User-Association and Resource-Allocation in Multi-cell Virtualized Wireless Networks

Rajesh Dawadi

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

September 2016

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

© 2016 Rajesh Dawadi

2016/09/25

Dedicated to my parents.

Abstract

The unprecedented growth in the number of wireless devices and applications, along with the advent of machine-type communication devices, has constrained current wireless networks. Virtualized wireless network (VWN) has been recently envisaged to allow multiple service providers to efficiently share the physical wireless communications infrastructure from the network operator. In a VWN, each service provider expects to have its allocated *slice* of wireless communications resources of the various infrastructure components (e.g., base-stations) in order to support its own end-users with different Quality-of-Service (QoS) requirements. VWN, therefore, presents unique challenges specifically related to user association, dynamic resource allocation, slice isolation and resource utility that must be addressed to make it practically effective. The overall objectives of this work are to study and develop computationally efficient algorithms for user-association (UA) and resource-allocation (RA) in multi-cell VWNs.

To improve the network spectral efficiency, we study and formulate a joint UA and RA problem that assigns users to base-stations (BSs) and allocates power and sub-carriers to users for sum-rate maximization, while satisfying the minimum rate reserved for each service provider (slice). By employing successive convex approximation (SCA) and complementary geometric programming (CGP), we convert this non-convex problem into the convex form, and develop a computationally efficient solution. We also study and develop an efficient solution for UA and RA in a multi-cell VWN with the aim of maximizing the transmit power efficiency, while still preserving the slice isolation requirements. Simulation results demonstrate the performance improvement of the developed solutions as compared to the conventional approach of associating users to BS based on the received signal-to-interference-plus-noise power ratio (SINR). We further explore the resource allocation in a VWN with massive-MIMO BS to assign antennas and pilot duration to users in addition to sub-carriers and power for sum-rate maximization while maintaining slice isolation. Simulation results verify that the proposed scheme significantly improves the feasibility of the VWN leading to a higher spectral efficiency.

Dynamic resource allocation algorithms in a VWN require knowledge of the channel state information for all users across all slices and base stations which effectively requires co-ordination among the base stations. One potential solution to address the need for the co-ordination is the use of a cloud radio access network (C-RAN). A resource allocation scheme is developed for a C-RAN-based VWN that jointly assigns users to remote radio heads and baseband units, and allocates power and antennas with the aim of maximizing the total sum-rate. Numerical results

compare the performance of the VWN in terms of the spectral efficiency in both perfect and imperfect channel estimation scenarios with the proposed scheme. Finally, in order to maximize the power efficiency of the VWN, non-orthogonal multiple access (NOMA) is considered. Specifically, we study and develop a resource allocation scheme that allocates power coefficients to users in all slices with the aim of minimizing the total transmit power while satisfying the rate reservations of the slices. Through numerical results, it is shown that the proposed NOMA-based scheme provides a significant improvement in power efficiency, as compared to the OFDMA-based scheme.

Sommaire

La croissance sans précédent du nombre d'appareils et applications sans fil, avec l'ascension des machines communicantes, a façonné les réseaux sans fil actuels. De ce fait, les réseaux sans fil de nouvelle génération doivent augmenter leur capacité et maintenir une latence réduite et une grande efficacité énergétique pour supporter l'augmentation des appareils à fort débit de transmission et à puissance réduite. La virtualisation sans fil est l'une des architectures clé envisagée pour les réseaux de nouvelle génération afin de répondre à ces exigences. Un réseau sans fil virtualisé (VWN) présente des challenges uniques spécifiquement liés à l'association des utilisateurs, l'allocation dynamique des ressources et L'efficacité énergétique qui doivent être adressés avant que tel réseau ne devient pratique. L'objective de cette thèse est d'étudier et de proposer des schémas d'allocation de ressources qui répondent a ces défis dans un VWN.

Dans le but de maximiser l'efficacité spectrale du réseau, un schéma d'association d'utilisateurs et d'allocation de ressource est développé pour une VWN multicellulaires basée sur l'OFDMA qui assigne les utilisateurs aux BSs et alloue la puissance et les sous-porteuses afin de maximiser le débit atteint par l'ensemble des utilisateurs, tout en assurant la réservation du débit minimal pour chaque fournisseur de services. Un algorithme est présenté qui emploie des approximations convexes successives et la programmation géométrique complémentaire pour convertir le problème non-convexe à une forme convexe. De plus, afin d'améliorer la faisabilité du VWN, un schéma a allocation de ressource pour une VWN a MIMO massif est proposé qui alloue les sous-porteuses, la puissance, les antennes et la durée des pilotes aux utilisateurs dans le but de maximiser l'efficacité spectrale. Les résultats numériques vérifient qui les schémas proposés améliorent signifiâmes la faisabilité du VWN conduisant à une meilleur efficacité spectrale.

Les algorithmes d'allocation dynamique des ressources dans un VWN exigent la connaissance de l'état du canal pour tous les utilisateurs dans toutes les tranches et les stations de base. Une possible solution pour répondre à la nécessité de maintenir efficacement la coordination entre les BSs est l'utilisation d'un réseau d'accès de cloud. Un schéma d'allocation de ressource est développé qui affecte jointement les utilisateurs aux stations radio et aux unités de bande de base, et alloue la puissance et les antennes dans le but de maximiser le débit total. Les résultats numériques comparent les performances du VWN en termes d'efficacité spectrale dans les deux cas d'estimation parfaite et imparfaite du canal utilisant le schéma proposé. Finalement, afin de maximiser l'efficacité en termes de puissance du VWN, un schéma d'accès multiple non orthogonal est appliqué. Précisément, un schéma d'allocation des ressources est développé qui alloue

les coefficients de puissance aux utilisateurs and toutes les tranches dans le but de minimiser la puissance de transmission totale tout en satisfaisant le débit réservé aux tranches. A travers des résultats numériques, il est démontré que le schéma proposé offre une amélioration significative en termes d'efficacité énergétique, comparé au schéma OFDMA.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Tho Le-Ngoc for his continuous support during my MEng. studies and research and for his sincere guidance and encouragement throughout. His vast knowledge and experience along with his strong dedication to top-quality research were truly motivational. It has been an honor to work under his guidance.

My special thanks goes to Dr. Saeedeh Parsaeefard and Dr. Mahsa Derakhshani for their excellent research coaching and collaboration. I feel greatly indebted for the chance given to me to learn from their research experience and technical expertise. I would also like to acknowledge the financial support received from the Natural Sciences and Engineering Research Council of Canada (NSERC).

I am thankful to my friends at McGill University and at the Broadband Communications Research Laboratory, especially, Vikas, Atoosa, Alfred, Rouzhou, Xiaowei, Sanjeewa for the helpful and engaging discussions during the lab meetings and for making the journey smooth and memorable. I am grateful to Daniel for undertaking the painstaking task of proofreading my thesis and for providing valuable suggestions. I am thankful to Ahmed for the French translation of the thesis abstract.

I would like to thank my family for all their love and encouragement and for supporting me in all my endeavors. Thank you for instilling in me the spirit of hard work and dedication. This would not have been possible without your support.

Contents

List of Figures

List of Acronyms

AWGN	Additive White Gaussian Noise
BBU	Baseband unit
bps	Bits per second
BS	Base station
cdf	Cumulative Distribution Function
CGP	Complementary Geometric Programming
CSI	Channel State Information
C-RAN	Cloud Radio Access Network
dB	Decibels
UA	User Association
RA	Resource Allocation
EE	Energy Efficiency
EPA	Equal Power Allocation
GP	Geometric Programming
Hz	Hertz
i.i.d.	Independent and Identically Distributed
Imperf	Imperfect CSI estimation
KKT	Karush–Kuhn–Tucker
LTE	Long Term Evolution
MAC	Media Access Control
MIMO	Multiple Input and Multiple Output
MMSE	Minimum Mean Square Error
MRC	Maximum Ratio Combining
NOMA	Non-orthogonal Multiple Access

OFDMA	Orthogonal Frequency Division Multiple Access
pdf	Probability density function
Perf	Perfect CSI estimation
Pr (Outage)	Outage Probability
PRB	Physical Resource Block
QoS	Quality-of-Service
RAN	Radio Access Network
RRH	Remote Radio Head
RV	Random Variable
SCA	Successive Convex Approximation
SE	Spectal Efficiency
SINR	Signal-to-Interference-plus-Noise Ratio
UMTS	Universal Mobile Telecommunications Service
VPN	Virtual Private Network
VWN	Virtualized Wireless Network
w.r.t.	with respect to
WiMAX	Wireless Interoperability for Microwave Access
WLAN	Wireless Local Area Network

