 = \chi_{k_s,n} d_{k_s}^{-\lambda}$, where $\chi \sim$ Exp(1) and $\lambda = 3$ is the path loss exponent. The BS coverage area of radius $\mathcal{R}$ is divided into short- and long-range regions at radius $r' = 0.5\mathcal{R}$, with short-range users having $d_{k_s} <= r'$ and long-range users having $d_{k_s} > r'$. We assume that a rich scattering environment and uncorrelated antennas and thus have $R_{H_{k_s,n}} = M_{t,k_s}$ independent channels from the SVD $G_{k_s,n}$. Short-range users transmit M_{t,k_s} data streams and long-range users transmit a single data stream over all M_{t,k_s} antenna. The results shown are taken over the average of 36 channel realizations. The performance gains of NOMA versus OMA in MIMO systems is studied in more detail in Chapter 4.

The required UE transmit power as a function of slice reservations is plotted in Fig. 3.6, with the SISO data from Fig. 3.1 added for comparison. In Fig. 3.6 (a) the impact of increasing rate reservations is plotted, for $N^{\mathbf{rsv}} = 1$. As expected, we see more power is required for higher reservations under both OMA and NOMA in both SISO and MIMO. Notwithstanding that in the SISO system implementation had $N = 16$ sub-carriers and each slice had $K_s = 4$ users, and for the MIMO simulation we set $N = 8$ and $K_s = 2$, lower power is needed for the MIMO users to meet the rate reservations in all cases. For very high rate reservations, $R^{\mathbf{rsv}} \geq 4$ bps/Hz, i.e., due to the limited number of sub-carriers and users contributing to each slice's achieved rate, there is reduced feasibility of meeting the reservation for both MIMO systems and we see similar power is required. For instance, if $R^{\mathbf{rsv}} = 4$ bps/Hz, at minimum one user per slice must achieve



(a) (b)

Fig. 3.6 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$, for MIMO NOMA

greater than 2 bps/Hz. At lower reservations, the power savings with MIMO NOMA are clearly demonstrated. For example, MIMO NOMA requires 3.56 and 2.01 dB lower power than MIMO OMA, for $R^{\mathbf{rsv}} = 1$ and 3 bps/Hz, respectively.

Fig. 3.6 (b) plots the required UE transmit power versus reserved sub-carriers, $N^{\mathbf{rsv}}$, with $R^{\mathbf{rsv}} = 1$ bps/Hz. For $N^{\mathbf{rsv}} \geq 3$, we see that MIMO OMA quickly approaches the maximum allowable UE transmit power, $P_{\mathbf{max}}$ as the system is limited to $N = 8$ sub-carriers. The plot shows SISO-OMA outperforming MIMO OMA at higher reservations as a result of the different number of available sub-carriers between these two systems. However, MIMO NOMA is able to meet the slice rate and sub-carrier reservations at lower UE transmit power than both MIMO OMA or the SISO system, even when the number of reserved sub-carriers exceeds the number of available sub-carriers in the system, i.e, multiplexing of users is required. For examples, MIMO NOMA requires 3.56 and 6.36 dB lower power than MIMO OMA, for $N^{\mathbf{rsv}} = 1$ and 4, respectively.

3.6 Chapter Summary

In this chapter, we have investigated the performance of UL NOMA in supporting VWNs with the goal of minimizing required transmit power for battery dependent devices, while ensuring slice isolation and minimum system performance. We have proposed an iterative algorithm based on SCA and CGP to solve the resulting non-convex and computationally intractable resource allocation problem. Via simulation results, the performance of the proposed algorithm is evaluated for both rate and resource-based slice reservations and, compared to OMA, demonstrates significant power savings for users and the ability to support more users on fewer sub-carriers in both single- and multi-antenna systems.

Chapter 4

Resource Allocation for Uplink MIMO NOMA VWN with Imperfect SIC

In this chapter, we investigate the resource allocation problem for UL MIMO NOMA VWN in the presence of increased inter-user interference resulting from imperfections in performing SIC. We limit our investigation to the multi-antenna system developed in the previous chapter and extend the system model and proposed iterative power and sub-carrier allocation algorithm to account for SIC imperfections. Simulation are performed to demonstrate the effectiveness of the proposed algorithm and a comparison to both OMA and NOMA with perfect SIC is presented.

4.1 Introduction

The performance gains in terms of power efficiency and traffic density which are available under NOMA systems are promising and the combination of NOMA and VWN can further increase those gains through resource sharing across service providers. With NOMA relying on SIC to resolve user signals, errors in performing SIC can significantly impact system performance as residual cancellation errors increase inter-user interference on shared sub-carriers.

In this chapter, we extend the multi-antenna system model presented in Chapter 3 to evaluate MIMO NOMA VWN performance with increased inter-user interference from residual cancellation errors due to imperfect SIC. We assume SIC imperfections are random and may result from a variety of sources, including inaccurate CSI or user synchronization, to examine the performance degradation under worst-case interference levels from such imperfections. We again consider the

problem of minimizing required transmit power for battery dependent UE subject to slice reservations and system constraints and the problem remains non-convex and NP-hard. The iterative approach based on SCA and CGP used in Chapter 3 is again taken to develop an efficient power and sub-carrier allocation algorithm.

Simulation results are presented which demonstrate that compared to MIMO OMA, under all but the most extreme cases for SIC errors, MIMO NOMA provides better performance in terms of UE power minimization and support of more users over fewer sub-carriers. As this chapter represents an extension of the previous work considering NOMA without SIC errors, much of the equations are reproduced here for the convenience of the reader, with the original numbering preserved where appropriate.

The remainder of this chapter is organized as follows: Section 4.2 presents the system and signal models and defines the effective channel gain power used to simplify the MIMO NOMA resource allocation. The problem formulation is presented in Section 4.3 and the derivation and definition of the proposed algorithm is provided in Section 4.4. Section 4.5 presents simulation results and analysis, followed by concluding remarks in Section 4.6.

4.2 System Model

Consider a BS equipped with M_r receive antennas serving a set of slices $\mathcal{S}$ where each slice $s \in \mathcal{S}$ has negotiated QoS as a minimum reserved slice rate $R_s^{\mathbf{rsv}}$ and a minimum reserved number of sub-carriers $N_s^{\mathbf{rsv}}$ to meet the needs of its users. For each slice, the set of users is $\mathcal{K}_s = \{1, \ldots, K_s\}$ and each user $k_s \in \mathcal{K}_s$, placed randomly in the BS coverage area, is equipped with M_{t,k_s} transmit antennas. The total number of users in the system is $K = \sum_{s\in\mathcal{S}} K_s$ and each of the available sub-carriers, $n \in \mathcal{N} = \{1, \ldots, N\}$, each of which can be shared by at most $K_n^{\mathbf{max}}$ users.

4.2.1 Signal Model

The signals transmitted by each user can be expressed

$$\mathbf{x}_{k_s,n} = \mathbf{P}_{k_s}\tilde{\mathbf{x}}_{k_s,n}, \tag{4.1}$$

where $\tilde{\mathbf{x}}_{k_s,n} = \sqrt{\boldsymbol{\beta}_{k_s,n}}\ \mathbf{s}_{k_s,n} \in \mathbb{C}^{M_{t,k_s}\times 1}$ is the transmitted data vector, $\mathbf{s}_{k_s,n}$ the information bearing signal transmitted by user k_s on sub-carrier n with transmission power coefficients

$\boldsymbol{\beta}_{k_s,n} = [\beta_{k_s,n,1}, \beta_{k_s,n,2}, \ldots \beta_{k_s,n,M_{t,k_s}}]^T$, and $\mathbf{P}_{k_s} \in \mathbb{C}^{M_{t,k_s} \times M_{t,k_s}}$ used for precoding at the transmitter.

As discussed, for UL NOMA, users are grouped on sub-carriers and SIC is performed to resolve each of the superimposed transmissions received at the BS. In each sub-carrier n shared by K users, they can be ranked according to their experienced channel gains to determine the decoding order from strongest $i = 1$ to weakest $i = K$ received signal strength. For MIMO systems, in general, user k_s with its $M_{t,k_s} \geq 1$ antennas can transmit up to M_{t,k_s} independent data streams. Unfortunately, such a general approach will require impractically complex user grouping and decoding ordering operations and in Section 3.5 we divided the BS coverage area which into short-range and long-range regions at radius r', as depicted in Fig. 3.5. Users in the long-range region uses their $M_{t,k_s} \geq 1$ antennas to transmit just 1 data stream to achieve the maximum diversity gain with relatively low average channel power gain. Users in the short-range region transmit M_{t,k_s} independent parallel data streams to achieve the maximum multiplexing gain with relatively higher average channel power gain. We now extend the system model presented in Section 3.5.1 to consider errors from imperfect SIC.

Without loss of generality, let us consider the user ranked i in the decoding process. Signals from users ranked lower than i are removed by SIC and signals from users ranked higher than i are treated as unresolvable interference. Let $\mathbf{G}_{i,n} \in \mathbb{C}^{M_r \times M_{t,i}}$ be the MIMO channel matrix for user i and $\mathbf{V}_{i,n} \in \mathbb{C}^{M_r \times M_r}$ be the matrix used for receiver shaping and detection for this user at the BS. Thus, the observation at the BS for user i can be expressed

$$\mathbf{y}_{i,n} = \mathbf{V}_{i,n}\mathbf{G}_{i,n}\mathbf{x}_{i,n} + I^r_{i,n} + I^e_{i,n} + \mathbf{V}_{i,n}\mathbf{z}_{i,n}, \tag{4.2}$$

where $\mathbf{z}_{i,n}$ is an additive white Gaussian noise (AWGN) vector, $I^r_{i,n}$ is the interference from users which are ranked higher than i and not removed by SIC, defined as

$$I^r_{i,n} = \sum_{j=i+1}^{K} \mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{x}_{j,n}, \tag{4.3}$$

and $I^e_{i,n}$ is residual interference from cancellation errors in performing SIC to remove signals from users ranked lower then i, which we define as

$$I^e_{i,n} = \sum_{j=1}^{i-1} ||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}||\mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{x}_{j,n}. \tag{4.4}$$

Here $||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}||$ is residual interfering signal power due to differences between the actual and estimated signals for user j. In the case of perfect SIC, $\hat{\mathbf{x}}_{j,n} = \mathbf{x}_{j,n}$ and then $I^e_{i,n} = 0$ for all users. Otherwise, some portion of the received power remains as interference. The magnitude of SIC error is dependent on the type of SIC employed, the number of signals being cancelled, and channel and user mobility conditions. To account for all sources of error, we define the expected level of cancellation achieved by SIC as $\sigma_e^2 = \mathbf{E}\left[||\mathbf{x}_{j,n} - \hat{\mathbf{x}}_{j,n}||\right]$, e.g. 20 dB cancellation is equivalent to $\sigma_e^2 = 0.01$ [51]. Thus, we have

$$I^e_{i,n} = \sigma_e^2 \sum_{j=1}^{i-1} \mathbf{V}_{i,n}\mathbf{G}_{j,n}\mathbf{P}_j\tilde{\mathbf{x}}_{j,n}. \tag{4.5}$$

4.2.2 Effective Power Gain

As described in Sec. 3.5.1, to simplify the user grouping and decoding order problem we consider the effective channel with precoding and detection vectors derived the SVD of the MIMO channel. The effective channel power gains used are reproduced here for reader convenience.

The SVD of the MIMO channel response matrix yields,

$$\mathbf{G}_{k_s,n} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{H}}, \tag{4.6}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices used for receive separation and transmit precoding, respectively. The non-zero elements of the rank $R_{G_{k_s,n}} \leq \min(M_r, M_{t,k_s})$ matrix $\mathbf{\Sigma}$ are denoted $\sqrt{\sigma_{r,k_s,n}},\ r = 1, 2, \ldots R_{G_{k_s,n}}$, and $\sigma_{r,k_s,n}$ are the eigenvalues of the matrix $G^H_{k_s,n}G_{k_s,n}$ which represent the power gains of the $R_{G_{k_s,n}}$ parallel virtual SISO channels after transmit precoding and receive separating. Therefore, the user ranked i with channel response matrix $G_{i,n}$ of rank $R_{G_{i,n}}$ has

- One virtual channel with power gain of $\tilde{h}_{i,n} = \sum_{r=1}^{R_{G_{i,n}}} \sigma_{r,i,n}$, if user i in the long-range region, or
- $R_{G_{i,n}}$ virtual channels with power gains of $\tilde{h}_{i^{(r)},n} = \sigma_{r,i,n},\ r = 1, 2, .., R_{G_{i,n}}$, if user i in the short-range region.

As before, the multiple parallel independent data streams generated by short-range users over their virtual channels must be treated independently in the resource allocation, subject to the constraint that total power across all antennas is within the maximum power available to the user

and that sub-carrier assignment groups these data streams appropriately. Defining K' to be the total number of virtual channels for all users in the system for sub-carrier n, virtual channels can be ranked in descending order as follows

$$\tilde{h}_{1,n} > \tilde{h}_{2,n} > \cdots > \tilde{h}_{K',n}.$$

4.3 Problem Formulation

With precoding and detection as defined for user grouping and decoding order purposes, we must consider power allocation per data stream. Letting $\tilde{\beta}_{k_s,n}$ denote the power allocation coefficient for data stream k_s on sub-carrier n we can define power for each data virtual channel as

$$\tilde{\beta}_{k_s,n} = \begin{cases} \beta_{k_s,n,m_t}, & \text{Users in the short-range region,} \\ \sum_{m_t=1}^{M_{t,k_s}} \beta_{k_s,n,m_t}, & \text{Users in the long-range region.} \end{cases} \tag{4.7}$$

Then, SINR for each data stream i can be written as

$$\tilde{\gamma}_{i,n} = \frac{\tilde{h}_{i,n}\tilde{\beta}_{i,n}}{\mathbf{z}_{i,n} + I^r_{i,n} + I^e_{i,n}}, \tag{4.8}$$

and the achieved rate per data streams can be expressed

$$R_{i,n} = \log_2\Big(1 + \tilde{\gamma}_{i,n}\Big). \tag{4.9}$$

Let us define $\alpha_{k_s,n} \in \{0, 1\}$ to be the sub-carrier allocation indicator, where $\alpha_{k_s,n} = 1$ means user k_s is allocated to sub-carrier n. In this context, $\alpha_{k_s,n}$ is applied to all the $R_{G_{k_s,n}}$ virtual channels for user k_s in sub-carrier n to ensure proper grouping. Then, defining $\mathcal{K}'_s$ to be the total number of virtual channels over all users of slice s, the sum rate achieved by each slice s is

$$R_s = \sum\nolimits_{k_s \in \mathcal{K}'_s} \sum\nolimits_{n \in \mathcal{N}} \alpha_{k_s,n} R_{i,n}, \tag{4.10}$$

Finally, we can formalize the constraint on slice rate reservation as

$$\text{C1: } R_s \geq R_s^{\mathbf{rsv}}, \qquad \forall s \in \mathcal{S}, \tag{4.11}$$

and the constraint on slice sub-carrier reservations can be expressed

$$\text{C2: } \sum\nolimits_{n\in\mathcal{N}} \sum\nolimits_{k_s\in\mathcal{K}_s} \alpha_{k_s,n} \geq N_s^{\mathbf{rsv}}, \qquad \forall s \in \mathcal{S}. \quad (4.12)$$

To respect UE transmission power restrictions over all their M_{t,k_s} antennas and limit the complexity of performing SIC at the BS we add two additional constraints as follows

$$\text{C3: } \sum\nolimits_{n\in\mathcal{N}} \sum\nolimits_{m_t=1}^{M_{t,k_s}} \beta_{k_s,n,m_t} \leq P_{\mathbf{max}}, \qquad \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \quad (4.13)$$

$$\text{C4: } \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_s} \alpha_{k_s,n} \leq K_n^{\mathbf{max}}, \qquad \forall n \in \mathcal{N}. \quad (4.14)$$

Finally, to ensure that no power is allocation to sub-carriers which the UE is not assigned, we force the power allocation to zero when the sub-carrier allocation indicator is zero

$$\text{C5: } \sum\nolimits_{m_t=1}^{M_{t,k_s}} \beta_{k_s,n,m_t} - \alpha_{k_s,n} \times P_{\mathbf{max}} \leq 0, \qquad \forall n \in \mathcal{N}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s. \quad (4.15)$$

Thus, the minimum power needed to meet slice reservations within practical system limitations is obtained by solving the following optimization problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \max_{\substack{\forall s\in\mathcal{S} \\ \forall k_s\in\mathcal{K}'_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\tilde{\beta}_{k_s,n} \quad (4.16)$$

Subject to: C1–5 .

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the $K' \times N$ matrices of $\alpha_{k_s,n}$ sub-carrier allocation indicators and $\tilde{\beta}_{k_s,n}$ user transmit power factors, respectively.

4.4 Proposed Algorithm

As in problem in (3.6), constraints C1, C2, and C5 are all non-convex in (4.16) and the complexity induced by the binary variable yields a problem which is NP-hard. In order to develop an efficient algorithm to solve (4.16), an iterative approach is again taken and the problem decomposed into separate sub-carrier and power allocation problems, as described in Algorithm 4.1, each of which can be solved iteratively using the techniques of CGP and AGMA described in Appendix A.3.

Algorithm 4.1 Iterative Power and Sub-carrier Allocation

Initialize: Set $t = 0$, $\boldsymbol{\alpha}^*(0) = [1]_{K' \times N}$ and $\boldsymbol{\beta}^*(0) = [P_{\max}/N]_{K' \times N}$

repeat

$t = t + 1$

Step 1: Derive sub-carrier allocation $\boldsymbol{\alpha}^*(t)$ according to Algorithm 4.1.1 with input $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$

Step 2: Derive power allocation $\boldsymbol{\beta}^*(t)$ according to Algorithm 4.1.2 with input $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$

until $\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}^*(t-1)\| \leq \varepsilon_1$ and $\|\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)\| \leq \varepsilon_2$, $0 < \varepsilon_1, \varepsilon_2 \ll 1$

4.4.1 Sub-carrier Allocation[1]

Given a fixed power allocation, we have the following optimization problem

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n} \tag{4.17}$$

Subject to: C1, C2, C4.

Each of C1 and C2 is non-convex and, due to the binary variable $\boldsymbol{\alpha}$, this problem suffers from high computational complexity. First we reduce the complexity by relaxing $\alpha_{k_s,n} \in [0,1]$. In implementation, $\boldsymbol{\alpha}$ can recovered to binary via integer rounding, as described in Sec. 3.3. For each iteration, t_1, we can then approximate the non-convex constraints as convex functions by applying AGMA.

C1 can again be written as $\frac{R_s^{\mathbf{rsv}}}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} R_{k_s,n}} \leq 1$, the left-hand side of which can be approximated by the following convex function

$$x_s(t_1) = R_s^{\mathbf{rsv}} \times \prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1) R_{k_s,n}}{\eta_{k_s,n}(t_1)} \right)^{-\eta_{k_s,n}(t_1)}, \tag{3.8}$$

where

$$\eta_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1 - 1) R_{k_s,n}}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1 - 1) R_{k_s,n}}. \tag{3.9}$$

[1] Other than accounting for all virtual channels in the optimization and calculation of R_s in C1, (4.17) is equivalent to (3.7) and has been reproduced here for the convenience of the reader.

Algorithm 4.1.1 Sub-carrier Allocation

Require: Power allocation $\boldsymbol{\beta}^*(t-1)$, $\boldsymbol{\alpha}^*(t-1)$
Initialize: Set $t_1 = 1$, $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}^*(t-1)$
repeat
Step 1: Update $\eta_{k_s,n}(t_1)$ and $\kappa_{k_s,n}(t_1)$ from (3.9), (3.11)
Step 2: Derive $\boldsymbol{\alpha}(t_1)$, according to (4.18) using CVX [46]
until $\|\boldsymbol{\alpha}(t_1) - \boldsymbol{\alpha}(t_1-1)\| \leq \epsilon_1$, otherwise set $t_1 = t_1 + 1$
return $\boldsymbol{\alpha}^*(t) = \boldsymbol{\alpha}(t_1)$

Similarly, we can transform C2 and define the convex function

$$y_s(t_1) = N_s^{\mathbf{rsv}} \times \prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_1)}{\kappa_{k_s,n}(t_1)} \right)^{-\kappa_{k_s,n}(t_1)}, \tag{3.10}$$

where

$$\kappa_{k_s,n}(t_1) = \frac{\alpha_{k_s,n}(t_1-1)}{\sum_{k_s \in \mathcal{K}'_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1-1)}. \tag{3.11}$$

Thus, at each iteration, t_1, we solve

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(t_1) \tilde{\beta}_{k_s,n} \tag{4.18}$$

Subject to: C4

$$x_s(t_1) \leq 1 \qquad , \forall s \in \mathcal{S},$$
$$y_s(t_1) \leq 1 \qquad , \forall s \in \mathcal{S}.$$

as described in Algorithm 4.1.1.

4.4.2 Power Allocation Sub-problem

Given a sub-carrier allocation, we solve the following optimization problem

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n} \tag{4.19}$$

Subject to: C1, C3, C5.

Here, C3 is in the proper form but C5 is a convex constraint for fixed sub-carrier assignment and C1 remains non-convex due to the presence of the interference terms in the expression for $R_{k_s,n}$. Extending the approach taken in Section 3.3.2 to account for imperfect SIC, we can rewrite the rate expression as

$$R_{i,n} = \log_2\Big(\frac{z_{i,n} + I^r_{i,n} + I^e_{i,n} + \tilde{h}_{i,n}\tilde{\beta}_{i,n}}{z_{i,n} + I^r_{i,n} + I^e_{i,n}}\Big). \tag{4.20}$$

Removing the logarithm, C1 can then be expressed $\forall s \in \mathcal{S}$ as

$$\prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} \Big(\frac{z_{i,n} + I^r_{i,n} + I^e_{i,n}}{z_{i,n} + I^r_{i,n} + I^e_{i,n} + \tilde{h}_{i,n}\tilde{\beta}_{i,n}}\Big)^{\alpha_{k_s,n}} \leq 2^{-R^{\mathbf{rsv}}_s}. \tag{4.21}$$

Then, expanding all matrix operations and calculating interference resulting from each data stream, at iteration t_2 we can approximate the product terms with the following convex function

$$\begin{aligned} x_{i,n}(t_2) = \big(z_{i,n} + I^r_{i,n}(t_2) + I^e_{i,n}(t_2)\big) \times \Big(\frac{z_{i,n}}{\phi_{i,n}(t_2)}\Big)^{-\phi_{i,n}(t_2)} \times \prod_{j=1}^{i-1} \Big(\frac{\tilde{\beta}_{j,n}(t_2)\tilde{h}_{j,n}\sigma^2_e}{\gamma_{j,n}(t_2)}\Big)^{-\gamma_{j,n}(t_2)} \\ \times \prod_{j=i+1}^{K'} \Big(\frac{\tilde{\beta}_{j,n}(t_2)\tilde{h}_{j,n}}{\mu_{j,n}(t_2)}\Big)^{-\mu_{j,n}(t_2)} \times \Big(\frac{\tilde{\beta}_{i,n}(t_2)\tilde{h}_{i,n}}{\rho_{i,n}(t_2)}\Big)^{-\rho_{i,n}(t_2)}, \end{aligned} \tag{4.22}$$

where

$$\phi_{i,n}(t_2) = \frac{z_{i,n}}{\zeta_{i,n}(t_2)} \tag{4.23}$$

$$\gamma_{j,n}(t_2) = \frac{\tilde{\beta}_{j,n}(t_2 - 1)\tilde{h}_{j,n}\sigma^2_e}{\zeta_{i,n}(t_2)}, \tag{4.24}$$

$$\mu_{j,n}(t_2) = \frac{\tilde{\beta}_{j,n}(t_2 - 1)\tilde{h}_{j,n}}{\zeta_{i,n}(t_2)}, \tag{4.25}$$

$$\rho_{i,n}(t_2) = \frac{\tilde{\beta}_{i,n}(t_2 - 1)\tilde{h}_{i,n}}{\zeta_{i,n}(t_2)}, \tag{4.26}$$

$$\zeta_{i,n}(t_2) = z_{i,n} + I^r_{j,n}(t_2 - 1) + I^e_{j,n}(t_2 - 1) + \tilde{h}_{i,n}\tilde{\beta}_{i,n}. \tag{4.27}$$

Algorithm 4.1.2 Power Allocation

Require: Sub-carrier allocation $\boldsymbol{\alpha}^*(t)$, $\boldsymbol{\beta}^*(t-1)$
  Initialize: Set $t_2 = 1$, $\boldsymbol{\beta}(0) = \boldsymbol{\beta}^*(t-1)$
  repeat
    Step 1.1: Update $\phi_{i,n}(t_2)$, $\gamma_{j,n}(t_2)$, $\mu_{j,n}(t_2)$, $\rho_{i,n}(t_2)$, and $\zeta_{i,n}(t_2)$ from (4.23-4.27)
    Step 1.2: Derive $\boldsymbol{\beta}(t_2)$ according to (4.28) using CVX [46]
  until $\|\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_2-1)\| \leq \epsilon_2$, otherwise set $t_2 = t_2 + 1$
  return $\boldsymbol{\beta}^*(t) = \boldsymbol{\beta}(t_2)$

Then at each iteration, t_2, we solve

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}'_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \tilde{\beta}_{k_s,n}(t_2)$$

Subject to: C3, C5

$$\prod_{k_s \in \mathcal{K}'_s} \prod_{n \in \mathcal{N}} [x_{i,n}(t_2)]^{\alpha_{k_s,n}} \leq 2^{-R_s^{\mathbf{rsv}}} \quad ,\forall s \in \mathcal{S}, \tag{4.28}$$

as described in Algorithm 4.1.2.

4.4.3 Complexity Analysis

As discussed in Section 3.3.3, an interior point method is used to solved the problems in Algorithms 4.1.1 and 4.1.2 and the required number of iterations to solve by this method is given as $\frac{\log(c/t^0\delta)}{\log \xi}$ [47]. Other than the extension to account for SIC error, the problem constraints are the same in (3.6) and (4.16), and remain so in the decomposed sub-problems. The difference in complexity for (4.18) and (4.28) comes from three sources:

- With $M_{t,k_s}, M_r > 1$, the possible UL channels for each user over which we must optimize is now $M_r M_{t,k_s}$ per sub-carrier, or $N M_r M_{t,k_s}$
- For short-range users, each datastream is allocated power independently, constrained by the sum power over all antennas less than $P_{\mathbf{max}}$ and grouping, thus each datastream increases the effective number of users in the system to K'
- The inclusion of terms for imperfect SIC in the rate calculation and considering all transmit receive antenna pairs, increases further the worst case number of required computations to convert to the GP form using AGMA

For (4.18) the total number of constraints remains $c_1 = 2S + N$, as in (3.12), but considering these sources of additional complexity, for (4.28) we have to consider all real and virtual users and now have $c_2' = S + K + K'N$. Therefore, now the number of iterations required to solve each algorithm is now

$$\text{Algorithm 4.1.1:} \qquad i_1' \times \frac{\log(c_1/(t_1^0\delta_1))}{\log(\xi_1)}, \tag{4.29}$$

$$\text{Algorithm 4.1.2:} \qquad i_2' \times \frac{\log(c_2'/(t_2^0\delta_2))}{\log(\xi_2)}, \tag{4.30}$$

which still has a polynomial presentation and the number of iterations (Newton steps) grows as $\mathcal{O}(\sqrt{c})$ [47, 48].

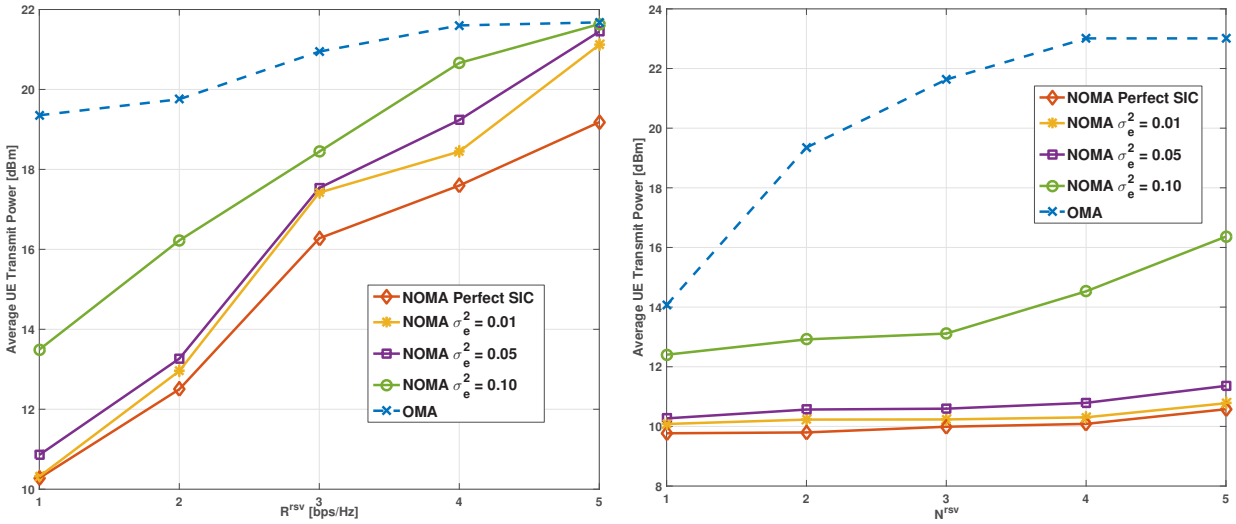
Additionally, because we must optimize over all virtual channels, the number of variables in both problems is now $\varrho = K'N$. The total computational complexity of solving the GP problems is increased but remains polynomial, namely,

$$C_{\text{GP}} = \begin{cases} \mathcal{O}(c_1^{1.5}\varrho^2) = \mathcal{O}((K')^2N^{3.5}), & \text{Algorithm 3.1.1,} \\ \mathcal{O}(c_2^{1.5}\varrho^2) = \mathcal{O}((K')^{3.5}N^{3.5}), & \text{Algorithm 3.1.2.} \end{cases} \tag{4.31}$$

Considering the possible UL virtual channels, the worst-case number of computations required to convert to the GP form using AGMA at each iteration is also increased. For 4.1.2, defining $M_t^* = \max\{M_{t,k_s}\}$, we have $i_2' = (K')^2NM_rM_t^* + 2K'N + 3K'NM_rM_t^* = \mathcal{O}((K')^2N)$. For 4.1.1, it is now dependent on K' and we have $i_1' = 4K'N = \mathcal{O}(K'N)$. In both cases, this is of lower order than C_{GP} and thus the overall complexity per iteration of each of the Algorithms is $C_{\text{I}} = C_{\text{GP}}$.