List of Symbols

$\mathcal{S}, S$	The set and total number of slices in the VWN
$\mathcal{L}, L$	The set and total number of BSs in the system
$\mathcal{K}, K$	The set and total number of users in the system
$\mathcal{K}_s$	Set of users in slice $s \in \mathcal{S}$
W	Total bandwidth available in VWN
$\mathcal{N}, N$	Set and total number of sub-carriers in the system
W_c	Coherence bandwidth
ζ	Path loss exponent
M	Total number of antennas available at the BS
$c_s^{\mathrm{M}}, c_s^{\mathrm{P}}, c_s^{\tau}$	Pricing factors for the number of allocated antennas, transmit power and pilot duration of slice s, respectively
$\alpha_{k_s,n}, \boldsymbol{\alpha}_{k_s}, \boldsymbol{\alpha}_s, \boldsymbol{\alpha}$	Sub-carrier assignment indicators: variable for sub-carrier n to user k_s in slice s , vector for user k_s, vector for slice s and vector for the whole VWN, respectively
$p_{k_s,n}, \mathbf{P}_{k_s}, \mathbf{P}_s, \mathbf{P}$	Power allocation indicators: variable representing power allocated to user k_s on sub-carrier n and power allocation vectors for user k_s, slice s and VWN, respectively
d_{l,k_s}	Distance of user k_s from BS l
δ_ν	Step size for Lagrange multiplier ν
ϵ	A constant with small value
$\mathbf{f}_{k_s,n}^{\mathrm{Imperf}}$	Precoding vector of user k_s on sub-carrier n for imperfect CSI
$\mathbf{f}_{k_s,n}^{\mathrm{Perf}}$	Precoding vector of user k_s on sub-carrier n for perfect CSI
$\gamma_{l,k_s,n}$	SINR of user k_s associated to BS l and on sub-carrier n

$\gamma_{l,k_s,n}^{\text{Perf}}$	SINR of user k_s on sub-carrier n for perfect CSI case without any pilot contamination error
$\gamma_{l,k_s,n}^{\text{Imperf}}$	SINR of user k_s on sub-carrier n for imperfect CSI case due to pilot contamination error
$h_{l,k_s,n}, \mathbf{h}_{k_s}, \mathbf{h}_s, \mathbf{h}$	Channel gain of user k_s associated with BS l on sub-carrier n, channel vectors of user k_s, slice s and VWN, respectively
$\widetilde{h}_{k_s,n}, \widetilde{\mathbf{h}}_{k_s,n}$	Uncertain channel gain of user k_s on sub-carrier n and uncertain channel vector of user k_s, respectively.
$\mathbf{h}_{k_s,n}^{\text{Imperf}}$	Estimated channel vector including pilot contamination error of user k_s on sub-carrier n
$I_{l,k_s,n}$	Interference to the user k_s in BS l from other users at sub-carrier n
R_s^{rsv}	Minimum reserved rate for slice s
M_s^{rsv}	Minimum reserved antennas for slice s
$\lambda, \nu_n, \nu_s, \psi_s, \phi_s, \zeta_{k_s}, \rho_n$	Lagrangian multipliers
$M_{k_s,n}$	Number of allocated antennas to user k_s in slice s and on sub-carrier n
$\mathbf{M}_s^{\min}, \mathbf{M}_s^{\max}$	Vectors with elements being the minimum and maximum numbers of antennas from the BS allocated to slice s, respectively
$P_l^{\max}$	Maximum downlink transmit power of BS l
$P_{k_s}^{\max}$	Maximum uplink transmit power limit of user k_s in slice s
R_{l,k_s}	Rate of user k_s in BS l on all sub-carriers in any time slot
$R_{l,k_s,n}$	Rate of user k_s in BS l on sub-carrier n
$R_{l,k_s,n}^{\text{Perf}}$	Rate of user k_s in BS l on sub-carrier n for perfect CSI case
$R_{l,k_s,n}^{\text{Imperf}}$	Rate of user k_s in BS l on sub-carrier n for imperfect CSI case
σ	Noise power of all users over all sub-carrier
$\tau_{k_s,n}$	Uplink pilot duration of user k in slice s in sub-carrier n
T_c	Coherence interval time
U_s	Utility of slice s
χ	Exponential random variable with mean of 1
$\mathbf{z}_{k_s,n}$	Vector of AWGN
$\rho_{k_s,n}$	Power allocation coefficient in NOMA for user k_s in sub-carrier n
$f_{k_s,b}$	Indicator for association between user k_s and BBU b
$\omega_{k_s,b}$	Load balancing factor BBU b to the user k_s
$o_b^{\max}$	Maximum number of users supported by BBU b

$c_{l,b}^{\max}$	Capacity of the fronthaul link between RRH l and BBU b

Mathematical Operators

$\{a, b\}$	Set containing two discrete elements a, b
$[a, b]$	Range of continuous values from a to b
$\mathcal{A} \cup \mathcal{B}$	Union of sets $\mathcal{A}$ and $\mathcal{B}$
$a \in \mathcal{A}$	a belongs to $\mathcal{A}$
$\sum_{a \in \mathcal{A}}$	Sum over all entities in set $\mathcal{A}$
$f(x)$	Function of x
$\mathcal{L}(\cdot)$	Lagrange function
$\mathcal{D}(\cdot)$	Dual function
$\frac{\partial f}{\partial x}$	Partial derivative of f with respect to x
$\ln(\cdot)$	Natural logarithm
$\log_a(\cdot)$	Logarithm to the base a
$\max[a, b]$	Maximum of the values of a and b
$\min[a, b]$	Minimum of the values of a and b
$[x]_a^b$	$\min[b, \max[x, a]]$
$[x]^+$	$\max[x, 0]$
$a \ll b$	a is very smaller than b
$\mathbb{C}^{a \times b}$	A space of all $a \times b$ matrices with complex entries
$\mathbb{E}_x\{\cdot\}$	Expectation operator over variable x
$\sqrt{a}$	Square root of a
$\lfloor a \rfloor$	The largest integer less than or equal to the value of a
$\|\|x\|\|$	Euclidean norm of matrix or vector x
$\xrightarrow{a.s}$	Almost sure convergence
$\Pr\{a\}$	Probability of event a

Chapter 1

Introduction

1.1 Motivation

1.1.1 Wireless Networks: Trends and Challenges

The tremendous increase in multimedia applications and high-end communication devices like smart phones over the last decade has driven the capacity demand for the currently deployed wireless networks. As such, the future generation of wireless networks need to address a number of challenges. First, with the proliferation of the high data-rate applications that have become ubiquitous nowadays, there should be a huge increase in the data rate supported by the next generation network. Secondly, with the advent of machine-to-machine type communications and the Internet of Things (IoT), the future network will need to support devices that are not only remotely controlled in real time but will also communicate with each other. This presents the requirement for ultra low latency, sometimes of the order of sub-milliseconds, as well as higher reliability beyond what the current network can support. Moreover, as usual, there is the requirement to increase the energy and infrastructure efficiencies. At a high level, these challenges and requirements can be succinctly listed as:

- Supporting very high data rates with better Quality of Service (QoS) in order to accommodate the ever-increasing data-intensive applications,
- Provisioning stringent latency and reliability requirements in order to support the critical machine-to-machine type communication devices and the IoT,

- Enabling network scalability with flexibility so that wide range of devices can be supported with higher energy efficiency,
- Reducing the infrastructure and operational expenditure of the network.

Various techniques and architectural enhancements have been proposed in order to address these challenges. One of the key aspects that has been recently promoted is the idea of dense heterogeneous network (HetNet), which consists of multiple types of small cells in addition to the traditional macro cells and is likely to become the dominant theme in the next-generation wireless network. With the deployment of low-power nodes with smaller coverage area, higher capacity density can be provided for dense areas through traffic off-loading of the macro cell to the small cells. Moreover, the small cell network can be used to serve the coverage holes that cannot be otherwise accommodated by the macro cells. In addition, the energy consumption of the network is significantly reduced as well due to the deployment of low-power nodes.

An important technique envisioned for the next-generation network that can potentially address the above challenges is massive MIMO where the base station (BS) employs low power antenna arrays consisting of potentially a few hundred antennas that simultaneously serves multiple users by relying on spatial multiplexing. By aggressively exploiting the spatial multiplexing, the capacity gains with massive MIMO systems has been shown to be more than 10 times the conventional systems. Moreover, through the coherent superposition of the waveforms emitted by multiple antenna terminals at the intended users, significant improvement in energy efficiency is achieved with massive MIMO.

Another key aspect in the evolution toward the next generation network is the idea of virtualized wireless network (VWN) where the service providers lease the underlying physical network from the network operator and share the available resources in order to provide service to their end-users. Unlike the traditional network, the resources (spectrum, power) in a VWN are not allotted on a fixed basis but rather dynamically allocated to the service providers by considering their requirements and channel conditions that can result in better infrastructure and spectral efficiency while lowering the capital and operating costs as well. However, unlike the traditional macrocell network, due to the additional requirement of maintaining the isolation among the service providers, the user association and dynamic resource allocation schemes require coordination among the BSs in a VWN.

Cloud radio access network (C-RAN) is another key enabling technology for the next-generation network that can leverage the cloud functionality in the baseband unit (BBU) and allows for co-

ordination among the BSs. Basically, with the baseband processing moved to the cloud of BBUs, there is the potential for deploying a central resource allocation scheme that has access to the channel conditions of all users in the whole system and can increase efficient resource utilization. Moreover, deployment of C-RAN allows for simple front-end remote radio heads (RRHs) with the complex processing moved to highly efficient BBU pool, which serves the dual benefit of enabling both scalability and energy efficiency in the network.