4.5 Numerical Results and Discussions

To evaluate the proposed algorithm for UL MIMO NOMA, we simulate a single cell VWN serving two slices each with $K_s = 4$ users, except where otherwise noted. The BS is equipped with $M_r = 4$ receive antennas and the $N = 16$ sub-carriers can each be shared by at most $K_n^{\mathbf{max}} = 4$ users each of which is equipped with $M_{t,k_s} = 2$, except for one random user in each slice for which $M_{t,k_s} = 1$. In all trials we have set $P_{\mathbf{max}} = 23$ dBm, and $R_s^{\mathbf{rsv}} = R^{\mathbf{rsv}}$ and $N_s^{\mathbf{rsv}} = N^{\mathbf{rsv}}$ for all slices. The users are placed randomly within the BS coverage area following a uniform



(a) (b)

Fig. 4.1 Average UE transmit power versus (a) reserved rate, $R^{\mathbf{rsv}}$, and (b) reserved sub-carriers, $N^{\mathbf{rsv}}$

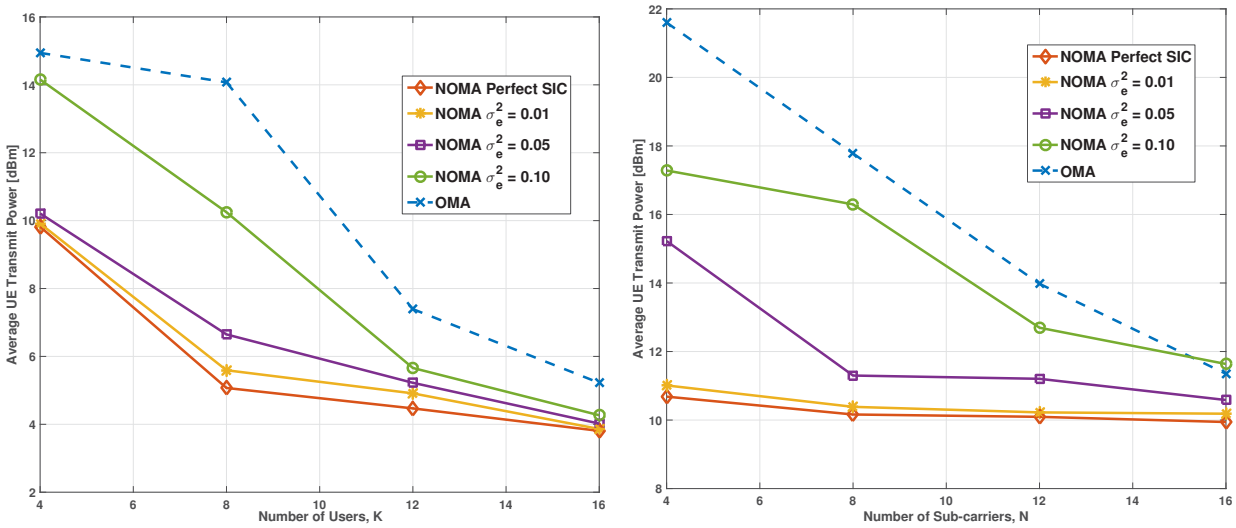
distribution with distances d_{k_s} normalized to the radius of the coverage area, $\mathcal{R}$. The coverage area is divided at radius $r' = 0.5 \times \mathcal{R}$ with short-range users having $d_{k_s} <= r'$ and long-range users having $d_{k_s} > r'$. We assume that a rich scattering environment and uncorrelated antennas and thus have $R_{H_{k_s,n}} = M_{t,k_s}$ independent channels from the SVD of the MIMO channel matrix. Short-range users transmit M_{t,k_s} data streams and long-range users transmit a single data stream over all M_{t,k_s} antenna. With precoding and detection as defined in Section 4.2, the channel is modelled as a Rayleigh fading channel such that $||V_{k_s,n} G_{k_s,n} P_{k_s}||^2 = \chi_{k_s,n} d_{k_s}^{-\lambda}$, where $\chi \sim \text{Exp}(1)$ and $\lambda = 3$ is the path loss exponent. Errors in performing SIC are introduced as residual interfering signal power with $\sigma_e^2 \in \{0.01, 0.05, 0.1\}$, i.e., up to 10% residual power remaining as inter-user interference after performing SIC at the BS, with results for perfect SIC, where $\sigma_e^2 = 0$, and for MIMO OMA, where $K_n^{\mathbf{max}} = 1$, presented for comparison. The results shown are taken over the average of 100 channel realizations.

In Fig. 4.1 (a) the average transmit power per UE is plotted versus $R^{\mathbf{rsv}}$ with $N^{\mathbf{rsv}} = 1$ for OMA and NOMA with varied levels of SIC error. Power increases with increasing $R^{\mathbf{rsv}}$ and SIC error levels since higher power is required to meet the slice rate reservation and overcome the increased inter-user interference from SIC imperfections. In all cases, NOMA is more power efficient than OMA though for $R^{\mathbf{rsv}} =$ bps/Hz and $\sigma_e^2 = 0.10$ the power savings are negligible.

For example, for $R^{\mathbf{rsv}} = 1$ bps/Hz, NOMA requires an average UE transmit power of 6.32 dB, 6.28 dB, 5.74 dB, and 3.11 dB lower than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. For $R^{\mathbf{rsv}} = 5$ bps/Hz, power savings over OMA fall to 2.15 dB, 0.56 dB, 0.23 dB, and 0.05 dB for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively.

The relationship between average transmit and slice sub-carrier reservations, $N^{\mathbf{rsv}}$ is plotted in Fig. 4.1 (b) for $R^{\mathbf{rsv}} = 1$ bps/Hz. As with increasing rate reservations, higher transmit power is required for higher sub-carrier reservations. The effect of varying $N^{\mathbf{rsv}}$ is particularly impactful for OMA which cannot avoid weak sub-carriers in favour of shared strong sub-carriers. For $\sigma_e^2 = 0.10$ we see a similarly sharp increase in transmit power for $N^{\mathbf{rsv}} > 3$ as either high levels of interference or the use of weaker sub-carriers require higher transmit power. However, even for high levels of SIC error, NOMA remains more power efficient than OMA. For example, with $N^{\mathbf{rsv}} = 1$, NOMA requires an average UE transmit power of 6.83 dB, 6.52 dB, 6.33 dB, and 4.20 dB lower than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. At $N^{\mathbf{rsv}} = 4$, NOMA is able to meet rate reservations with 11.93 dB, 11.71 dB, 11.22 dB, and 7.48 dB lower power than OMA for $\sigma_e^2 = 0, 0.01, 0.05, 0.10$, respectively. In comparison to increased rate reservations, where the optimal solution can allocate sufficient transmit power to a few users on strong sub-carriers, for OMA the results for $N^{\mathbf{rsv}} \geq 3$ are significantly impacted by reduced feasibility of finding a suitable solution as the sub-carrier reservation forces the use of weaker sub-carriers.

The performance impact of system density, i.e., the ratio of users to sub-carriers in the system, is depicted in Fig. 3.3. In Fig. 3.3 (a), the average transmit power versus the number of users K is plotted for $N = 16$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. As expected, with increasing SIC error levels required power increases as there will be cases where shared sub-carriers are needed to meet the slice rate reservation. Since the system will allow the users to be pushed into outage as long as the reserved slice rate is met, as the number of users increases each slice can leverage users with stronger sub-carriers or spread users over more sub-carriers at lower power. In general, this causes the required power to meet slice reservation decreases with more users per slice and the difference in power between the different levels of SIC error is reduced as sub-carrier sharing becomes less beneficial. The effect of varying the total number of sub-carriers, N, in the system for fixed number of users is depicted in Fig. 3.3 (b), for $K = 4$ with $N^{\mathbf{rsv}} = 1$ and $R^{\mathbf{rsv}} = 1$ bps/Hz. As before, we see that with decreasing system density, in this case for increasing number of sub-carriers, required average required power for both OMA and NOMA goes down due to the increased flexibility in ignoring weaker sub-carriers and decreased reliance on power-domain multiplexing. Required power under NOMA is lower than for OMA in general, but we note that,



(a) (b)

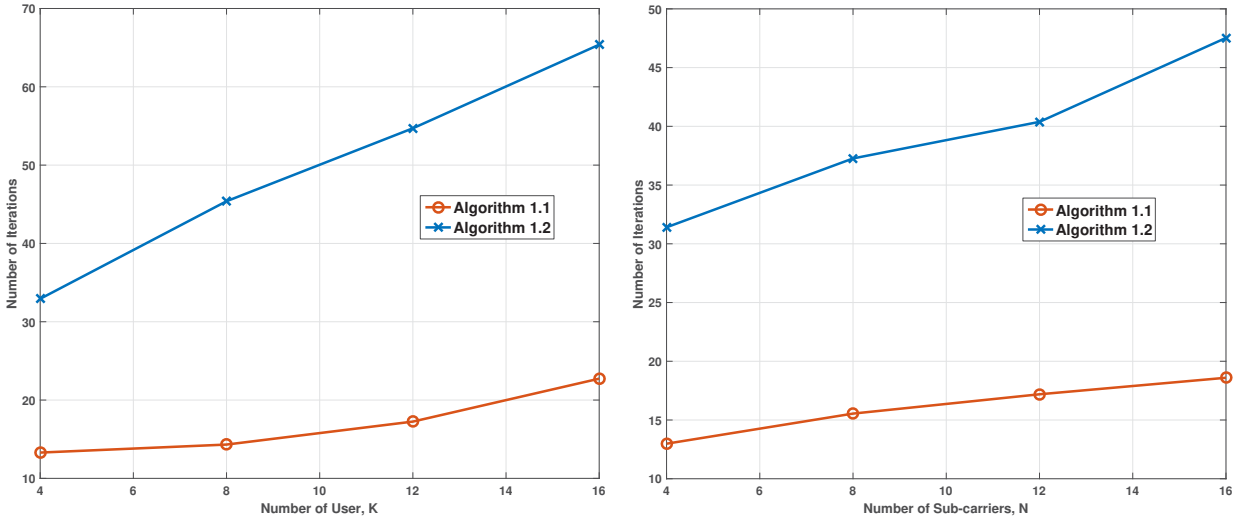
Fig. 4.2 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N

as the system density increases, for high levels of SIC error the performance gains over OMA may disappear as the optimal solution reverts to a single user per sub-carrier.

To analyse the convergence and complexity of the proposed algorithms, Fig. 4.3 shows the number of iterations for convergence of Algorithms 4.1.1 and 4.1.2 with $N^{\text{rsv}} = 1$ and $R^{\text{rsv}} = 1$ bps/Hz. In Fig. 4.3 (a) we see the required iterations as a function of the number of users, K, and in Fig. 4.3 (b) as a function of the number of sub-carriers, N. In the analysis in Section 4.4.3, it was found that the complexity of Algorithm 4.1.2 was higher than that of Algorithm 4.1.1, with Algorithm 4.1.2 more sensitive to system parameters K and N, and particularly to increasing K. This is consistent with the required number of iterations for each algorithm to converge plotted in Fig. 4.3.

4.6 Chapter Summary

In this chapter, we have investigated the performance of UL MIMO NOMA in supporting VWNs in the presence of increased inter-user interference from errors in performing SIC. With the goal of minimizing required transmit power for battery dependent devices, while ensuring slice isolation and minimum system performance, we proposed an iterative algorithm based on SCA and CGP to solve the resulting non-convex and computationally intractable resource allocation prob-



(a) (b)

Fig. 4.3 Number of required iterations for each Algorithm versus (a) users, K, and (b) sub-carriers, N

lem. Simulation results were presented validating the proposed algorithm and demonstrating that, even for high level of SIC error, MIMO NOMA outperforms MIMO OMA in terms of required user transmit power and overall system density.

Chapter 5

Outage Constrained Resource Allocation for Uplink NOMA VWN

In this chapter, we examine the NOMA resource allocation problem for VWN with power-restricted but high priority users supporting critical applications such as health and public monitoring. Such systems require high reliability and robust resource allocation techniques are needed to ensure performance. The proposed algorithm for power and sub-carrier allocation is derived from the non-convex optimization minimizing power subject to rate and probability of outage constraints. To tackle the high computational complexity of such a problem, we apply the chance-constrained robust optimization approach to reach a tractable formulation and develop an efficient iterative algorithm based on CGP and SCA. Simulations are performed to validate the proposed algorithm and a comparison to both OMA and non-robust NOMA is presented.

5.1 Introduction

The application of NOMA with imperfect SIC to support VWN was investigated in Chapters 3 and 4. System performance was examined for NOMA in the presence of increased inter-user interference from imperfect SIC and dynamic and efficient resource allocation algorithms based on CGP were presented demonstrating the performance gains available with VWN and NOMA. As was noted, many critical use-cases for IoT devices have strict QoS requirements and slice level QoS guarantees the performance for groups of users. For these types applications, user-level performance must be considered and robust optimization approaches are needed.

In this chapter, we consider the resource allocation problem for UL NOMA VWN whose users represent power-restricted but high priority devices. As noted, DL NOMA has been extensively studied and UL NOMA is of increasing interest due to the greater resources available to the BS in performing SIC over more users, but also due to the expected increase in the number of devices subject to severe power-constraints and for which UL traffic will significantly outweigh DL traffic, such as distributed sensor networks, asset tracking, and health and public safety monitoring applications. Such systems require high reliability and robust resource allocation techniques are needed to protect against device outage. Low-power sensors and devices used in asset tracking are typically equipped with a single antenna, so we limit our investigation in this chapter to UL SISO channels, though we have demonstrated a mechanism to extend resource allocation for NOMA to multi-antenna systems in the preceding chapters. In this context, we consider slices as groups of users serving a particular critical application, with QoS requirements driven by the specific needs of that application. To mathematically represent these issues of slicing and isolation in resource management problems, the minimum rate of each user per slice is preserved. However, preserving minimum rate in dynamic wireless networks can be challenging and in the context of UL NOMA, errors in performing SIC can degrade system and user performance [5].

As discussed, errors in performing SIC can result from any of several source and, as was done in Chapter 4, we assume SIC imperfections are random. To consider the sources of SIC error and protect the QoS and isolation of slices we apply the techniques of chance-constrained optimization theory, where the maximum outage probability of each critical application is kept below a predefined value [52, 53]. Since the resource allocation problem with this type of constraint suffers from significant computational complexity, we first apply the Chebyshev approximation via the Chebyshev-Cantelli inequality to reach a more tractable formulation. Then, we use this result to formulate a robust outage-constrained resource allocation problem which minimizes transmit power of users, which is highly desirable for critical applications which rely on low-power sensors, while ensuring slice isolation and outage performance. The robust problem constrains on a lower-bound on achieved rate to ensure outage performance and remains both non-convex and NP-hard [7]. To tackle this complexity, we apply the techniques of variable relaxation and CGP, discussed in Appendix A.3, to develop a computationally tractable two-step iterative algorithm which can be solved efficiently. The proposed algorithm is then evaluated in terms of user transmit power, outage performance, and system density performance.

Simulation results demonstrate that, even for high levels of SIC error variance, the proposed algorithm for NOMA outperforms the traditional OMA case in terms of both required user trans-

mit power and overall system density, i.e., serving more users over fewer sub-carriers at lower power. The outage-constrained solution necessitates a power-robustness trade-off compared to non-robust NOMA but simulation results demonstrate that, notwithstanding this trade-off that even for high levels of SIC error variance the proposed algorithm for UL NOMA can support more users at lower average transmit power on fewer sub-carriers than the corresponding OMA system.

The remainder of this chapter is organized as follows: In Section 5.2, a system model for a virtualized UL NOMA system is presented and the problem formulation based on an outage probability analysis is presented in Section 5.3. The derivation of the proposed joint resource allocation algorithm is provided in Section 5.4. Section 5.5 presents simulation results and their related analysis, followed by the conclusion in Section 5.6.

5.2 System Model

Consider a BS employing UL NOMA to support a VWN in which for each slice $s \in \mathcal{S}$ serves battery-dependent users are supporting a particular critical application. For each slice, its group of users is $\mathcal{K}_s = \{1, \ldots, K_s\}$ and the total number of users in the system is $K = \sum_{s\in\mathcal{S}} K_s$. Each slice has negotiated QoS based on the application priority and requirements of its users as a minimum rate, $R_s^{\mathbf{rsv}}$, which must be ensured for each user in supporting critical applications. The set of available sub-carriers, $\mathcal{N} = \{1, \ldots, N\}$, is shared by all K users and SIC is applied at the BS to resolve the individual signals when a given sub-carrier is used by more than one user concurrently.

As discussed, to perform SIC users are ranked in each sub-carrier with signals of lower ranking user removed by SIC and those of higher ranking users treated as unresolvable interference. Thus, for the user ranked i transmitting with power $\beta_{i,n}$, the SINR experienced for this user on sub-carrier n is

$$\gamma_{i,n} = \frac{\beta_{i,n} h_{i,n}}{\sigma_{i,n}^2 + \sum_{j=i+1}^{K} \beta_{j,n} h_{j,n} + I_{i,n}^e}, \forall i, n \tag{5.1}$$

where $\sigma_{i,n}^2$ is additive white Gaussian noise (AWGN) and $I_{i,n}^e$ is the residual interference which resulting from imperfect cancellation of the transmissions of users $1 \leq j < i$.

Defining $\hat{\gamma}_{i,n} = 1 + \gamma_{i,n}$, to simplify the derivation to follow, and $\alpha_{k_s,n} \in \{0, 1\}$ to be the sub-carrier allocation indicator, where $\alpha_{k_s,n} = 1$ means that sub-carrier n is allocated to user k_s,

the achieved rate for user k_s, whose rank i is determined independently for each sub-carrier n, is

$$R_{k_s} = \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \log\left(\hat{\gamma}\right). \tag{5.2}$$

Then, the constraint on achieved rate[1] to meet reservations can then be expressed as

$$\text{C1} : R_{k_s} \geq R_s^{\mathbf{rsv}}, \qquad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}. \tag{5.3}$$

For practicality, we limit the user transmit power to $P_{\mathbf{max}}$ and the number of users per sub-carrier is constrained to $K_n^{\mathbf{max}}$ by the following two constraints

$$\text{C2} : \sum_{n \in \mathcal{N}} \beta_{k_s,n} \leq P_{\mathbf{max}}, \qquad \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.4}$$

$$\text{C3} : \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} \alpha_{k_s,n} \leq K_n^{\mathbf{max}}, \qquad \forall n \in \mathcal{N}. \tag{5.5}$$

Further, we restrict power allocation to only those sub-carriers which are allocated to users with

$$\text{C4} : \beta_{k_s,n} - \alpha_{k_s,n} \times P_{\mathbf{max}} \leq 0, \forall n \in \mathcal{N}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s. \tag{5.6}$$

Then, the minimum power needed to meet slice reservations within practical system limitations is obtained by solving the following optimization problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \max_{\substack{\forall s \in \mathcal{S} \\ \forall k_s \in \mathcal{K}_s}} \sum_{n \in \mathcal{N}} \alpha_{k_s,n} \beta_{k_s,n}, \tag{5.7}$$

Subject to: C1–4

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the $K \times N$ matrices of $\alpha_{k_s,n}$ sub-carrier allocation indicators and $\beta_{k_s,n}$ user transmit power factors, respectively.

[1] To limit notational complexity in the derivations in Sections 5.3 and 5.4, we use $\log$ to be the natural logarithm and subsequently in this chapter user and reserved rates are expressed in units of nats/s/Hz.

5.3 Robust Formulation with Outage Probability Analysis

The optimization in (5.7) assumes that SINR can be calculated from accurate CSI and a known level of SIC residual cancellation error; however, CSI may not be accurately known and, assuming that SIC errors occur, the occurrence and magnitude of residual cancellation error is non-deterministic and depends on one or more factors including, but not limited to, the type of SIC employed, thermal and environmental noise, system parameters such as number of users, user mobility and synchronization of received signals. From (5.1), we see that residual cancellation errors from imperfect SIC will degrade the achieved SINR and may reduce users achieved rate below their reserved $R_s^{\mathbf{rsv}}$, i.e., the user will be in outage.

Residual cancellation errors from imperfect SIC results in a non-zero residual interference from cancelled signals which we model as

$$I_{i,n}^{e} = \sum_{j=1}^{i-1} \beta_{j,n} h_{j,n} \|e_{j,n}\|^2, \tag{5.8}$$

where we assume that $e_{j,n} \sim \mathcal{CN}(0, \sigma_e^2)$ and as a result $\frac{1}{\sigma_e^2}\|e_{j,n}\|^2$ is a random variable which has a chi-squared distribution with 2 degrees of freedom. The assumption on $e_{j,n}$ is made by considering the potential sources of error, i.e., thermal noise, CSI inaccuracy, and asynchronization. For each source, and others not explicitly accounted for, the magnitude of the resulting cancellation error either follows a Gaussian distribution, e.g., thermal noise and asynchronicity received signals, as shown in [30], or at worst can be assumed to be independent and identically distributed (i.i.d) for any sources of error resulting from co-channel interference of other users in the system. For each source of error, the convolution of i.i.d. random variables quickly converges to the Gaussian distribution.

To address the uncertainty in achieved SINR due to the uncertain parameter, $e_{j,n}$ in (5.8), we apply the techniques of robust optimization theory where the nominal optimization problem, i.e., (5.7), is mapped to its own robust counterpart considering the uncertain parameter [53]. Since the uncertain parameter affects the constraint of our optimization problem, i.e., C1 in (5.3), and not the objective function, we apply the chance-constrained approach where the probability of violation of the outage constraint is limited to a certain value [52]. This technique considers the expected values of the data and accepts sub-optimal solutions which remain feasible if the data changes; however, this introduces a trade-off between robustness and optimality via changing the

limit of violation of constraints [54]. To apply this approach, we reformulate C1 and consider the maximum outage probability, $0 \leq \epsilon_s \leq 1$, which has been negotiated by each slice based on service levels required by their users. Therefore, C1 can be re-expressed for all users $k_s \in \mathcal{K}_s$ as its own robust counterpart

$$\Pr\left[R_{k_s} \leq R_s^{\mathbf{rsv}}\right] \leq \epsilon_s, \tag{5.9}$$

which is equal to

$$\Pr\left[R_{k_s} \geq R_s^{\mathbf{rsv}}\right] \geq 1 - \epsilon_s. \tag{5.10}$$

Note that (5.10) can be considered as a maximum outage probability of reserved rate of each user. Depending on the SLA and the request of critical applications, these threshold can be adjusted, e.g., for highly critical mission applications, one can use very small value of ϵ_s, resulting in significant computational complexity and less optimality [52, 53]. However, one can also allow for increased efficiencies in cases when looser constraints are acceptable.

In this context, to tackle the computational complexity and reach a more tractable formulation, the constraint with uncertain parameters is relaxed and a more tractable formulation is used in place of the original one [52, 53]. Here, we apply the Chebyshev approximation using the Chebyshev-Cantelli inequality, defined as

$$\Pr\left(X - \mathbf{E}[X] \geq \eta\right) \leq \frac{\mathbf{Var}[X]}{\mathbf{Var}[X] + \eta^2}, \tag{5.11}$$

to replace (5.10). Substituting $X = R_{k_s}$ and $\eta = R_s^{\mathbf{rsv}} - \mathbf{E}[R_{k_s}]$, we have

$$\Pr\big[R_{k_s} - \mathbf{E}[R_{k_s}] \geq R_s^{\mathbf{rsv}} - \mathbf{E}[R_{k_s}]\big] \leq \frac{\mathbf{Var}[R_{k_s}]}{\mathbf{Var}[R_{k_s}] + (R_s^{\mathbf{rsv}} - \mathbf{E}[R_{k_s}])^2}\,. \tag{5.12}$$

From the inequality in (5.12), the constraint in (5.10), which does not have a closed form solution, can be relaxed to a deterministic form as

$$1 - \epsilon_s \leq \frac{\mathbf{Var}[R_{k_s}]}{\mathbf{Var}[R_{k_s}] + (R_s^{\mathbf{rsv}} - \mathbf{E}[R_{k_s}])^2}\,, \tag{5.13}$$

which can be re-arranged as

$$\mathbf{E}[R_{k_s}] + \sqrt{\mathbf{Var}[R_{k_s}]\frac{\epsilon_s}{1 - \epsilon_s}} \geq R_s^{\mathbf{rsv}}. \tag{5.14}$$

While (5.14) is more tractable than (5.10), we have to calculate the mean and variance of R_{k_s}. In order to obtain the required statistics, recall that the achieved rate on each sub-carrier is independent. Then, from (5.2), we have

$$\mathbf{E}[R_{k_s}] = \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\mathbf{E}[\log(\hat{\gamma}_{i,n})], \tag{5.15}$$

$$\mathbf{Var}[R_{k_s}] = \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\mathbf{Var}[\log(\hat{\gamma}_{i,n})]. \tag{5.16}$$

To find $\mathbf{E}[\log(\hat{\gamma}_{i,n})]$ and $\mathbf{Var}[\log(\hat{\gamma}_{i,n})]$, we will approximate the rate function with the help of the Taylor series of a logarithmic function. Approximating with two terms, we have

$$\log(\hat{\gamma}_{i,n}) \approx \log(\mathbf{E}[\hat{\gamma}_{i,n}]) + \frac{1}{\mathbf{E}[\hat{\gamma}_{i,n}]}(\hat{\gamma}_{i,n} - \mathbf{E}[\hat{\gamma}_{i,n}]). \tag{5.17}$$

We can take the expected value and variance of the both sides to obtain

$$\mathbf{E}[\log(\hat{\gamma}_{i,n})] \approx \log(\mathbf{E}[\hat{\gamma}_{i,n}]), \tag{5.18}$$

$$\mathbf{Var}[\log(\hat{\gamma}_{i,n})] \approx \frac{\mathbf{Var}[\hat{\gamma}_{i,n}]}{\mathbf{E}^2[\hat{\gamma}_{i,n}]}. \tag{5.19}$$

Now, we need to calculate $\mathbf{E}[\hat{\gamma}_{i,n}]$ and $\mathbf{Var}[\hat{\gamma}_{i,n}]$. Similarly, we use the Taylor series to approximate the one plus SINR expression as a function of $I^e_{i,n}$. With two terms, we have

$$\hat{\gamma}_{i,n}(I^e_{i,n}) \approx 1 + \frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]} - \frac{a_{i,n}}{(b_{i,n} + \mathbf{E}[I^e_{i,n}])^2}(I^e_{i,n} - \mathbf{E}[I^e_{i,n}]) \tag{5.20}$$

and consequently

$$\mathbf{E}[\hat{\gamma}_{i,n}] \approx 1 + \frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]}, \tag{5.21}$$

$$\mathbf{Var}[\hat{\gamma}_{i,n}] \approx \frac{a^2_{i,n}}{(b_{i,n} + \mathbf{E}[I^e_{i,n}])^4}\mathbf{Var}[I^e_{i,n}], \tag{5.22}$$

where, for the sake of notational simplicity, we have defined

$$a_{i,n} = \beta_{i,n}h_{i,n}\ , \qquad\qquad b_{i,n} = \sigma^2_{i,n} + \sum_{j=i+1}^{K} \beta_{j,n}h_{j,n}. \tag{5.23}$$

As a result, considering (5.15), (5.16), (5.18), (5.19), (5.21) and (5.22), we can express the mean and variance of $R_{k_s,n}$ as

$$\mathbf{E}[R_{k_s,n}] \approx \alpha_{k_s,n} \log\left(1 + \frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]}\right), \tag{5.24}$$

$$\mathbf{Var}[R_{k_s,n}] \approx \alpha_{k_s,n} \left(\frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]}\right)^2 \times \frac{\mathbf{Var}[I^e_{i,n}]}{(a_{i,n} + b_{i,n} + \mathbf{E}[I^e_{i,n}])^2}, \tag{5.25}$$

From the definition of $I^e_{i,n}$ and $e_{j,n}$ in (5.8), we then have

$$\mathbf{E}[I^e_{i,n}] = \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2, \tag{5.26}$$

$$\mathbf{Var}[I^e_{i,n}] = \sum_{j=1}^{i-1} 4a_{j,n}^2\sigma_e^4, \tag{5.27}$$

Finally, substituting (5.24) and (5.25) in (5.14), the approximation of the outage probability constraint in (5.10), $\forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}$ can be written as

$$\widetilde{\text{C1}}: \sum_{n\in\mathcal{N}} \alpha_{k_s,n} \log\left(1 + \frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]}\right) + \sqrt{\frac{\epsilon_s}{1-\epsilon_s}} \times \sqrt{\sum_{n\in\mathcal{N}} \alpha_{k_s,n} \left(\frac{a_{i,n}}{b_{i,n} + \mathbf{E}[I^e_{i,n}]}\right)^2 \frac{\mathbf{Var}[I^e_{i,n}]}{(a_{i,n} + b_{i,n} + \mathbf{E}[I^e_{i,n}])^2}} \geq R_s^{\mathbf{rsv}}. \tag{5.28}$$

Then, subject to $\widetilde{\text{C1}}$, C2–4, the chance-constrained counterpart of (5.7) is

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \max_{\substack{\forall s\in\mathcal{S} \\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n}. \tag{5.29}$$

Note that, since $\widetilde{\text{C1}}$ is not a linear constraint, we cannot apply the conventional approaches of chance-constrained optimization, such as the Bernstein approximation [53]. Therefore, in the sequel we apply the techniques of CGP discussed in Appendix A.3 to reach a more tractable formulation.