With these key enabling technologies, the next-generation network will be able to support high data rate, low latency applications over a wide range of devices with a better energy efficiency. Moreover, by leveraging the functionalities of VWN, the service providers will be able to quickly come up in the market and start service provisioning to the end-users without having to set up the entire physical network, which in turn improves the service provisioning to the end-users as well.

1.1.2 VWN: Requirements

Virtualization, in the general sense, means the abstraction and sharing of the available resources by multiple virtual entities. In the wired networks, virtualization has already been successfully implemented, e.g., virtual private networks (VPNs) and virtual LAN in wide area networks. Recently, virtualization has been proposed in the context of wireless networks and has gained considerable research interest. A virtualized wireless network (VWN) allows multiple service providers (called slices, hereafter) to lease the physical infrastructure from the network operator in order to support their end-users. [1]. By sharing the common physical infrastructure among multiple slices, the physical network is, in effect, decoupled from the end-user service provisioning. This ensures the maximum resource utilization in the network [2]. It also allows new service providers to quickly start the service provisioning without having to setup and maintain the physical network while the network operator can focus on the network efficiency improvement and maintenance aspects. The basic requirements of a VWN can be described as below:

- Isolation among slices: The basic idea in a VWN is to allow multiple slices to utilize the physical infrastructure in order to provide services to their end-users in a reliable manner regardless of the other co-existing slices. Hence, variations in one slice caused by dynamic channel conditions and mobility of users should not affect the services provided to users in other slices. Effective isolation among slices is a prime requirement in a VWN and has been discussed in the literature. For example, [1] proposes a VWN scheme with slices

requiring either bandwidth-based or slot-based reservations from the network operator so that there is isolation among the slices. Similarly, [3] considers an approach with slices having minimum bandwidth as well as power-based reservations.

- Dynamic resource allocation: Unlike the wired networks, the capacity in a wireless network depends on the dynamic channel conditions, and thus, the resources allocated for the slice should be dynamically updated in order to guarantee QoS to the end-users. Moreover, since the ultimate goal of the service providers is to effectively utilize the available resources in order to maximize their revenue, efficient resource allocation is at the core of their interests. To this end, a computationally simple and effective resource allocation algorithm that can adaptively assign available resources to slices in order to meet their requirements is essential.

- Mobility among slices: Apart from the traditional mobility of users from one geographical area to the other within the same service provider, a VWN should support mobility within slices, so that users can experience reliable QoS, and also, the available resources can be utilized to the maximum [4]. In order to support this, advanced user association algorithms are needed that allocate users to BS by not only considering the received SINR but also taking into account the slice requirements including per-slice resource reservations.

1.1.3 VWN: User-Association and Resource-Allocation Issues

As noted, although a VWN is an extension of the concept of virtualization in the wired network, there are a number of challenges that are unique for the wireless network. By considering the above basic requirements, there are specific issues that must be addressed in order to practically realize a VWN which are described next.

User association in a wireless network basically describes how the users are assigned to BSs. In the conventional wireless networks, the basic metric used for user-BS association is the received signal-to-noise-plus-interference ratio (SINR) that a user receives from the BS. With this technique, a user is associated with a BS if the received SINR over a certain time interval is above a pre-defined threshold [5]. Though this criterion is intuitive and simple to implement in order to determine the user association in current networks, in a VWN, there are specific issues with this approach. Unlike the current wireless networks, in a VWN, the physical infrastructure, namely the BS and the backhaul network will be shared among the slices. Moreover, each slice has a

specific resource reservation from the network and by simply assigning users to the BS with the highest received SINR, the constraints on other slices may not be met leading to an infeasible condition. Also, with a dense network and in the high-mobility users scenario, the conventional technique based on received SINR may result in frequent handovers that impose overheads and additional delays [5]. Thus, a dynamic user association policy is required in the VWN that not only takes into account the received signal strength but also considers the isolation constraints of the slices in the system.

One of the key motivations in implementing a VWN is to improve the spectral and energy efficiency of the network [2]. With the current wireless network, since the physical infrastructure and spectrum are owned and operated by the service providers and there is no provision for sharing the resources, the allocated resources to a service provider may not be fully utilized at all times. With a VWN, however, the slices should be allocated spectrum and other resources dynamically so that the resources are efficiently utilized [4]. However, in order to realize this, a dynamic resource allocation algorithm is of utmost importance. Unlike the current wireless networks, in a VWN, the resource allocation policy should not only focus on the overall spectral and energy efficiency, but also the resource reservations of the individual slices in order to provide reliable QoS to the users [3]. Implementing a dynamic resource allocation algorithm that considers all these aspects is a challenging task given the complexity in the optimization problem involving the resource allocation. Various optimization techniques to convexify the resource allocation problems and relaxation techniques are required to formulate computationally tractable approaches.

In order to implement the user association and dynamic resource allocation algorithms in a VWN, co-ordination among the BSs is required so that the algorithms can present efficient solutions by considering the dynamic channel conditions of all users in the system. Moreover, due to the computational complexity of the algorithms, higher processing capability is required in the BS to implement them practically. In order to realize these, one potential solution is to maintain a centralized processing unit that can effectively implement the algorithms and has control over all the BSs. This can be realized by leveraging the cloud functionality through the C-RAN architecture where the processing modules of the BSs are located in the cloud of highly efficient BBU pool [6,7] that has access to the dynamic channel conditions of users in the system.

1.2 Literature Survey

The work presented in this thesis addresses the user association and resource allocation in multi-cell VWNs. In this section, we will briefly present the literature review of the various topics considered in this thesis while further detailed literature survey on each specific topic will be presented in the introduction section of the corresponding chapter.

1.2.1 User-Association and Resource-Allocation

User association and resource allocation in wireless communications have been hot research topics in the last couple of years. However, most of the research works were performed in the context of traditional wireless networks, as VWN is a recently proposed concept. A number of recent research works have addressed the issue of user association specifically in a heterogeneous network (HetNet). Since dense HetNets are likely to become the dominant theme in the next-generation network, conventional max SINR-based approach is unsuitable due to the higher transmit power of the macro BS that forces users to associate with only the macro BS. In order to address this issue, [8] considers a biased user association in HetNets where the received power from the small cell to users is increased by adding a bias so that more users are associated with small cells in order to maintain load balancing. However, this scheme leads to sub-optimal performance as demonstrated in [9] due to the high interference experienced by users associated with the small cell. [10] considers the dynamic user association problem to maximize the total sum-rate of all users in the HetNet and verifies the performance improvement compared with the max SINR approach. Moreover, in [11], a user association scheme is proposed for maximizing the user data rate related utility that tries to maintain the load balancing in HetNet and ensures user fairness. The original problem is relaxed to a fractional association problem and the problem is solved using convex optimization theory. Also, [12] formulates the user association policy from the user-centric view in a stochastic game where the users compete with each other so that they can be served by the BS with the highest bandwidth resources. The algorithm is formulated using the Markov decision process that tracks the long-term reward for the users which leads to a reliable performance in the dynamic channel.

The user association approaches followed in the aforementioned works do not directly apply to a VWN due to a couple of reasons. First, the user association problem in a VWN is different than the traditional network due to the isolation constraint among slices that ensures the QoS guarantee for users in all slices. Moreover, the problems considered in these works generally

deal with only the user association with the BS and do not impose the constraint on the sub-carrier allocation. However, in a practical network deployment with OFDMA as the multiple access scheme, the users need to be allocated sub-carriers from the corresponding BS which has an extra complexity of exclusive sub-carrier constraint within a BS. As per our knowledge, the issue of user association in a VWN with OFDMA has not been considered in the literature.

Existing works on resource allocation in traditional networks have considered the problem of sub-carrier and power allocation in an OFDMA network with an implicit assumption that user association is performed using the max SINR approach. For instance, joint cell, channel and power allocation in multi-cell relay networks is explored in [13], where each user is assigned to the BS with the highest channel gain. [14] proposes a proportional fair resource allocation in a multi-cell OFDMA network aiming to maintain the quality of experience of users by considering a utility function based on the mean opinion score. A similar problem in OFDMA cognitive radio networks is studied in [15], where an iterative algorithm is proposed to solve the sub-carrier and power allocation problem with the aim of maximizing the spectral efficiency of the system. In [16], a resource allocation algorithm is proposed for a two-cell downlink OFDMA network with a fractional frequency reuse scheme between BSs.