Algorithm 5.3 Iterative Power and Sub-carrier Allocation

Initialize: Set $t = t_1 = t_2 = 0$, $\boldsymbol{\beta}^*(0) = [P_{\max}/N]_{K\times N}$ and $\boldsymbol{\alpha}^*(0) = [\mathbf{1}]_{K\times N}$.
repeat
  Step 1: Derive power allocation matrix, $\boldsymbol{\beta}^*(t)$
  repeat
    $t_1 = t_1 + 1$
    Step 1.1: Update $\Gamma_{k_s,n}(t_1)$, $\theta_{k_s}(t_1)$, $\nu_{i,n}(t_1)$, $\psi_{j,n}(t_1)$, $\rho_{i,n}(t_1)$, $\mu_{j,n}(t_1)$, $\zeta_{i,n}(t_1)$, $\Delta_{k_s,n}(t_1)$, and $\Lambda_{j,n}(t_1)$ from (5.36), (5.37), (5.42-5.46), (5.50), (5.52)
    Step 1.2: Find $\boldsymbol{\beta}^*(t_1)$ according to (5.53) using CVX [46]
  until $\|\boldsymbol{\beta}^*(t_1) - \boldsymbol{\beta}^*(t_1 - 1)\| \leq \varepsilon_1$
  Step 2: Derive sub-carrier allocation matrix, $\boldsymbol{\alpha}^*(t)$
  repeat
    $t_2 = t_2 + 1$
    Step 2.1: Update $\tau_{k_s,n}(t_2)$, $\phi_{k_s,n}(t_2)$, $\upsilon_{k_s,n}(t_2)$, and $\omega_{k_s,n}(t_2)$ from (5.60-5.62), (5.64)
    Step 2.2: Find $\boldsymbol{\alpha}^*(t_2)$ according to (5.65) using CVX [46]
  until $\|\boldsymbol{\alpha}^*(t_2) - \boldsymbol{\alpha}^*(t_2 - 1)\| \leq \varepsilon_2$
until $\|\boldsymbol{\beta}^*(t) - \boldsymbol{\beta}^*(t-1)\| \leq \varepsilon_1$ && $\|\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)\| \leq \varepsilon_2$, otherwise $t = t + 1$

5.4 Proposed Algorithm

Both $\widetilde{\text{C1}}$ and C4 are non-convex and the optimization in (5.29) is computationally intractable due to the binary sub-carrier allocation indicator, $\boldsymbol{\alpha}$. In order to develop an efficient resource allocation algorithm, we first relax the elements of $\boldsymbol{\alpha}$ to be continuous on the interval $[0, 1]$, and then decompose (5.29) into separate power and sub-carrier allocation problems which, while simpler than the original problem, remain challenging due to the nature of the constraints. For each sub-problem, we use SCA and apply AGMA to approximate non-convex constraints with monomial functions on each iteration until the solution converges, as discussed in Appendix A.3. Overall, the resource allocation algorithm is shown in Algorithm 5.3. Once the algorithm converges, the sub-carrier allocation $\boldsymbol{\alpha}$ can be recovered to binary via integer rounding, as described in Sec. 3.3.

5.4.1 Power Allocation

With a fixed sub-carrier allocation, the optimization problem becomes

$$\min_{\boldsymbol{\beta}} \max_{\substack{\forall s\in\mathcal{S} \\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n} \tag{5.30}$$

Subject to: $\widetilde{\text{C1}}$, C2, C4.

For fixed sub-carrier allocation, both C2 and C4 are in the proper GP form but $\widetilde{\text{C1}}$ remains non-convex and appropriate approximations are required. To deal with $\widetilde{\text{C1}}$ we define two new variables, $\mathbf{X} = [X_{k_s,n}]_{K\times N}$ and $\mathbf{Y} = [Y_{k_s}]_{K\times 1}$, and transform $\widetilde{\text{C1}}$ into three constraints as follows

$$\text{C1.1} : \sum_{n\in\mathcal{N}} \alpha_{k_s,n} X_{k_s,n} + \sqrt{\frac{\epsilon_s}{1-\epsilon_s}} Y_{k_s} \geq R_s^{\mathbf{rsv}}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.31}$$

$$\text{C1.2} : X_{k_s,n} \leq \log\left(1 + \frac{a_{i,n}}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}\right) \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \forall n \in \mathcal{N}, \tag{5.32}$$

$$\text{C1.3} : Y_{k_s}^2 \leq \sum_{n\in\mathcal{N}} \alpha_{k_s,n} \left(\frac{a_{i,n}}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}\right)^2 \times \frac{\sum_{j=1}^{i-1} 4a_{j,n}^2\sigma_e^4}{(a_{i,n} + b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2)^2} , \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s. \tag{5.33}$$

For simplicity of notation, let $C = \sqrt{\frac{\epsilon_s}{(1-\epsilon_s)}}$, and we can re-arrange C1.1 as

$$\frac{R_s^{\mathbf{rsv}}}{\sum_{n\in\mathcal{N}} \alpha_{k_s,n} X_{k_s,n} + C Y_{k_s}} \leq 1. \tag{5.34}$$

Then, for iteration index t_1, we can apply AGMA to approximate C1.1 with the following convex constraint

$$\widetilde{\text{C1.1}} : R_s^{\mathbf{rsv}} \times \prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n} X_{k_s,n}(t_1)}{\Gamma_{k_s,n}(t_1)}\right)^{-\Gamma_{k_s,n}(t_1)} \times \left(\frac{C Y_{k_s}(t_1)}{\theta_{k_s}(t_1)}\right)^{-\theta_{k_s}(t_1)} \leq 1, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.35}$$

where

$$\Gamma_{k_s,n}(t_1) = \frac{\alpha_{k_s,n} X_{k_s,n}(t_1 - 1)}{\sum_{n\in\mathcal{N}} \alpha_{k_s,n} X_{k_s,n}(t_1 - 1) + C Y_{k_s}(t_1 - 1)}, \tag{5.36}$$

and

$$\theta_{k_s}(t_1) = \frac{C Y_{k_s}(t_1 - 1)}{\sum_{n\in\mathcal{N}} \alpha_{k_s,n} X_{k_s,n}(t_1 - 1) + C Y_{k_s}(t_1 - 1)}. \tag{5.37}$$

To eliminate the logarithm, C1.2 can be expressed as

$$e^{X_{k_s,n}} \leq 1 + \frac{a_{i,n}}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}. \tag{5.38}$$

Then, approximating $e^{X_{k_s,n}}$ using the truncated Taylor series and re-arranging, we have

$$\begin{aligned} e^{X_{k_s,n}} &\approx \sum_{m=0}^{10} \frac{X_{k_s,n}^m}{m!} \leq \left(\frac{a_{i,n} + b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2} \right) \\ &\Longrightarrow \sum_{m=0}^{10} \frac{X_{k_s,n}^m}{m!} \times \left(\frac{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}{a_{i,n} + b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2} \right) \leq 1. \end{aligned} \tag{5.39}$$

This approximation remains non-convex but can be transformed by AGMA into a convex constraint on each iteration as follows

$$\widetilde{\text{C1.2}} : \sum_{m=0}^{10} \frac{X_{k_s,n}^m(t_1)}{m!} \times \left(b_{i,n}(t_1) + \sum_{j=1}^{i-1} 2a_{j,n}(t_1)\sigma_e^2 \right) \times z_{i,n}(t_1) \leq 1, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \forall n \in \mathcal{N}. \tag{5.40}$$

Where, substituting from the definition of b in (5.23) and applying AGMA and simplifying, we have defined $z_{i,n}(t_1)$ as the following convex function

$$\begin{aligned} z_{i,n}(t_1) = \left(\frac{\sigma_{i,n}^2}{\nu_{i,n}(t_1)} \right)^{-\nu_{i,n}(t_1)} &\times \prod_{j=1}^{i-1} \left(\frac{2a_{j,n}(t_1)\sigma_e^2}{\psi_{j,n}(t_1)} \right)^{-\psi_{j,n}(t_1)} \\ &\times \prod_{j=i+1}^{K} \left(\frac{a_{j,n}(t_1)}{\rho_{j,n}(t_1)} \right)^{-\rho_{j,n}(t_1)} \times \left(\frac{a_{i,n}(t_1)}{\mu_{i,n}(t_1)} \right)^{-\mu_{i,n}(t_1)}, \end{aligned} \tag{5.41}$$

with

$$\nu_{i,n}(t_1) = {}^{\sigma_{i,n}^2}/_{\zeta_{i,n}(t_1)}, \tag{5.42}$$

$$\psi_{j,n}(t_1) = {}^{2a_{j,n}(t_1-1)\sigma_e^2}/_{\zeta_{i,n}(t_1)}, \tag{5.43}$$

$$\rho_{j,n}(t_1) = {}^{a_{j,n}(t_1-1)}/_{\zeta_{i,n}(t_1)}, \tag{5.44}$$

$$\mu_{i,n}(t_1) = {}^{a_{i,n}(t_1-1)}/_{\zeta_{i,n}(t_1)}, \tag{5.45}$$

and

$$\zeta_{i,n}(t_1) = \sigma_{i,n}^2 + a_{i,n}(t_1 - 1) + \sum_{j=1}^{i-1} 2a_{j,n}(t_1 - 1)\sigma_e^2 + \sum_{j=i+1}^{K} a_{j,n}(t_1 - 1). \tag{5.46}$$

For C1.3, we introduce auxiliary variable $\mathbf{W} = [W_{i,n}]_{K\times N}$ and define two new non-convex constraints

$$\text{C1.3.1} : Y_{k_s}^2 \leq \sum_{n\in\mathcal{N}} \alpha_{k_s,n} a_{i,n}^2 W_{i,n}, \tag{5.47}$$

$$\text{C1.3.2} : W_{i,n} \leq \frac{\sum_{j=1}^{i-1} 4a_{j,n}^2\sigma_e^4}{\left(b_{i,n}+\sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2\right)^2\left(a_{i,n}+b_{i,n}+\sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2\right)^2}. \tag{5.48}$$

We can again apply AGMA and on each iteration approximate C1.3.1 by

$$\widehat{\text{C1.3.1}} : Y_{k_s}^2(t_1) \times \prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n} a_{i,n}^2(t_1) W_{i,n}(t_1)}{\Delta_{k_s,n}(t_1)}\right)^{-\Delta_{k_s,n}(t_1)} \leq 1, \forall s\in\mathcal{S}, \forall k_s\in\mathcal{K}_s, \tag{5.49}$$

with

$$\Delta_{k_s,n}(t_1) = \frac{\alpha_{k_s,n} a_{i,n}^2(t_1-1) W_{i,n}(t_1-1)}{\sum_{n\in\mathcal{N}} \alpha_{k_s,n} a_{i,n}^2(t_1-1) W_{i,n}(t_1-1)}, \tag{5.50}$$

and C1.3.2 by

$$\begin{aligned}\widetilde{\text{C1.3.2}} : \frac{W_{i,n}(t_1)}{4\sigma_e^4}\times &\left(b_{i,n}(t_1)+\sum_{j=1}^{i-1} 2a_{j,n}(t_1)\sigma_e^2\right)^2 \times \left(a_{i,n}(t_1)+b_{i,n}(t_1)+\sum_{j=1}^{i-1} 2a_{j,n}(t_1)\sigma_e^2\right)^2 \\ &\times \prod_{j=1}^{i-1}\left(\frac{a_{j,n}^2(t_1)}{\Lambda_{j,n}(t_1)}\right)^{-\Lambda_{j,n}(t_1)} \leq 1, \forall s\in\mathcal{S}, \forall k_s\in\mathcal{K}_s, \forall n\in\mathcal{N},\end{aligned} \tag{5.51}$$

with

$$\Lambda_{j,n}(t_1) = \frac{a_{j,n}^2(t_1-1)}{\sum_{l=1}^{i-1} a_{l,n}^2(t_1-1)}. \tag{5.52}$$

Then, at each iteration t_1, we solve

$$\begin{aligned}&\min_{\boldsymbol{\beta},\boldsymbol{W},\boldsymbol{X},\boldsymbol{Y}} \max_{\substack{\forall s\in\mathcal{S}\\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n}(t_1) \\ &\text{Subject to: } \widetilde{\text{C1.1}}, \widetilde{\text{C1.2}}, \widehat{\text{C1.3.1}}, \widetilde{\text{C1.3.2}}, \text{C2}, \text{C4},\end{aligned} \tag{5.53}$$

which is in GP form and can be solved efficiently with standard convex optimization tools such as CVX [46].

5.4.2 Sub-carrier Allocation

With a fixed power allocation, the optimization problem becomes

$$\min_{\boldsymbol{\alpha}} \max_{\substack{\forall s\in\mathcal{S}\\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\beta_{k_s,n} \tag{5.54}$$
$$\text{Subject to: } \widetilde{\text{C1}}, \text{C3}.$$

C3 is in the proper GP form, but $\widetilde{\text{C1}}$ is non-convex and needs to be transformed into GP form. For simplicity of notation, we define

$$L_{i,n} = \log\left(1 + \frac{a_{i,n}}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}\right), \tag{5.55}$$
$$M_{i,n} = \left(\frac{a_{i,n}}{b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2}\right)^2 \times \frac{\sum_{j=1}^{i-1} 4a_{j,n}^2\sigma_e^4}{(a_{i,n} + b_{i,n} + \sum_{j=1}^{i-1} 2a_{j,n}\sigma_e^2)^2}, \tag{5.56}$$

which are constants for fixed power allocation. With $C = \sqrt{\frac{\epsilon_s}{(1-\epsilon_s)}}$ as before, we again introduce auxiliary variable Y_{k_s} but can now express $\widetilde{\text{C1}}$ as two non-convex constraints

$$\text{C1.4} : \sum_{n\in\mathcal{N}} \alpha_{k_s,n}L_{i,n} + CY_{k_s} \geq R_s^{\mathbf{rsv}}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.57}$$
$$\text{C1.5} : Y_{k_s}^2 \leq \sum_{n\in\mathcal{N}} \alpha_{k_s,n}M_{i,n}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s. \tag{5.58}$$

Applying AGMA, C1.4 can be approximated by a convex constraint as

$$\widetilde{\text{C1.4}} : R_s^{\mathbf{rsv}} \times \prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_2)L_{i,n}}{\tau_{k_s,n}(t_2)}\right)^{-\tau_{k_s,n}(t_2)} \times \left(\frac{CY_{k_s}(t_2)}{\phi_{k_s}(t_2)}\right)^{-\phi_{k_s}(t_2)} \leq 1, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.59}$$

where

$$\tau_{k_s,n}(t_2) = {}^{\alpha_{k_s,n}(t_2-1)L_{i,n}}/_{\upsilon_{k_s}(t_2)}, \tag{5.60}$$
$$\phi_{k_s}(t_2) = {}^{CY_{k_s}(t_2-1)}/_{\upsilon_{k_s}(t_2)}, \tag{5.61}$$

and

$$v_{k_s}(t_2) = \sum_{n\in\mathcal{N}} \alpha_{k_s,n}(t_2-1)L_{i,n} + CY_{k_s}(t_2-1). \tag{5.62}$$

Similarly, we can approximate C1.5 by

$$\widetilde{\text{C1.5}} : Y_{k_s}^2(t_2) \times \prod_{n\in\mathcal{N}} \left(\frac{\alpha_{k_s,n}(t_2)M_{i,n}}{\omega_{k_s,n}(t_2)}\right)^{-\omega_{k_s,n}(t_2)} \leq 1, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_s, \tag{5.63}$$

where

$$\omega_{k_s,n}(t_2) = \frac{\alpha_{k_s,n}(t_2-1)M_{i,n}}{\sum_{n\in\mathcal{N}} \alpha_{k_s,n}(t_2-1)M_{i,n}}. \tag{5.64}$$

Then at each iteration t_2 solve,

$$\begin{aligned} &\min_{\boldsymbol{\alpha},\boldsymbol{Y}} \max_{\substack{\forall s\in\mathcal{S}\\ \forall k_s\in\mathcal{K}_s}} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}(t_2)\beta_{k_s,n} \\ &\text{Subject to: } \widetilde{\text{C1.4}}, \widetilde{\text{C1.5}}, \text{C3}, \end{aligned} \tag{5.65}$$

which is in GP form and can be solved efficiently with standard convex optimization tools such as CVX [46].

5.4.3 Complexity Analysis

5.4.3.1 *Convergence*

As discussed in the preceding chapters, an interior point method is used to solved the problems in (5.53) and (5.65) and the required number of iterations is given by $\frac{\log(c/t^0\delta)}{\log\xi}$. In the problems of this chapter, for (5.53) the total number of constraints is $c_1 = 3K + 3KN$ and for (5.65) we have $c_2 = 2K + N$. The problem presents as polynomial and the number of iterations again grows as $\mathcal{O}(\sqrt{c})$ [47, 48]. The required iterations for each sub-problem to converge is

$$\begin{cases} \frac{\log(c_1/(t_1^0\delta_1))}{\log(\xi_1)}, & \text{Power (5.53)}, \\ \frac{\log(c_2/(t_2^0\delta_2))}{\log(\xi_2)}, & \text{Sub-carrier (5.65)}. \end{cases} \tag{5.66}$$

We see that power allocation (5.53) has higher complexity than the sub-carrier allocation (5.65) and is more sensitive to K and N. The number of iterations required to achieve convergence is studied further in Section 5.5.

5.4.3.2 *Computational Complexity*

As previously discussed, for each iteration a Newton step of an interior-point method costs $\mathcal{O}(c\varrho^2)$ operations where c is the number of constraints and ϱ is the number of variables [48]. For (5.53), the total number of variables is $\varrho_1 = 3KN + K = \mathcal{O}(KN)$ and for (5.65) we have $\varrho_2 = KN + K = \mathcal{O}(KN)$. Thus, the total computational complexity of solving the GP problem is polynomial, namely,

$$C_{\text{GP}} = \begin{cases} \mathcal{O}(c_1^{1.5}\varrho_1^2) = \mathcal{O}(K^{3.5}N^{3.5}), & \text{Power (5.53)}, \\ \mathcal{O}(c_2^{1.5}\varrho_2^2) = \mathcal{O}(K^{2}N^{3.5}), & \text{Sub-carrier (5.65)}. \end{cases} \tag{5.67}$$

In Algorithm 5.3, the worst-case number of computations required to convert to the GP form using AGMA is $i_1 = 2K + 5KN + 3N$ and $i_2 = 8 + 3K + 8KN + N$, for steps 1 and 2, respectively. Each is $\mathcal{O}(KN)$ and of lower order than C_{GP} yielding an overall complexity per iteration of each of the steps in Algorithm 5.3 of $C_{\text{I}} = C_{\text{GP}}$.

5.5 Numerical Results and Discussions[2]

The performance of the proposed algorithm was evaluated considering a single cell VWN with $N = 16$ sub-carriers which can each be shared by at most $K_n^{\mathbf{max}} = 4$ users, supporting two slices each with $K_s = 4$ users, except where otherwise noted. In all trials we have $P_{\mathbf{max}} = 23$ dBm, $\epsilon_s = \epsilon$, and $R_s^{\mathbf{rsv}} = R^{\mathbf{rsv}}$. The users are placed randomly within the BS coverage area following a uniform distribution and the channels gains are derived according to the Rayleigh fading model with $h_{k_s,n} = \chi_{k_s,n} d_{k_s}^{-\lambda}$ where $\lambda = 3$ is the path loss exponent, d_{k_s} is the distance between user k_s and the BS normalized to the radius of the coverage area, and $\chi_{k_s,n} \sim \text{Exp}(1)$. We evaluate SIC imperfections for several levels of SIC error variance with $\sigma_e^2 \in \{0.01, 0.025, 0.05, 0.10\}$. For comparison, we present results for non-robust NOMA with both perfect and imperfect SIC, and for OMA. For non-robust NOMA we consider $||e_{j,n}||^2 = \sigma_e^2$ in (5.8), i.e., is deterministic with σ_e^2 representing the imperfect level of achieved cancellation in SIC, and $\sigma_e^2 = 0$ represents perfect SIC. For OMA, we set $K_n^{\mathbf{max}} = 1$ to enforce orthogonality between the sub-carriers. In all formulations, where no feasible solutions exists for a given channel realization, user power is set to $P_{\mathbf{max}}$ and all users are considered to be in outage.

[2]For consistency between the derivation in the preceding sections based on the natural logarithm and the numerical results presented in this section, we have maintained the units of rates in nats/s/Hz.

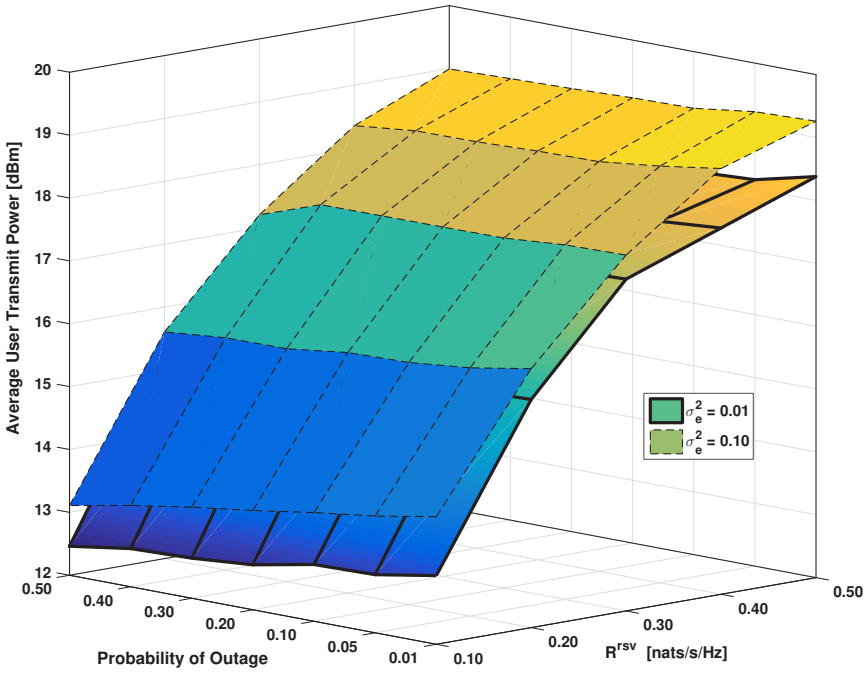


Fig. 5.1 Average UE transmit power versus reserved rate, R^{rsv}, and Pr(outage), ϵ

To visualize the overall relationship between probability of outage, reserved rate, and required transmit power, Fig. 5.1 shows the average transmit power per user versus R^{rsv} and ϵ for $\sigma_e^2 = 0.01$ and 0.10. Only the highest and lowest values of SIC error variance are shown in the figure for clarity. As expected, power increases with increased rate reservations and levels of SIC error and decreases with increasing probability of outage. For any specific value of ϵ, power increases sharply for increasing user rate reservations due to the decreased feasibility, and is always higher for increased SIC error. For any specific user rate reservation, lower power is required for less stringent user outage constraints as this increases the flexibility in finding a feasible solution which will provide the required maximum outage protection. The relationships over two of the three axes are depicted in the subsequent figures for specific cases.

Average transmit power versus R^{rsv} is plotted in Fig. 5.2 for some cases of user outage constraint, ϵ. As expected, average required transmit power increases proportionally to rate reservations, R^{rsv}, and SIC error variance, σ_e^2, because users must transmit with higher power to achieve their desired rates. Power levels for all values of σ_e^2 are shown for $\epsilon = 0.01$ and 0.50. There is a power trade-off for robustness and the results for non-robust power optimization are presented for comparison. For example, with $R^{\mathrm{rsv}} = 0.2$ nats/s/Hz and $\epsilon = 0.5$, i.e. power minimization

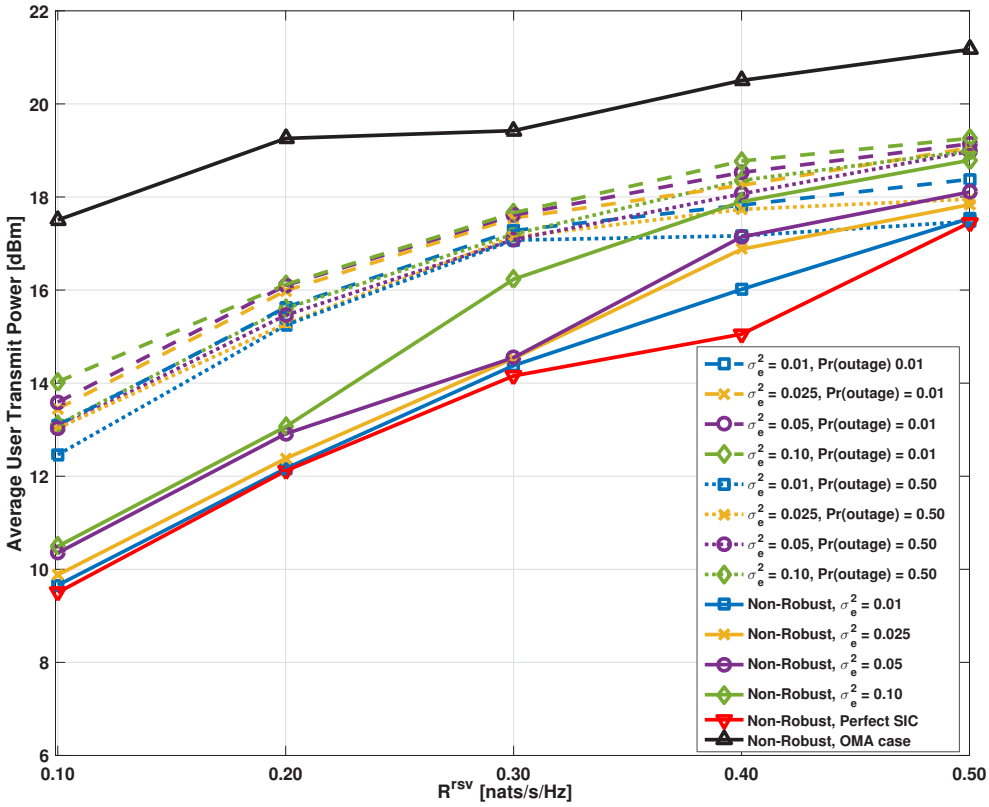


Fig. 5.2 Average UE transmit power versus reserved rate, R^{rsv}

allowing up to 50% probability of user outage, 3.08 dB, 2.89 dB, 2.60 dB, and 2.55 dB higher power is allocated by the robust optimization versus non-robust, for $\sigma_e^2 = 0.01$, 0.025, 0.01, and 0.10, respectively, and 3.28 dB higher than the perfect SIC case. However, for both robust and non-robust cases, NOMA outperforms OMA based on the inability to multiplex users on strong channels and the requirement to use weaker channels to maintain orthogonality between users and meet rate reservations. For the robust case and $R^{\mathrm{rsv}} = 0.2$ nats/s/Hz and $\epsilon = 0.01$, OMA requires 3.63 dB, 3.28 dB, 3.16 dB, and 3.14 dB, higher power compared to OMA, with $\sigma_e^2 = 0.01$, 0.025, 0.05, and 0.10, respectively.

For the robust optimization, average required transmit power increases with increasing SIC error variance and rate reservations but decreases with less strict outage constraints. This can be seen clearly in Fig. 5.3, which plots average transmit power versus probability of outage, ϵ, for $R^{\mathrm{rsv}} = 0.1$ nats/s/Hz and 0.5 nats/s/Hz. From the figure, we note that as the outage constraint is loosened, average power decreases but for higher levels of SIC error variance and R^{rsv} for

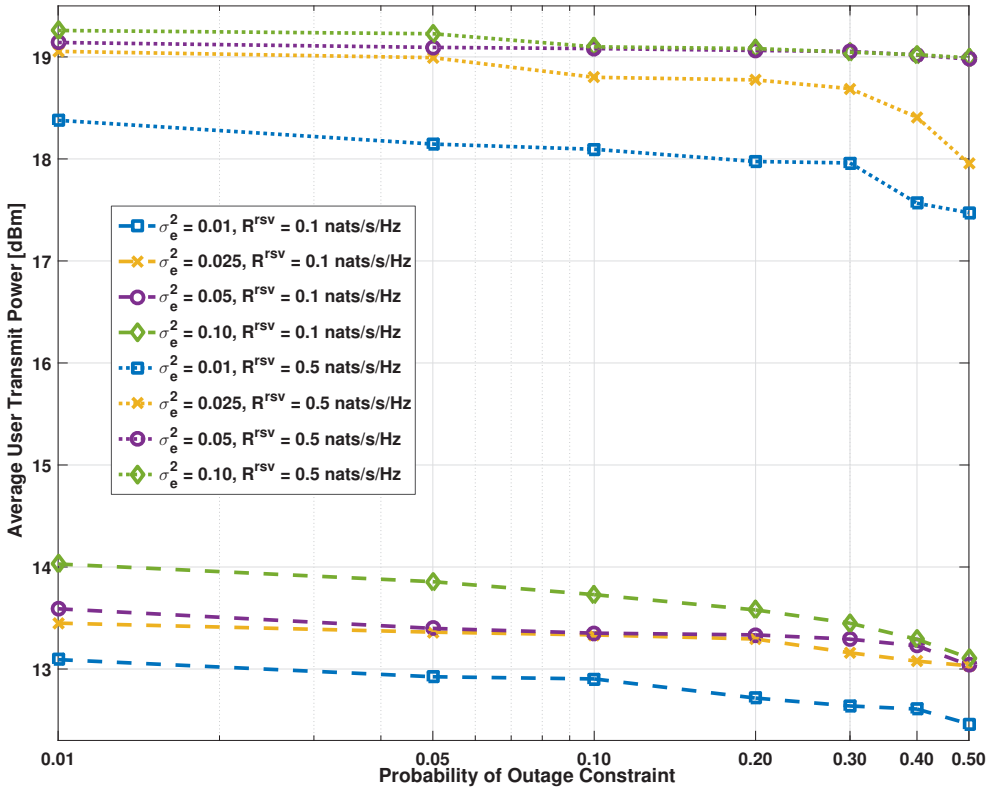


Fig. 5.3 Average UE transmit power versus Pr(outage), ϵ

$\epsilon \geq 0.10$ the plots converge as more users can be pushed into outage in favour reducing required user transmit power. For example, for $R^{\text{rsv}} = 0.1$ nats/s/Hz and $\epsilon = 0.10$ versus 0.01, required transmit power decreases by 0.19 dB, 0.12 dB, 0.24 dB, and 0.30 dB, for $\sigma_e^2 = 0.01, 0.025, 0.05,$ and 0.10, respectively. For $R^{\text{rsv}} = 0.5$ nats/s/Hz and $\epsilon = 0.50$ versus 0.01, required transmit power decreases by 0.91 dB, 0.42 dB, 0.55 dB, and 0.92 dB, for $\sigma_e^2 = 0.01, 0.025, 0.05,$ and 0.10, respectively.