1.2.2 Massive MIMO

Recently, the application of massive MIMO has been proposed for the next generation wireless network in order to increase the spectral and energy efficiencies by making use of the degrees of freedom gained from the large number of antennas [17]. There have been a lot of works focusing on the resource allocation in a massive MIMO system. For instance, [18] presents the energy efficiency achievable by a massive MIMO system and discusses various implementation issues related to channel estimation, detection and precoding schemes in a massive MIMO system. Since pilot duration is a significant parameter in the resource allocation problem in a massive MIMO system due to the pilot contamination effect, some research works have addressed the issue of formulating the optimal pilot duration. In [19], an optimal resource allocation algorithm is proposed in a conventional wireless network with massive MIMO that computes the pilot signal power, data signal power as well as pilot duration by considering different transmit power schemes for data signal and training signal transmission. Moreover, in [20], the maximum number of admissible users in a downlink pilot-contaminated time division duplexing (TDD) massive MIMO system is derived and an algorithm is proposed to achieve the maximum individual user

capacity. [21] considers the user association problem in a massive MIMO network and proposes a centralized algorithm based on the network utility maximization. Moreover, the performance of the proposed centralized scheme compared with a decentralized scheme where users connect to BSs with a higher promised throughput demonstrates the improvement obtained from the centralized scheme.

1.2.3 C-RAN

As the C-RAN architecture has been recently proposed to leverage the centralized high processing power of the cloud BBU pool, a number of works have addressed the resource allocation problem by considering the limitations in a C-RAN network. The basic challenges in a C-RAN network involve the limitations imposed on the fronthaul link that connects the RRHs with the BBUs, the user association problem with the RRHs and the limitations on the maximum number of users supported by the BBUs. For instance, in [6], a resource allocation algorithm in a C-RAN with limited fronthaul capacity is proposed by considering a cooperative transmission scheme. Specifically, an optimal resource allocation scheme to minimize the total transmit power in the downlink by considering the QoS constraint of each user is analyzed. Similarly, in [7], the potential performance improvement as well as energy saving benefits of the C-RAN architecture has been demonstrated. In [22], the authors have proposed techniques to improve energy and spectral efficiency by considering a cooperative transmission in a HetNet. Specifically, a computationally efficient precoding scheme has been proposed that tries to reduce the associated power consumption in the network. In [23], an algorithm based on relaxed integer programming (RIP) to jointly associate users to RRHs in a C-RAN is proposed in order to increase the energy efficiency. Specifically, the authors have considered joint user association and beamforming to coordinate interference in the C-RAN and propose an algorithm that tries to maximize the energy efficiency in the network.

1.2.4 Power Allocation in NOMA

Non-orthogonal multiple access (NOMA) has been recently explored as the potential scheme to increase the spectral efficiency and to support massive connectivity in the future network. For instance, [24] proposes NOMA in the context of the next-generation network and highlights the limitations of OFDMA in massive user connectivity and optimal spectral efficiency. Since users are allocated different power allocation coefficients in NOMA, computationally efficient power

allocation algorithms are significant for the practical implementation. [25] proposes a NOMA scheme with SIC at the receiver as the baseline receiver for robust multiple access. Basically, [25] considers two schemes for power allocation, namely the fractional transmission power allocation (FTPA) and the tree-search based transmission power allocation (TTPA). FTPA is similar to the brute-force search where the user sets are selected to maximize the scheduling metric of sum-rate. In order to reduce the complexity of brute-force search, TTPA is proposed where instead of searching for all possible power combination ratios, the redundant ones are discarded during the tree-search. Similarly, [26] considers the dynamic user selection and power allocation for the users in the NOMA with multiple antennas at the BS. The authors propose an iterative algorithm that tries to find the optimal user set that can be served by each beam and then formulate the power allocation algorithm. Moreover, [27] proposes another user selection and power allocation algorithm in NOMA-based system with multiple sub-carriers where users in each sub-carrier are superimposed upon each other. The authors consider the greedy algorithm to assign users to sub-carriers and then use the difference-of-convex (DC) programming approach to formulate the power allocation algorithm within each sub-carrier.

1.2.5 Resource allocation in VWN

Resource allocation in VWN, though being a recently proposed technique, has received significant interest in the literature. Some of the earlier works have focused on the architecture and the design considerations [1], [4], [28]. [1] describes the design and implementation of a network virtualization substrate (NVS) that enables virtualizing the BS resources into slices that can have different policies of resource and rate reservations from the network. Moreover, a dynamic slice provisioning algorithm based on the utility function that assigns a penalty for the slices based on their resource reservations is proposed. In [4], different aspects of VWN including resource discovery and allocation as well as the research challenges have been discussed. Similarly, in [28], the concept of wireless virtualization is extended to the LTE network by formulating a resource sharing algorithm to enable the provision of slices. Moreover, as discussed in [29], virtualization can be extended to the wireless spectrum by the provision of a spectrum virtualization layer (SVL) below the physical layer which enables the sharing of the same RF front end on different parts of the spectrum. Furthermore, in [30], interactions among slices, network operator, and users are modeled as an auction game where the network operator is responsible for spectrum management on a higher level, and each slice focuses on QoS management for its own users. As

the slice-based reservations may not be satisfied sometimes due to the dynamically varying channel conditions, a novel admission control policy is proposed in [3] by considering the channel state information (CSI) of users in each slice to support the QoS of users. In order to guarantee resources to the slices while still maintaining efficient spectrum utilization, [31] proposes an opportunistic algorithm to allocate the resources to virtual operators by differentiating the resource requirements among operators as baseline and fluctuate requirements which ensures that the minimum required QoS is achieved in each virtual operator. In [32], the advantages of full-duplex transmission relay in VWN is investigated. In [33], another resource allocation algorithm is proposed by considering both time and space division multiplexing so that effective isolation is ensured between virtual operators, while the resource utilization is optimized.

1.3 Research Contributions and Thesis Organization

Although virtualization of wireless network has its potential benefits and improvements, its practical realization is possible only after addressing the various technical challenges. In this thesis, we explore various techniques of user association and dynamic resource allocation in a VWN. In particular, as the resource allocation problem in a VWN is inherently complicated and computationally intractable, we use various mathematical tools and convexification techniques to approximate and propose the following efficient algorithms and solutions in different scenarios.

As discussed, user association and dynamic resource allocation in VWN is a challenging issue and is significantly more complicated than in the traditional wireless network. Chapter 2 deals with this issue in a multi-cell OFDMA VWN by formulating the resource allocation problem that jointly associates users to BS and allocates sub-carriers and power while maintaining the minimum reserved rate per slice. By using variable relaxation and various techniques like successive convex approximation (SCA), complementary geometric programming (CGP), we propose a computationally efficient algorithm to maximize the total sum-rate of the system under the given constraints. In order to verify the effectiveness of the proposed algorithm, via simulation results, we compare the performance against the conventional max SINR approach that is used for user association in the traditional networks.

One of the main requirements of a VWN is to maximize the power efficiency of the system. In order to do so, in Chapter 3, we extend the user association and resource allocation problem in a multi-cell VWN with the aim of maximizing the transmit power efficiency, while still preserving the slice isolation requirements. Specifically, we consider the resource allocation problem that

jointly associates users to BS and allocates sub-carriers and power with the objective of minimizing the total transmit power. Numerical results demonstrate the performance improvement of the proposed algorithm compared to the conventional approach of associating users to BS based on the received SINR.

As the current trend of the increasing data rate requirement suggests, the future network needs to support a tremendous increase in the capacity. As such, we extend the resource allocation problem in a VWN where the BS is equipped with massive MIMO in order to utilize the degree of freedom gained from the large number of antennas. However, with massive MIMO, there is increased complexity in channel estimation, which is performed using the pilot signals from the users. In Chapter 4, we focus on the case where the channel estimation is imperfect due to pilot contamination effect and formulate a resource allocation problem to dynamically allocate sub-carriers, power, antennas as well as pilot duration to users in order to maximize the total sum rate of the system. An iterative algorithm based on the dual decomposition method is proposed and the performance of the proposed algorithm is compared with the scenario where the users are allocated with fixed pilot duration.

Unlike the traditional network, the user association and resource allocation algorithms in a VWN need to consider the channel state information of users in all slices. This is only possible with the collaboration among the BSs during the resource allocation. However, with the current system architecture, it may not be feasible to have very low latency backhaul links to connect the BSs. This has led to the consideration of C-RAN, where the baseband processing units are located in the cloud. In Chapter 5, we explore the resource allocation problem in a C-RAN VWN by formulating a joint user association and resource allocation problem that associates users to RRHs/BBUs and allocates power and antennas to users, to maximize the total sum rate of the system. The performance of the proposed scheme is evaluated in both perfect and imperfect channel state information (CSI) estimation scenarios.

Since non-orthogonal multiple access (NOMA) has been recently proposed as a promising scheme to improve the spectral and energy efficiencies of the next-generation wireless network, in Chapter 6, we consider the resource allocation problem in the downlink of a VWN with NOMA. We propose a power-efficient algorithm that allocates power coefficients to users in order to minimize the total transmit power, while ensuring the isolation among the slices in terms of the minimum reserved rate per slice. Numerical results demonstrate the power efficiency improvement achieved with the NOMA scheme as compared to OFDMA.

Finally, Chapter 7 presents concluding remarks and potential extensions of the research works.