Experience user outage versus probability of outage constraint, ϵ, is plotted in Fig. 5.4. For clarity, only the two extreme values for SIC error variance are shown. For $\sigma_e^2 = 0.01$ and $R^{\text{rsv}} = 0.1$ nats/s/Hz, we see that the experienced outage is very close to the probability used to constrain the problem, with correlation coefficient R = 0.9997 and root mean square error (RMSE), as compared to the ideal values of experienced outage equal to ϵ, of 0.00322. As R^{rsv} is increased, the resulting outage increasingly differs from the constraint due to decreased feasibility of solutions but for $\sigma_e^2 = 0.01$ the worst-case correlation is for $R^{\text{rsv}} = 0.5$ nats/s/Hz at R = 0.9771 and RMSE of 0.0306. For higher levels of SIC error variance, feasibility of solutions

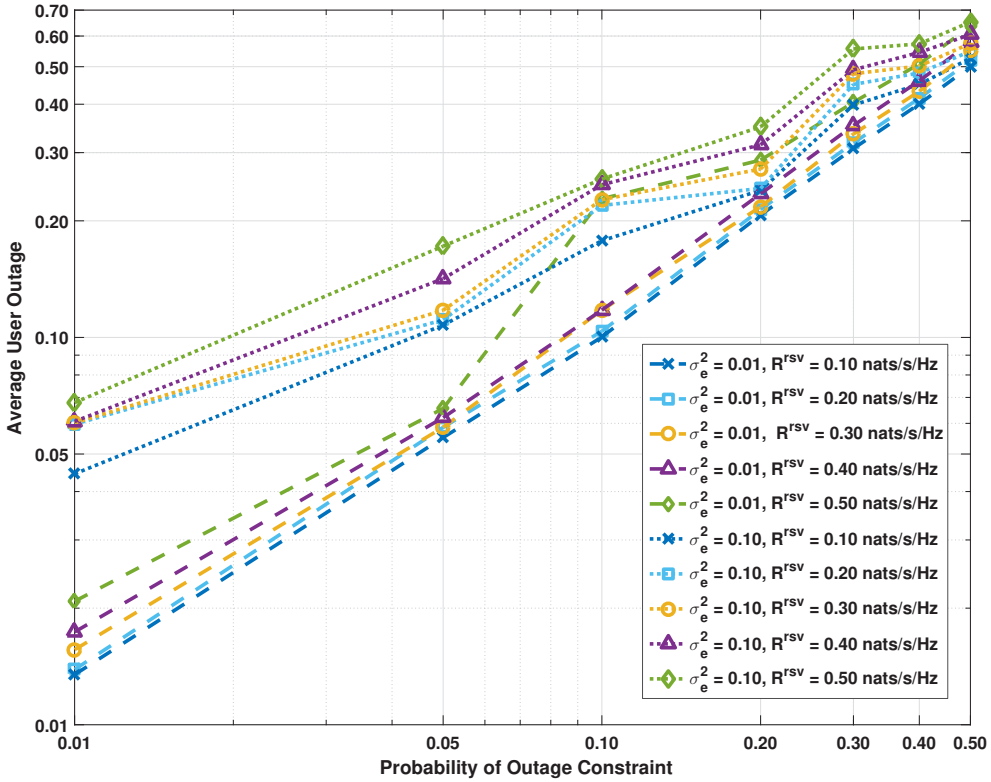


Fig. 5.4 Average user outage versus Pr(outage), ϵ

can be significantly impacted over the trials conducted and we see correlation coefficients of R = 0.9606 and 0.7276 and RMSE of 0.0402 and 0.106, for $\sigma_e^2 = 0.10$ and $R^{\text{rsv}} = 0.1$ and 0.5 nats/s/Hz, respectively.

Experienced user outage versus rate reservation, R^{rsv}, is plotted in Fig. 5.5. The experienced user outage aligns very well to the probability of outage constraint, ϵ, with reduced alignment as either R^{rsv} or σ_e^2 is increased, as was also seen in Fig. 5.4. Further, results for OMA and non-robust NOMA are presented where outage is not constrained, but when no feasible solution exists for a given channel realization all users are considered to be in outage. For the OMA case, at $R^{\text{rsv}} \geq 0.20$ nats/s/Hz experienced outage is higher than the robust cases with $\epsilon = 0.01, 0.05$ and higher than all non-robust NOMA cases. Non-robust NOMA with SIC error experiences higher levels outage as the value of σ_e^2 increases but does not exceed 5.25% even for the worst case level of error.

The ratio of users to available sub-carriers will reduce the flexibility of the system in both power and sub-carriers allocation. Under OMA, the most users which can be supported is equal

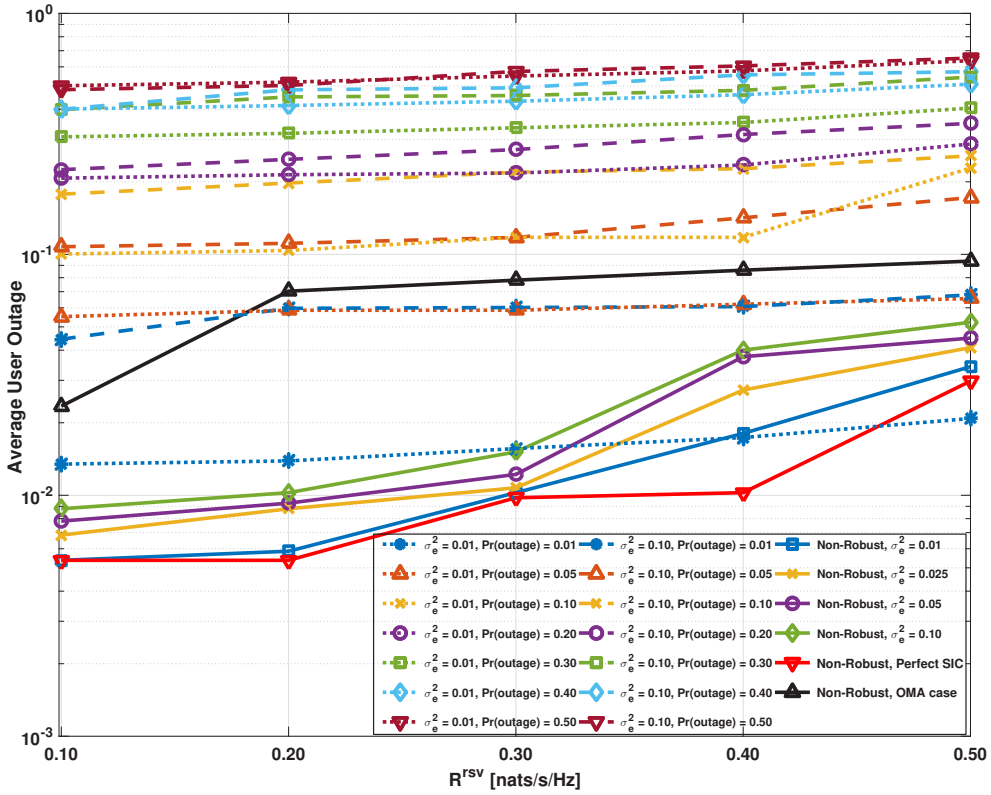
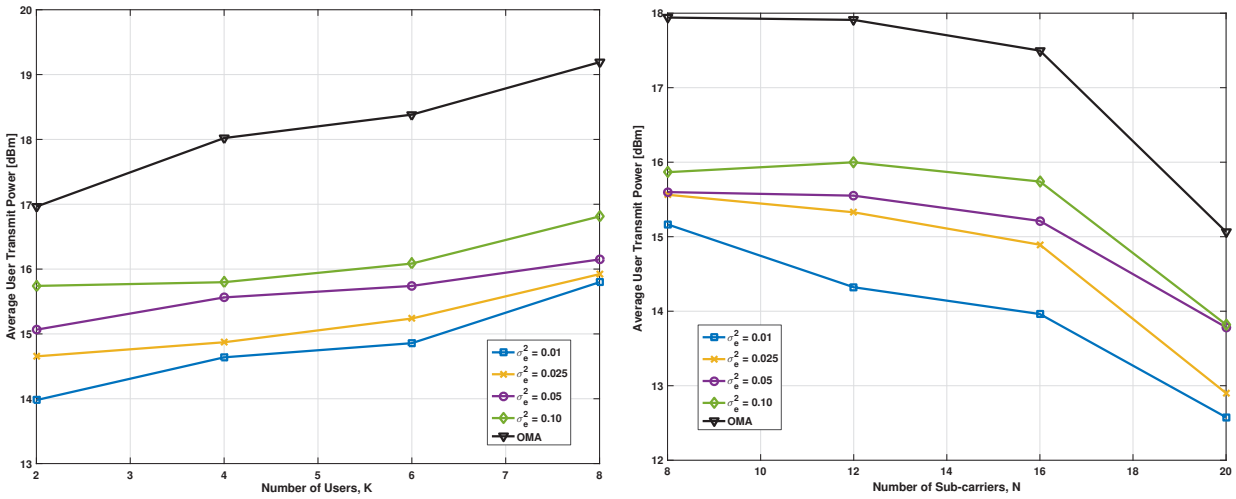


Fig. 5.5 Average user outage versus reserved rate, R^{rsv}

to the number of available sub-carriers, and only then if a suitably strong sub-carrier is available for each user to meet their rate reservation. Under NOMA, stronger sub-carriers can be leveraged by several users in order to minimize required transmit power to meet reservations, but with increased sub-carrier sharing average transmit power will necessarily increase, especially for higher levels of SIC error variance.

For a fixed number of sub-carriers, $N = 16$, Fig. 5.6 (a) plots average user transmit power as the total number of users in the system, i.e., system density, is increased, for $R^{\mathrm{rsv}} = 0.1$ nats/s/Hz and $\epsilon = 0.01$. With more users utilizing the same number of sub-carriers, required transmit power increases at all levels of SIC error variance as sub-carrier sharing results in increased inter-user interference. For low number of users, i.e. low system density, the user rate reservations can easily be met. As the density is increased, the required power also increases but at faster rate under OMA than NOMA. For example, as system density increases for $K = 2$ to 8, transmit power under OMA increases by 2.23 dB due to reduced flexibility in avoiding weak channels. The increase in required power under NOMA was only 1.82 dB, 1.27 dB, 1.08 dB, and

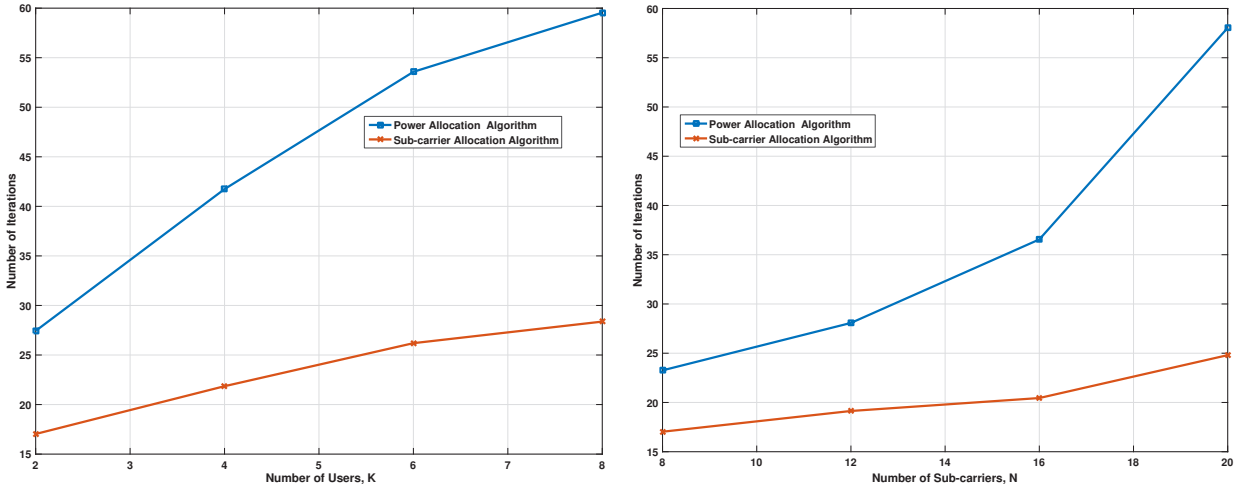


(a) (b)

Fig. 5.6 Average UE transmit power versus (a) users, K, and (b) sub-carriers, N

1.07 dB, for $\sigma_e^2 = 0.01$, 0.025, 0.05, and 0.10, respectively. In all cases, NOMA outperforms the OMA case. For example, with $K = 6$, required transmit power under NOMA is decreased compared to OMA by 3.52 dB, 3.14 dB, 2.64 dB, and 2.29 dB, for $\sigma_e^2 = 0.01$, 0.025, 0.05, and 0.10, respectively. For a fixed number of users, $K = 4$, Fig. 5.6 (b) plots average user transmit power as the total number of sub-carriers in the system is increased, for $R^{\text{rsv}} = 0.1$ nats/s/Hz and $\epsilon = 0.01$. With an increasing number of sub-carriers, i.e. decreasing system density, required transmit power decreases for both OMA and NOMA, with reduced requirement to utilize weaker channels or multiplex and experience inter-user interference, but NOMA still outperforms OMA in all cases. For example, with $N = 12$, required transmit power under NOMA is decreased compared to OMA by 3.59 dB, 2.58 dB, 2.36 dB, and 1.91 dB, for $\sigma_e^2 = 0.01$, 0.025, 0.05, and 0.10, respectively.

The convergence and complexity of the proposed algorithms was studied in Section 5.4.3 and the number of iterations for convergence for power allocation (5.53) and sub-carrier allocation (5.65) problems as a function of the number of users, K, and sub-carriers, N, are plotted in Fig. 5.7 (a) and 5.7 (b), respectively, for $R^{\text{rsv}} = 0.1$ nats/s and $\epsilon = 0.01$. As K and N increase the required number of iterations for each problem to converge increases. As expected, the power allocation problem is also more sensitive to changes in both K and N. The analysis found that the solution of the power allocation problem to be of higher complexity than the sub-carrier



(a) (b)

Fig. 5.7 Number of required iterations each sub-problem versus (a) users, K, and (b) sub-carriers, N

allocation, in (5.66), which is confirmed in the required number of iterations depicted in Fig. 5.5 and5.5.

5.6 Chapter Summary

In this chapter, we have investigated robust resource allocation in UL NOMA systems subject to residual cancellation errors from imperfect SIC. We evaluated a VWN with slices comprised of low power devices serving critical applications, for which minimum achieved rate and maximum user outage must be maintained. With the goal of maximizing battery life for such devices, we first derived the probability of outage as a function of SIC error variance and then used this result to formulate a robust resource allocation problem minimizing transmit power subject to slice and system constraints. The proposed iterative algorithm to solve the resulting non-convex and computationally intractable optimization is based on SCA and uses CGP and AGMA to transform to the convex GP form which can be solved efficiently at each iteration. Simulation results show the expected trade-off for robustness in terms of higher average transmit power compared to a non-robust approach. Despite this trade-off, the proposed algorithm outperforms the corresponding OMA system in terms of average user transmit power and overall system density due to the multiplexing gains available in NOMA systems.

Chapter 6

Conclusion

6.1 Summary

In this thesis, we have examined the use of UL NOMA within the VWN architecture paradigm and proposed resource allocation algorithms to minimize required UE transmit power and support QoS requirements of slices in several system models. We considered various QoS constraints, including minimum slice and user rates, minimum dedicated sub-carriers, and maximum user outage probability, for both SISO and MIMO channels to support the various expected use-cases proposed for future networks.

In Chapter 3, we examined the performance gains which are available with NOMA compared to traditional OMA in VWNs supporting slices with rate- and resource-based QoS. In this context, we proposed an efficient dynamic resource allocation algorithm in SISO channels and subsequently presented an extension of the algorithm to multi-antenna systems. The original non-convex optimization problem is converted to the GP form using CGP and AGMA, and SCA is used to iteratively solve for power and sub-carrier allocations. simulations results, we observed that, compared to OMA, significant power savings can be achieved via the power-domain multiplexing available under NOMA in both SISO and MIMO channels. Further, we found that such multiplexing can allow NOMA to support more users on fewer sub-carriers.

Chapter 4 extended the system model and iterative algorithm of the preceding chapter to formalize the extension to MIMO systems and examine the potential performance degradation from imperfect cancellation of user signals in NOMA. A sensitivity analysis of NOMA systems to residual cancellation errors from imperfect SIC was presented and via simulations results compared to both perfect SIC and OMA systems. It was discovered the as levels of SIC error increase,

performance degrades in terms of required user transmit power. Notwithstanding this degradation, even for residual interference of up to 10%, the gains with NOMA over OMA persist.

Finally, Chapter 5 introduced a robust resource allocation policy for NOMA VNW supporting critical applications. We considered high-priority applications such as health and public safety monitoring systems which rely on dense deployments of low-power sensors and high-reliability of communications. By considering the uncertainty in user performance with imperfect SIC, we developed a robust power and sub-carrier allocation algorithm which minimizes user transmit power while maintaining QoS via minimum rate and maximum outage probability. The simulation results demonstrated a trade-off for robustness in terms of higher average transmit power compared to a non-robust approach, however, the proposed algorithm outperforms the corresponding OMA system in terms of average user transmit power and overall system density due to the multiplexing gains available in NOMA systems. Moreover, the proposed algorithm can allow slices to set QoS targets to manage this power-robustness trade-off when a higher probability of outage is acceptable.

6.2 Potential Extensions and Future Work

As noted in this thesis, NOMA is a new technique and there are many avenues of research, such as UL and MIMO systems, for which there are limited works in the literature. VWN is also still an evolving paradigm and there are a number of possible extension of the works in this thesis; some of which are discussed below.

- In Chapters 3 and 4, we considered rate- and resource-based QoS at the slice level, with no consideration to user QoS. One possible extension to this would be to consider multi-level resource allocation wherein resources are allocated by the BS to slices to meet QoS requirements and then allocated to individual users based on user SLAs within each slice.
- We considered power-efficiency via resource allocation algorithms where the objective was to minimize user transmit power. Further studies can consider network efficiency by taking into account overhead required for reporting CSI to the BS and communicating power and sub-carrier allocations to users and power consumption at the BS to process both resource allocation algorithms and SIC to resolve individual signals.
- In all chapters, we limited our investigations to single-cell VWN. It would be interesting to extend the system models to consider intra-cell interference in multi-cell or heterogeneous

networks, especially to evaluate the NOMA performance at cell-edge and the potential benefits of multi-cell coordination and interference management and coordinated multipoint communication.

- In Chapter 4 we consider a generalized multi-antenna system with the BS coverage area divided into short- and long-range regions. The analysis of the optimal boundary between these regions would make for an interesting extension to this work.

- We consider the resource allocation problem in VWN from the context of BS virtualization to support multiple service providers. A possible extension of these works could include cloud-based radio access networks, wherein baseband processing and resource allocation algorithms are defined and executed remotely in the cloud, rather than locally at the BS, and examine the performance degradation of the proposed algorithms with additional constraints on processing power and back-haul network capacity are imposed.

- As outlined in Chapter 2, VWN and software-defined networking can allow for so-called "super BSs" which can support multiple-RATs. Additionally, the high cost of licensed spectrum and the advent of LAA-LTE to extend cellular networks into the unlicensed bands, co-existence of multiple RATs is an active research topic. One possible extension would be to examine the use of NOMA in the context of coordinated coexistence, possibly by considering both power- and code-domain NOMA hybrid systems.

Appendix: Overview of Complementary Geometric Programming

A.1 Preliminaries

Monomial and Posynomial Functions

If we have a vector $\mathbf{x} = (x_1, \ldots, x_n)$ with components $x_i \in \mathbb{R}^+$, for $c \in \mathbb{R}^+$ and $a_i \in \mathbb{R}$, a monomial is a function of the form

$$f(x) = c x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}. \tag{A.1.1}$$

A posynomial is a sum of monomial terms, i.e. a function of the form

$$g(x) = \sum_{k=1}^{K} c x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}. \tag{A.1.2}$$

Arithmetic-Geometric Mean Approximation

In approaching many optimization problems, it becomes necessary to approximate posynomial terms in either the objective function or constraints to convert the non-convex problem to an approximate form. A widely used approach is to convert posynomial terms into the approximated product of monomials is based on the arithmetic-geometric mean inequality, namely, for $a_j > 0$ and $d_j \geq 0$ s.t. $\sum_j d_j = 1$,

$$\sum_j a_j \geq \prod_j \left(\frac{a_j}{d_j} \right)^{d_j}. \tag{A.1.3}$$

Using this inequality, the arithmetic-geometric mean approximation (AGMA) can be used to approximate posynomial functions as products of monomial terms. For example, given posynomial function $f(\boldsymbol{x}) = \sum_k g_k(\boldsymbol{x})$ where $g_k(\boldsymbol{x})$ are monomials, applying the arithmetic-geometric mean inequality we have

$$f(\boldsymbol{x}) \geq \hat{f}(\boldsymbol{x}) = \prod_k \left(\frac{g_k(\boldsymbol{x})}{\alpha_k(\boldsymbol{x}_0)} \right)^{\alpha_k(\boldsymbol{x}_0)}, \tag{A.1.4}$$

where the weights $\alpha_k(\boldsymbol{x}_0)$ are computed as

$$\alpha_k(\boldsymbol{x}_0) = \frac{g_k(\boldsymbol{x}_0)}{f(\boldsymbol{x}_0)}, \forall k. \tag{A.1.5}$$

In the above approximation, $\boldsymbol{x}_0 > 0$ is the optimal solution obtained from the preceding iteration of the optimization. It has been proved that AGMA provides the best local monomial approximation of $f(\boldsymbol{x})$ near $\boldsymbol{x}_0$ [47].

A.2 Geometric Programming

Geometric programming (GP) is a class of non-linear optimization which can be efficiently solved and has many applications in science and engineering [55]. The standard form of a GP problem is

$$\begin{aligned} &\min_{\mathbf{x}} \ f_0(\mathbf{x}) \\ &\text{Subject to: } \ f_i(\mathbf{x}) \leq 1, i = 1, 2, \cdots, m \\ &\qquad\qquad g_j(\mathbf{x}) = 1, j = 1, 2, \cdots, M. \end{aligned} \tag{A.2.1}$$

for non-negative optimization variables $\mathbf{x} = [x_1, x_2, \cdots, x_N]$ and, for $c_{j,n} > 0$ and $a_{j,n} \in \mathbb{R}$, $g_j = \prod_{n=1}^{N} c_{j,n} x_n^{a_{j,n}}, \ \forall \ j$, i.e. are monomial functions, and f_i are products of monomial terms $\forall \ i$, i.e. are posynomials.

The restrictions on the objective function and constraints in GP problems cannot be met for many practical problems, e.g., some f_i are monomial, some g_j are posynomial, or are ratios of posynomials, for some i, j, or if f_i are lower bounds rather than upper bounds.

A.3 Complementary Geometric Programming

Complementary geometric programming (CGP) is an approach to solve problems formulated in terms of rational functions of posynomial terms [55,56]. Such problems can be solved iteratively using successive convex approximation (SCA) by substituting monomial approximations to convert the problem into the GP form via AGMA on each iteration. For example, given an upper bound constraint, $f_1(\mathbf{x}) : x_1 + x_2 \geq 1$, we can reformulate this as a lower bound

$$f_1(\mathbf{x}) : \frac{1}{x_1 + x_2} \leq 1, \tag{A.3.1}$$

which is a ratio of posynomial terms. Applying AGMA at iteration index t we approximate this constraint as a product of two monomial terms

$$\tilde{f}_1(\mathbf{x}(t)) : \left(\frac{x_1(t)}{\xi_1(t)}\right)^{\xi_1(t)} \left(\frac{x_2(t)}{\xi_2(t)}\right)^{\xi_2(t)} \leq 1, \tag{A.3.2}$$

with the weights $\xi_i(t)$ defined as

$$\xi_i(t) = \frac{x_i(t-1)}{(x_1(t-1) + x_2(t-1))}, \quad i = 1, 2. \tag{A.3.3}$$

Applying these approximations to all functions in the optimization problem, as needed, the resulting GP problem can be solved efficiently via numerical methods. The convergence of CGP has been proven in [56] and it has been shown that the output of algorithms based on SCA and AGMA converges to a local maximum of the original problem and have very close performance to the optimal solution [43, 57, 58].

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast up-date 2016-2021," whitepaper, Feb 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358–380, First quarter 2015.

[3] J. Feng, Q. Zhang, G. Dong, P. Cao, and Z. Feng, "An approach to 5G wireless network virtualization: Architecture and trial environment," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar 2017, pp. 1–6.

[4] Z. Feng, C. Qiu, Z. Feng, Z. Wei, W. Li, and P. Zhang, "An effective approach to 5G: Wireless network virtualization," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 53–59, Dec 2015.

[5] H. Tabassum, M. S. Ali, E. Hossain, M. J. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) in cellular uplink and downlink: Challenges and enabling techniques," *CoRR*, vol. abs/1608.05783, 2016. [Online]. Available: http://arxiv.org/abs/1608.05783

[6] 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 Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec 2014.

[7] L. Lei, D. Yuan, and P. Värbrand, "On power minimization for non-orthogonal multiple access (NOMA)," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2458–2461, Dec 2016.

[8] 3GPP, "Study on downlink multiuser superposition transmission (MUST) for LTE," *TR 36.859 Version 13.1.0 Release 13*, 2015.

[9] Q.-D. Ho, D. Tweed, and T. Le-Ngoc, *Long Term Evolution in Unlicensed Bands*. Springer International Publishings, 2016.

[10] D. Tweed, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Dynamic resource allocation for MC-NOMA virtualized wireless networks with imperfect SIC," in *Proc. IEEE Intl. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Oct. 2017.

[11] 3GPP, "Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-Advanced)," *TR 36.913 v13.0.0*, July 2016.

[12] 3GPP , "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description," *TR 36.300 v13.3.0*, 2016.

[13] ITU, "IMT vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," Recommendation M.2083, Sep 2015.

[14] "LTE," Online, 3GPP, White Paper, ND, accessed: June 21, 2016. [Online]. Available: www.3gpp.org/technologies/keywords-acronyms/98-lte

[15] "LTE-Advanced," Online, 3GPP, White Paper, 2013, accessed: June 21, 2016. [Online]. Available: www.3gpp.org/technologies/keywords-acronyms/97-lte-advanced

[16] M. Abdullah and A. Yonis, "Performance of LTE release 8 and release 10 in wireless communications," in *Proc. IEEE Intl. Conf. on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*, June 2012, pp. 236–241.

[17] H. J. Kwon, J. Jeon, A. Bhorkar, Q. Ye, H. Harada, Y. Jiang, L. Liu, S. Nagata, B. L. Ng, T. Novlan, J. Oh, and W. Yi, "Licensed-assisted access to unlicensed spectrum in LTE release 13," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 201–207, February 2017.

[18] "Requirements related to technical performance for IMT-Advanced radio interface (Rep. ITU-R M.2134)," ITU, Tech. Rep., 2008.

[19] 3GPP. (2016, July) 3GPP work programme (and associated reference links therein). Online. [Online]. Available: www.3gpp.org/dynareport/GanttChart-Level-2.htm

[20] E. Hossain and M. Hasan, "5G cellular: Key enabling technologies and research challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, June 2015.

[21] C. Liang and F. R. Yu, "Wireless virtualization for next generation mobile cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 61–69, February 2015.

[22] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 84–91, April 2016.

[23] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, June 2013, pp. 1–5.

[24] 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 Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.

[25] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, Mar 2016.

[26] R. Zhang, M. Wang, et al, "LTE-unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, June 2015.

[27] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for MC-NOMA systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2016.

[28] A. Mokdad, P. Azmi, and N. Mokari, "Radio resource allocation for heterogeneous traffic in GFDM-NOMA heterogeneous cellular networks," *IET Commun.*, vol. 10, no. 12, pp. 1444–1455, 2016.

[29] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov 2016.

[30] H. Haci, H. Zhu, and J. Wang, "Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique," *IEEE Trans. Wireless Commun.*, vol. 65, no. 3, pp. 1319–1335, Mar 2017.

[31] H. Tabassum, E. Hossain, and M. J. Hossain, "Modeling and analysis of uplink non-orthogonal multiple access (NOMA) in large-scale cellular networks using poisson cluster processes," *IEEE Trans. Commun.*, vol. PP, no. 99, 2017, pre-print.

[32] Z. Yang, W. Xu, H. Xu, J. Shi, and M. Chen, "Energy efficient non-orthogonal multiple access for machine-to-machine communications," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 817–820, April 2017.

[33] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan 2016.

[34] 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.

[35] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[36] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink nonorthogonal multiple access with statistical channel state information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4718–4731, Jul 2016.

[37] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with 1-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 4718–4731, Oct 2016.

[38] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1226–1230, Sept 2016.

[39] A. Mokdad, M. Moltafet, P. Azmi, and N. Mokari, "Robust radio resource allocation for heterogeneous traffic in pd-noma-based cellular systems," in *Proc. Iranian Conf. on Electrical Engineering (ICEE)*, May 2017, pp. 1796–1801.

[40] L. Gao, P. Li, Z. Pan, N. Liu, and X. You, "Virtualization framework and VCG based resource block allocation scheme for LTE virtualization," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, May 2016, pp. 1–6.

[41] C. Bao, Z. Zhao, X. Sui, and H. Zhang, "Energy-efficient user association and downlink power allocation in software defined hetnet," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, May 2016, pp. 1–5.

[42] T. D. Tran and L. B. Le, "Resource allocation for efficient bandwidth provisioning in virtualized wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar 2017, pp. 1–6.

[43] S. Parsaeefard, R. Dawadi, M. Derakhshani, and T. Le-Ngoc, "Joint user-association and resource-allocation in virtualized wireless networks," *IEEE Access*, vol. 4, pp. 2738–2750, 2016.

[44] R. Dawadi, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Power-efficient resource allocation in NOMA virtualized wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec 2016.

[45] P. H. Huang, Y. Gai, B. Krishnamachari, and A. Sridharan, "Subcarrier allocation in multiuser ofdm systems: Complexity and approximability," in *2010 IEEE Wireless Communication and Networking Conference*, April 2010, pp. 1–6.

[46] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1, build 1116," http://cvxr.com/cvx, Mar 2017.

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[48] A. Nemirovski, "Interior point polynomial time methods in convex programming," *Lecture Notes*, 2004.

[49] Z. Ding, R. Schober, and H. V. Poor, "A general mimo framework for noma downlink and uplink transmission based on signal alignment," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, June 2016.

[50] A. Goldsmith, *Wireless Comunications*. Cambridge University Press, 2004.

[51] S. Lim and K. Ko, "Non-orthogonal multiple access (NOMA) to enhance capacity in 5G," *Intl. Journal of Contents*, vol. 11, no. 4, pp. 38–43, Dec 2015.

[52] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2007.

[53] S. Parsaeefard, A. R. Sharafat, and N. Mokari, *Robust Resource Allocation in Future Wireless Networks*. Springer Nature, 2017.

[54] B. Dimitris and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, Feb 2004.

[55] M. Chiang, "Geometric programming for communications systems," *Commun. Inf. Theory*, vol. 2, no. 1, pp. 1–154, Jul 2005.

[56] M. Avriel and A. Williams, "Complementary geometric programming," *SIAM Journal on Applied Mathematics*, vol. 19, no. 1, pp. 125–141, Jul 1970.

[57] M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul 2007.

[58] G. Xu, "Global optimization of signomial geometric programming problems," *European Journal of Operational Research*, vol. 233, no. 3, pp. 500–510, 2014.