Chapter 2

Bandwidth-Efficient Joint User-Association and Resource-Allocation in Multi-cell VWN

This chapter addresses the user-association and resource-allocation problem in a multi-cell VWN where users belong to different slices with each slice requiring a minimum reserved rate from the network. We formulate a bandwidth-efficient joint user-association and resource-allocation scheme that assigns users to BS and allocates the sub-carriers and power to maximize the total achieved sum-rate of the network subject to the rate reservation per slices. The original problem is converted into computationally tractable form using variable relaxation and complementary geometric programming (CGP). The simulation results demonstrate the performance improvement obtained with the proposed iterative algorithm compared to the conventional max SINR-based approach for user association. [1]

2.1 Introduction

In a *virtualized wireless network* (VWN), the physical resources of one network provider, e.g., spectrum, power, and infrastructure are shared among different service providers, also called slices [1, 4]. Generally, each slice comprises of a set of users, and has its own QoS requirements. To harvest the potential advantages of VWN, effective resource allocation is a major concern

[1] Part of this chapter has been published in [34].

which has been addressed in a number of works.

For instance, in [1], a resource management scheme is studied by introducing two types of slices, including rate-based and resource-based slices, where the minimum rate and minimum network resources are preserved for each slice, respectively. Furthermore, in [30], interactions among slices, network operator, and users are modeled as an auction game where the network operator is responsible for spectrum management on a higher level, and each slice focuses on QoS management for its own users. To preserve the QoS of slices from wireless channel fading, the admission control policy is proposed in [3] where the requirement of each slice is adjusted by the channel state information (CSI) of its own users. To extend the feasibility condition of VWN in order to support diverse QoS requirements, [35] considers the use of massive MIMO where the access point is equipped with a large number of antennas. In [33], the combination of time, space and elastic resource allocation for OFDMA system is considered. Advantages of full-duplex transmission relay in VWN have been investigated in [32].

Generally, these works have focused on analyzing the resource allocation problem in a single-cell VWN scenario. However, in practice, the coverage of a specific region may require a set of BSs, in a multi-cell VWN scenario. The key question in such a multi-cell VWN scenario is how to allocate the resources to maintain the QoS of each slice, while improving the total performance of VWN over a specific region. In this chapter, we consider a multi-cell OFDMA based VWN where the coverage of a specific region is provided by a set of BSs serving different groups of users belonging to different slices. The QoS of each slice is represented by its minimum reserved rate. Each user of each slice can be only served by one BS and this BS is not predetermined by the distance or by measuring the average received signal strength of BSs. Consequently, in this setup, the resource sets in the related optimization problem involve the *set of BSs, sub-carriers* and *power* for each user belonging to each slice.

In the problem considered in this chapter, the objective of proposed resource allocation problem is to maximize the total throughput of VWN in the specific region subject to power limitation of BSs, minimum required rate of each slice, and sub-carrier and BS assignment limitations. Based on the limitations of downlink OFDMA transmission, each sub-carrier can be assigned to one user within a cell and each user can be associated to only one BS. Since in this optimization problem, the sub-carrier assignment and BS association are inter-related, we introduce the user association factor (UAF) that jointly determines the BS assignment and sub-carrier allocation as the optimization variable vector. Due to this user-association constraint and the inter-cell interference, the proposed optimization problem is non-convex and NP-hard, suffering from high com-

putational complexity [36]. We apply the frameworks of complementary geometric programming (CGP) and the successive convex approximation (SCA) [37–40] to develop an efficient, iterative, two-step algorithm to solve the proposed problem. For a given power-allocation, the first step derives the optimum user-association solution, and subsequently, with this obtained user-association solution, the second step finds the optimal power allocation. This 2-step optimization process is repeated until convergence. It can be shown that the simplified problem of each step still involves non-convex optimization problem. By applying various transformation and convexification techniques, we develop the analytical framework to transform the non-convex optimization problems encountered in each step into the equivalent lower-bound geometric programming (GP) problems, [39], which can be solved by available software packages, e.g., CVX [41].

Simulation results demonstrate that the proposed approach can significantly outperform the traditional scenario where the BS assignment is based on the largest average SINR, and subsequent sub-carrier and power allocation is derived for the users of each cell. The simulation results reveal that considering UAF can increase the feasibility of resource allocation problem (the required rate of each slice will be satisfied with a higher probability as compared to the traditional approach). Specifically, the proposed algorithm can significantly increase the probability of achieving higher rates for the cell-edge users, resulting better coverage for the VWN.

The rest of this chapter is organized as follows: Section 2.2 presents the system model considered in this problem along with the problem formulation. The proposed algorithm is discussed in Section 2.3 followed by Section 2.4 where we present the numerical results with concluding remarks in Section 2.5.

2.2 System Model and Problem Formulation

We consider the downlink transmission of a VWN where the coverage of a specific area is provided by a set of BSs, i.e., $\mathcal{L} = \{1, \ldots, L\}$. The total bandwidth of W Hz is divided into a set of sub-carriers, $\mathcal{N} = \{1, \ldots, N\}$ and shared by all BSs through orthogonal frequency-division multiple-access (OFDMA). The bandwidth of each sub-carrier, i.e., $\frac{W}{N}$, is assumed to be much smaller than the coherent bandwidth, W_c, of the wireless channel so that the channel response in each sub-carrier is flat. This set of BSs serves a set of slices, $\mathcal{S} = \{1, \ldots, S\}$, where the slice s has a set of users $\mathcal{K}_s = \{1, \ldots, K_s\}$ and requests for a minimum reserved rate of R_s^{rsv} and $K = \sum_{s \in \mathcal{S}} K_s$ is the total number of users. An illustration of the system model with 3 BSs is presented in Fig. 2.1.

Let $h_{l,k_s,n}$ and $P_{l,k_s,n}$ be the channel power gain (also representing the channel state information (CSI))), and the allocated power, respectively, of the link from BS $l \in \mathcal{L}$ to user k_s of slice s on sub-carrier n. Due to the OFDMA limitation, each user is assigned to one BS, and to avoid intra-cell interference, orthogonal sub-carrier assignment is assumed among users in a cell. The binary-valued user association factor (UAF) $\alpha_{l,k_s,n} \in \{0,1\}$ represents both sub-carrier allocation and BS assignment indicator for user k_s of slice s on sub-carrier n of BS l, i.e., $\alpha_{l,k_s,n} = 1$ when BS l allocates sub-carrier n to user k_s, and $\alpha_{l,k_s,n} = 0$, otherwise.

Consider $\mathbf{P} = [P_{l,k_s,n}]_{\forall l,s,k_s,n}$ and $\boldsymbol{\alpha} = [\alpha_{l,k_s,n}]_{\forall l,s,k_s,n}$ as the vectors of all transmit powers and UAFs of users, respectively. The rate of user k_s at sub-carrier n of BS l can be expressed as

$$R_{l,k_s,n}(\mathbf{P}) = \log_2\left[1 + \frac{P_{l,k_s,n}h_{l,k_s,n}}{\sigma^2 + I_{l,k_s,n}}\right], \tag{2.1}$$

where

$$I_{l,k_s,n} = \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in \mathcal{S}} \sum_{\forall k'_s \in \mathcal{K}_s, k'_s \neq k_s} P_{l',k'_s,n} h_{l,k_s,n}$$

is the interference to user k_s in BS l and sub-carrier n, and σ^2 is the noise power. Without loss of generality, noise power is assumed to be equal for all users in all sub-carriers and BSs. From (2.1), the required minimum rate of slice $s \in \mathcal{S}$ can be represented as

$$\text{C2.1}: \quad \sum_{l \in \mathcal{L}} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}) \geq R_s^{\text{rsv}}, \forall s \in \mathcal{S}.$$

We consider the maximum transmit power limitation of each BS as

$$\text{C2.2}: \quad \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in N} P_{l,k_s,n} \leq P_l^{\max}, \quad \forall l \in \mathcal{L},$$

where $P_l^{\max}$ is the maximum transmit power of BS l. Furthermore, the OFDMA exclusive sub-carrier allocation within each cell l can be expressed as

$$\text{C2.3}: \quad \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} \alpha_{l,k_s,n} \leq 1, \qquad \forall l \in \mathcal{L}, \quad \forall n \in \mathcal{N}.$$

In this setup, due to the limitation of multi-cell OFDMA, we restrict the access of each user by

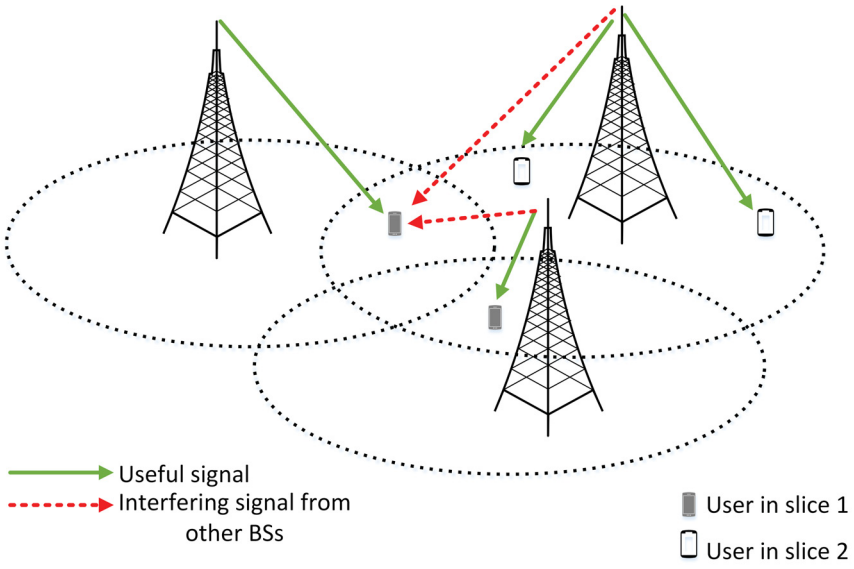


Fig. 2.1 Illustration of a multi-cell VWN with BSs serving users of different slices

the following constraint

$$\text{C2.4}: \quad \Big[\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}\Big]\Big[\sum_{\forall l'\neq l}\sum_{n\in\mathcal{N}} \alpha_{l',k_s,n}\Big] = 0, \quad \forall k_s \in \mathcal{K}_s,\ \forall s \in \mathcal{S},\ \forall l \in \mathcal{L}.$$

C2.4 implies that each user can be associated to only one BS. More specifically, C2.4 ensures when any sub-carrier n is assigned to user k_s by BS l, that user would not be assigned any sub-carriers by other BSs l'.

The joint power, sub-carrier and BS assignment can be formulated as

$$\max_{\boldsymbol{\alpha},\,\mathbf{P}} \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}), \tag{2.2}$$

subject to: C2.1 - C2.4.

The optimization problem (2.2) has a non-convex objective function due to inter-cell interference and involves non-linear constraints with combination of continuous and binary variables, i.e., $\mathbf{P}$ and $\boldsymbol{\alpha}$. In other words, (2.2) is a non-convex mixed-integer, NP-hard optimization problem [36]. Therefore, proposing an efficient algorithm with reasonable computational complexity is desirable.

2.3 Two-step Iterative Algorithm for Joint User-Association and Resource-Allocation

Algorithm 2.1 : Iterative Joint User-Association and Resource-Allocation Algorithm

Initialization: Set $t = 0$, and $\mathbf{P}(t = 0)$ such as power in each sub-carrier of BS l is $P_l^{\max}/K$.
Repeat: Set $t = t + 1$.
Step 1.A User Association:
Initialization for Step 1.A: Set $t_1 = 0, \boldsymbol{\alpha}(t_1) = \boldsymbol{\alpha}(t)$, $\mathbf{P}(t_1) = \mathbf{P}(t)$ and set arbitrary initial for $s_{k_s,n}(t_1)$.
  Repeat: Set $t_1 = t_1 + 1$.
    Step 1.A.1: Update $\lambda_{l,k_s}(t_1), \xi_{l,k_s}(t_1)$, $\nu_{l,k_s,n}(t_1)$, $\eta_{l,k_s,n}(t_1)$ and $\varphi_{l,k_s,n}(t_1)$ using (2.5)-(2.8) and (2.15)-(2.17),
    Step 1.A.2: Find optimal UAF in (2.14) using CVX [41][2],
  Until $||\boldsymbol{\alpha}^*(t_1) - \boldsymbol{\alpha}^*(t_1 - 1)|| \leq \varepsilon_1$,
      set $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}^*(t_1)$.
Step 1.B Power Allocation:
Initialization for Step 1.B: Set $t_2 = 0$, $\boldsymbol{\alpha}(t_2) = \boldsymbol{\alpha}(t)$.
  Repeat: Set $t_2 = t_2 + 1$.
    Step 1.B.1: Update $\kappa_{l,k_s,n}(t_2)$, $\kappa_o(t_2)$ using (2.22),
    Step 1.B.2: Find optimal power allocation according to (2.23) using CVX [41, 42]
  Until $||\mathbf{P}^*(t_2) - \mathbf{P}^*(t_2 - 1)|| \ll \varepsilon_2$,
      set $\mathbf{P}(t) = \mathbf{P}^*(t_2)$.
Until $||\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t - 1)|| \leq \varepsilon_1$, and $||\mathbf{P}^*(t) - \mathbf{P}^*(t - 1)|| \leq \varepsilon_2$.

To tackle the computational complexity of (2.2), we adopt the iterative approach to iteratively find the UAF and power allocation for each user in two steps as shown in Algorithm 2.1. In the first step, for a given power allocation vector, the UAF is considered as the variable of the user-association problem and solved by sub-algorithm 2.1.A (to be discussed in detail in Section 2.3.1).

This derived UAF is then used in in the second step to find the corresponding allocated power as the solution of the power-allocation optimization problem by sub-algorithm 2.1.B (to be discussed in detail in Section 2.3.2). Both these steps are iteratively executed until both the current UAF and power allocation vector solutions are not much different from their values obtained in the previous iteration. In other words, the sequence of the UAF and power allocation vector

[2]CVX chooses its own initialize value for vector $\boldsymbol{\alpha}$ [41], which is applied for our algorithm to check the convergence condition.

solutions can be expressed as

$$\underbrace{\boldsymbol{\alpha}(0) \rightarrow \mathbf{P}(0)}_{\text{Initialization}} \rightarrow \ldots \underbrace{\boldsymbol{\alpha}^*(t) \rightarrow \mathbf{P}^*(t)}_{\text{Iteration } t} \rightarrow \underbrace{\boldsymbol{\alpha}^* \rightarrow \mathbf{P}^*}_{\text{Optimal solution}}, \tag{2.3}$$

where $t > 0$ is the iteration number and $\boldsymbol{\alpha}^*(t)$ and $\mathbf{P}^*(t)$ are the optimal values at the iteration t from convex transformation of related optimization problems in each step. The iterative procedure is stopped when

$$||\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)|| \leq \varepsilon_1 \text{ and } ||\mathbf{P}^*(t) - \mathbf{P}^*(t-1)|| \leq \varepsilon_2$$

where $0 < \varepsilon_1 \ll 1$ and $0 < \varepsilon_2 \ll 1$.

Notably, both the user-association and power-allocation optimization problems are still non-convex and suffer from high computational complexity. To solve them efficiently, we apply complementary geometric programming (CGP) for each step [39] in which via different transformations and convexification approaches, the sequence of lower bound GP based approximation of relative optimization problem is solved as described next.

2.3.1 User-Association Problem

At the iteration t, with given $\mathbf{P}(t)$, we formulate the following user-association optimization problem to maximize the sum rate,

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}(t)), \\ &\text{subject to:} \quad \widetilde{\text{C}}2.1, \text{C2.3, C2.4,} \end{aligned} \tag{2.4}$$

where $R_{l,k_s,n}(\mathbf{P}(t))$ is computed by (1) with $\mathbf{P}(t)$ and

$$\widetilde{\text{C}}2.1 : \sum_{l\in\mathcal{L}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}(t)) \geq R_s^{\text{rsv}}, \ \forall s \in \mathcal{S}.$$

In (2.4), the only optimization variable is $\boldsymbol{\alpha}$, and therefore, (2.4) has lower computational complexity than (2.2). However, it still suffers from the integer optimization variable $\boldsymbol{\alpha}$. In addition, due to C2.4 and the objective function, (2.4) is still a non-convex optimization problem. To overcome the computational complexity of (2.4), we follow the following steps: We first relax

the UAF variable to be continuous as $\alpha_{l,k_s,n} \in [0,1]$. Then, we apply the technique as described in **Proposition 1** to convert (2.4) into the GP formulation by transforming C2.4 into its related linear constraints, and the objective function to the monomial function from **Proposition 2**. Specifically, to have a standard GP formulation, the equality constraint in C2.4 should only involve monomial functions. In the following, we first relax C2.4 and then apply iterative AGMA approximation (as shown in (A.1.9) and (A.1.10) in the Appendix) to get the monomial approximation for C2.4.

Proposition 1: At iteration t_1 in solving (2.4), define $x_{l,k_s}(t_1) = \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1)$ and $y_{k_s}(t_1) = \sum_{l\in\mathcal{L}} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1)$. C2.4 can be approximated by the following constraints.

$$\text{C2.4.1:} \quad z_{l,k_s}^{-1}(t_1) + x_{l,k_s}(t_1) y_{k_s}(t_1) z_{l,k_s}^{-1}(t_1) \leq 1, \quad \forall k_s \in \mathcal{K}_s,\ s \in \mathcal{S},\ \forall l \in \mathcal{L},$$

$$\text{C2.4.2:} \quad \Big[\frac{1}{\lambda_{l,k_s}(t_1)}\Big]^{-\lambda_{l,k_s}(t_1)} z_{l,k_s}(t_1) \times \left[\frac{x_{l,k_s}^2(t_1)}{\xi_{l,k_s}(t_1)}\right]^{-\xi_{l,k_s}(t_1)} \leq 1, \quad \forall k_s \in \mathcal{K}_s,\ \forall s \in \mathcal{S},\ \forall l,$$

$$\text{C2.4.3:} \quad x_{l,k_s}(t_1) \prod_{n\in N} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\nu_{l,k_s,n}(t_1)}\right]^{-\nu_{l,k_s,n}(t_1)} = 1, \quad \forall k_s \in \mathcal{K}_s,\ s \in \mathcal{S},\ \forall l \in \mathcal{L},$$

$$\text{C2.4.4:} \quad y_{k_s}(t_1) \prod_{l\in\mathcal{L}, n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\eta_{l,k_s,n}(t_1)}\right]^{-\eta_{l,k_s,n}(t_1)} = 1, \quad \forall k_s \in \mathcal{K}_s,\ s \in \mathcal{S},\ \forall l \in \mathcal{L},$$

where $z_{l,k_s}(t_1)$ is an auxiliary variable, and,

$$\lambda_{l,k_s}(t_1) = \frac{1}{x_{l,k_s}^2(t_1-1)+1}, \tag{2.5}$$

$$\xi_{l,k_s}(t_1) = \frac{x_{l,k_s}^2(t_1-1)}{x_{l,k_s}^2(t_1-1)+1}, \tag{2.6}$$

$$\nu_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)}{\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1-1)}, \quad , \tag{2.7}$$

$$\eta_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)}{\sum_{l\in\mathcal{L}} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1-1)}, \tag{2.8}$$

for all $k_s \in \mathcal{K}_s$, $s \in \mathcal{S}$, and $\forall l \in \mathcal{L}$.

Proof: From the definition of $x_{l,k_s}(t_1)$ and $y_{k_s}(t_1)$, C2.4 can be rewritten as for all $k_s \in \mathcal{K}_s$, $s \in \mathcal{S}$ and $l \in \mathcal{L}$,

$$x_{l,k_s}(t_1)[y_{k_s}(t_1) - x_{l,k_s}(t_1)] = 0, \tag{2.9}$$

which is not a monomial function. (2.9) can be rewritten as $x_{l,k_s}(t_1) y_{k_s}(t_1) = x_{l,k_s}^2(t_1)$ and by

adding 1 to both the left and right hand sides, we have $x_{l,k_s}(t_1)y_{k_s}(t_1)+1=1+x^2_{l,k_s}(t_1)$ for all $k_s\in\mathcal{K}_s$, $s\in\mathcal{S}$, and $l\in\mathcal{L}$. We define $z_{l,k_s}(t_1)\geq 0$ as an auxiliary variable to relax and convert (2.9) into the posynomial inequalities as follows [37]

$$x_{l,k_s}(t_1)y_{k_s}(t_1)+1\leq z_{l,k_s}(t_1)\leq 1+x^2_{l,k_s}(t_1),\quad \forall k_s\in\mathcal{K}_s,\ \forall s\in\mathcal{S},\ \forall l\in\mathcal{L}. \tag{2.10}$$

The above inequalities can be written as

$$\frac{x_{l,k_s}(t_1)y_{k_s}(t_1)+1}{z_{l,k_s}(t_1)}\leq 1,\quad \frac{z_{l,k_s}(t_1)}{1+x^2_{l,k_s}(t_1)}\leq 1.$$

Now, the above constraints can be approximated using AGMA approximation as

$$\text{C2.4.1:}\quad z^{-1}_{l,k_s}(t_1)+x_{l,k_s}(t_1)y_{k_s}(t_1)z^{-1}_{l,k_s}(t_1)\leq 1, \tag{2.11}$$

$$\text{C2.4.2:}\quad z_{l,k_s}(t_1)\Big[\frac{1}{\lambda_{l,k_s}(t_1)}\Big]^{-\lambda_{l,k_s}(t_1)}\left[\frac{x^2_{l,k_s}(t_1)}{\alpha_{l,k_s}(t_1)}\right]^{-\alpha_{l,k_s}(t_1)}\leq 1, \tag{2.12}$$

where $\lambda_{l,k_s}(t_1)=\frac{1}{x^2_{l,k_s}(t_1-1)+1}$ and $a_{l,k_s}(t_1)=\frac{x^2_{l,k_s}(t_1-1)}{x^2_{l,k_s}(t_1-1)+1}$. Now, C2.4 can be replaced by the following constraints

$$\begin{aligned}
&\text{C2.4.1:}\quad z^{-1}_{l,k_s}(t_1)+x_{l,k_s}(t_1)y_{k_s}(t_1)z^{-1}_{l,k_s}(t_1)\leq 1,\\
&\text{C2.4.2:}\quad z_{l,k_s}(t_1)\big[1/\lambda_{l,k_s}(t_1)\big]^{-\lambda_{l,k_s}(t_1)}\left[\frac{x^2_{l,k_s}(t_1)}{\alpha_{l,k_s}(t_1)}\right]^{-\alpha_{l,k_s}(t_1)}\leq 1,\\
&\widehat{\text{C}}\text{2.4.3}:\quad x_{l,k_s}(t_1)=\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1),\\
&\widehat{\text{C}}\text{2.4.4}:\quad y_{k_s}(t_1)=\sum_{l\in\mathcal{L},n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1),
\end{aligned}$$

Note that via (2.10), the positive condition for the constraints of GP is met [43]. However, the equality constraints in $\widehat{\text{C}}2.4.3$ and $\widehat{\text{C}}2.4.4$ are not monomial functions since we have $x_{l,k_s}(t_1)-\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1)=0$ and $y_{k_s}(t_1)-\sum_{l\in\mathcal{L},n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1)=0$, and, they have negative constraints. To convert $\widehat{\text{C}}2.4.3$ and $\widehat{\text{C}}2.4.4$ to the monomial functions, we again apply AGMA approximation as

$$\text{C2.4.3:} \quad x_{l,k_s}(t_1) \prod_{n\in N} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\nu_{l,k_s,n}(t_1)}\right]^{-\nu_{l,k_s,n}(t_1)} = 1, \tag{2.13}$$

$$\text{C2.4.4:} \quad y_{k_s}(t_1) \prod_{l\in\mathcal{L},n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\eta_{l,k_s,n}(t_1)}\right]^{-\eta_{l,k_s,n}(t_1)} = 1$$

where $\nu_{l,k_s,n}(t_1)$ and $\eta_{l,k_s,n}(t_1)$ are defined in (2.7) and (2.8), respectively.

□

Based on C2.4.1-C2.4.4, C2.4 is transformed and represented by the approximated monomial equalities and posynomial inequalities. Next, we show how we can transform the objective function into the monomial function to reach the GP-based formulation for (2.4).

Proposition 2: Consider the auxiliary variable $x_0 > 0$ and $\Xi_1 \gg 1$. The user association problem (2.4) at each iteration t_1 can be transformed into the following standard GP problem

$$\min_{\boldsymbol{\alpha}(t_1),\, x_0(t_1), z_{l,k_s}(t_1), x_{l,k_s}(t_1), y_{k_s}(t_1)} \quad x_0(t_1), \tag{2.14}$$

$$\text{subject to :} \quad \text{C2.4.1-C2.4.4},$$

$$\Xi_1 \left[\frac{x_0(t_1)}{c_0(t_1)}\right]^{-c_0(t_1)} \prod_{l\in\mathcal{L},s\in\mathcal{S},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1)R_{l,k_s,n}(\mathbf{P}(t))}{c_{l,k_s,n}(t_1)}\right]^{-c_{l,k_s,n}(t_1)} \leq 1,$$

$$\widetilde{\text{C}}2.1.1: \quad R_s^{\text{rsv}} \times \prod_{l\in\mathcal{L},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1)R_{l,k_s,n}(\mathbf{P}(t))}{\varphi_{l,k_s,n}(t_1)}\right]^{-\varphi_{l,k_s,n}(t_1)} \leq 1, \quad \forall s \in \mathcal{S},$$

$$\text{C2.3.1:} \quad \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_\mathcal{S}} \alpha_{l,k_s,n}(t_1) \leq 1, \quad \forall l \in \mathcal{L}, \forall n \in \mathcal{N},$$

where

$$\varphi_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}{\sum_{l\in\mathcal{L}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}, \quad \forall s \in \mathcal{S}, \tag{2.15}$$

$$c_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}{x_0(t_1-1) + \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}, \tag{2.16}$$

$$c_0(t_1) = \frac{x_0(t_1-1)}{x_0(t_1-1) + \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}. \tag{2.17}$$

Proof: To reach the GP based formula for (2.4), we should have minimization over the objective function, i.e.,

$$\min_{\boldsymbol{\alpha}(t_1)} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} -\alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}(t)).$$

Clearly, we have negative elements on the objective function similar to our general formulation of Appendix (A.1.8). To meet the positive conditions of objective function in GP, we consider $\Xi_1 \gg 1$ and rewrite objective function as

$$\Xi_1 - \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1) R_{l,k_s,n}(\mathbf{P}(t))$$

which is always positive. Then, consider a positive auxiliary variable x_0, and rewrite the objective function with this new auxiliary variables

$$\frac{\Xi_1}{x_0 + \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1) R_{l,k_s,n}(\mathbf{P}(t))} \leq 1. \tag{2.18}$$

Now, (2.18) can be rewritten as the product of monomial functions based on the AGMA (discussed in Appendix A.1.2) as

$$\Xi_1 \left[\frac{x_0}{c_0(t_1)}\right]^{c_0(t_1)} \prod_{l\in\mathcal{L},s\in\mathcal{S},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1) R_{l,k_s,n}(\mathbf{P}(t))}{c_{l,k_s,n(t_1)}}\right]^{c_{l,k_s,n}(t_1)} \leq 1, \tag{2.19}$$

where $c_{l,k_s,n}(t_1)$ and $c_0(t_1)$ are updated from (2.16) and (2.17), respectively. Therefore, the total optimization problem can be transformed into (2.14).

□

Now, at each iteration, the optimization problem can be replaced by its GP approximation in (2.14). Iteratively, (2.14) will be solved until achieving the optimal value of UAF value as shown in Step 1.A of Algorithm 2.1.

Proposition 3: With AGMA, the above mentioned approach (Step 1.A) converges to a locally optimal solution that satisfies the KKT conditions of the original problem.

Proof: In [38], it is shown that the conditions for the convergence of the SCA are satisfied and guarantee that the solutions of the series of approximations by AGMA converges to a point that satisfies the KKT conditions of 2.4, i.e., a local maximum is attained [44]. □

2.3.2 Power-Allocation Problem

For a given set of UAFs obtained from Step 1.A, the optimization problem can be formulated as

$$\max_{\mathbf{P}(t_2)} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t) R_{l,k_s,n}(\mathbf{P}(t_2)) \tag{2.20}$$

subject to:

$$\widetilde{\text{C}}2.1.2: \quad \sum_{l\in\mathcal{L}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t) R_{l,k_s,n}(\mathbf{P}(t_2)) \geq R_s^{\text{rsv}}, \forall s \in \mathcal{S},$$

$$\widetilde{\text{C}}2.2.2: \quad \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} P_{l,k_s,n}(t_2) \leq P_l^{\max}, \quad \forall l \in \mathcal{L},$$

where t_2 is the iteration index. Due to interference in the objective function of $R_{l,k_s,n}(\mathbf{P}(t_2))$, (2.20) is a non-convex optimization problem. In order to approximate it to the corresponding GP form, first, we rewrite the objective of (2.20) as

$$\max_{\mathbf{P}(t_2)} \prod_{l\in\mathcal{L},s\in\mathcal{S},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \gamma_{l,k_s,n}(\mathbf{P}(t_2)) \tag{2.21}$$

where

$$\gamma_{l,k_s,n}(\mathbf{P}(t_2)) = \frac{\sigma^2 + I_{l,k_s,n}(t_2) + P_{l,k_s,n}(t_2) h_{l,k_s,n}}{\sigma^2 + I_{l,k_s,n}(t_2)},$$

$$I_{l,k_s,n}(t_2) = \sum_{\forall l'\in\mathcal{L},l'\neq l} \sum_{\forall s\in\mathcal{S}} \sum_{\forall k_s'\in\mathcal{K}_s,k_s'\neq k_s} P_{l',k_s',n}(t_2) h_{l,k_s,n},.$$

Now, using AGMA, $\gamma_{l,k_s,n}^{-1}(\mathbf{P}(t_2))$ can be approximated as

$$\widehat{\gamma}_{l,k_s,n}(\mathbf{P}(t_2)) = (\sigma^2 + I_{l,k_s,n}(t_2)) \left(\frac{\sigma^2}{\kappa_o(t_2)}\right)^{-\kappa_o(t_2)} \prod_{\forall l,s,k_s,n} \left(\frac{P_{l,k_s,n}(t_2) h_{l,k_s,n}}{\kappa_{l,k_s,n}(t_2)}\right)^{-\kappa_{l,k_s,n}(t_2)},$$

where

$$\kappa_{l,k_s,n}(t_2) = \frac{P_{l,k_s,n}(t_2-1) h_{l,k_s,n}}{\sigma^2 + \sum_{l\in\mathcal{L},k_s\in\mathcal{K}_s,s\in\mathcal{S}} P_{l,k_s,n}(t_2-1) h_{l,k_s,n}}, \tag{2.22}$$

$$\kappa_o(t_2) = \frac{\sigma^2}{\sigma^2 + \sum_{s\in\mathcal{L},k_s\in\mathcal{K}_s,s\in\mathcal{S}} P_{l,k_s,n}(t_2-1) h_{l,k_s,n}}.$$

Consequently, (2.20) is transformed into the following standard GP problem

$$\min_{\mathbf{P}(t_2)} \prod_{l\in\mathcal{L},s\in\mathcal{S},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \widehat{\gamma}_{l,k_s,n}(\mathbf{P}(t_2)) \tag{2.23}$$

subject to:

$$\widetilde{\text{C}}2.1.2: \prod_{l\in\mathcal{L},,k_s\in\mathcal{K}_s,n\in\mathcal{N}} \widehat{\gamma}_{l,k_s,n}(\mathbf{P}(t_2)) \leq 2^{-R_s^{\text{rsv}}}, \qquad \forall s\in\mathcal{S},$$

$$\widetilde{\text{C}}2.2.2: \sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} P_{l,k_s,n}(t_2) \leq P_l^{\max}, \quad \forall l\in\mathcal{L},$$

The overall optimization problem is iteratively solved as described in Step 1.B until the power vector converges, i.e., $||\mathbf{P}(t_2)-\mathbf{P}(t_2-1)|| \leq \varepsilon_2$ where $0<\varepsilon_2\ll 1$. Note that Proposition III holds for Step 1.B.

2.4 Numerical Results and Discussions

2.4.1 System Parameters

We consider a multi-cell VWN scenario with $L=4$ BSs and $N=4$ sub-carriers serving $S=2$ slices in a 2×2 square area. The 4 BSs are located at coordinates: (0.5, 0.5), (0.5, 1.5), (1.5, 0.5) and (1.5, 1.5). The channel power gains are based on the path loss and Rayleigh fading model, i.e., $h_{l,k_s,n}=\chi_{l,k_s,n}d_{l,k_s}^{-\zeta}$ where $\zeta=3$ is the path loss exponent, $d_{l,k_s}\in[0.1,1]$ is the distance between the BS l and user k_s normalized to the cell radius and $\chi_{l,k_s,n}$ is the exponential random variable with mean of 1 [45]. We use the noise power in a sub-carrier bandwidth as reference (i.e., normalized to 1 or 0 dB) and hence express transmit power or interference power in dB relative to noise power. For all of the simulations, we set $\Xi_1=10^7$ and $\varepsilon_1=10^{-5}$, $\varepsilon_2=10^{-6}$. In all of the following simulations, for each realization of network, when there exists no feasible solution for the system, i.e., C2.1-C2.4 cannot be hold simultaneously, the corresponding total rate is set to be zero. The simulation results are taken over the average of 100 different channel realizations. For all the following simulations, we set $R^{\text{rsv}}=R_s^{\text{rsv}}$ for all $s\in\mathcal{S}$ and $P^{\max}=P_l^{\max}$ for all $l\in\mathcal{L}$.

2.4.2 Reference Resource-Allocation Algorithm 2.2

For performance comparison, we take as reference, the traditional SINR-based joint sub-carrier and power allocation algorithm as summarized in Algorithm 2.2. Under the SINR criterion, the users are assigned to the BSs that yields the highest average received SINR. In this case, the resource allocation problem is formulated as

$$\max_{\boldsymbol{\alpha}',\mathbf{P}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha'_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}) \tag{2.24}$$

subject to: C2.1 - C2.3,

where $\boldsymbol{\alpha}' = [\alpha'_{l,k_s,n}]_{\forall l,k_s,n}$ and $\alpha'_{l,k_s,n}$ shows the sub-carrier allocation of user k_s on sub-carrier n when it is allocated to the BS l. Clearly, (2.24) is still highly non-convex. In order to show the importance and effects of defining UAF in this context, we apply the similar approach based on CGP to solve (2.24). In other words, Algorithm 2.2 is based on CGP and similar to Algorithm 2.1, except that, in Algorithm 2.2, (2.14) contains only C2.1-C2.3, i.e., C2.4 is removed, since BS-user association is based on the highest average received SINR. When the sub-carrier assignment is solved, the optimal power is derived from Step 1.B similar to Algorithm 2.1, for (2.24). This iterative algorithm is terminated when the convergence conditions are met as summarized in Algorithm 2.2.

Algorithm 2.2 Reference Resource-Allocation Algorithm

Initialization: Set $t_3 = 0$, BS assignment: user k_s is assigned to BS l based on the average received SINR.

Repeat: Set $t_3 = t_3 + 1$.

Step 2.A: Compute $\boldsymbol{\alpha}'^{*}(t_3)$ by using Step **1.A** of Alg. 2.1 except that the BS is assigned based on the signal strength.

Step 2.B: For a fixed $\boldsymbol{\alpha}'^{*}(t_3)$, find the optimal power allocation $\mathbf{P}'(t_3)$ by using Step **1.B** of Alg. 2.1.

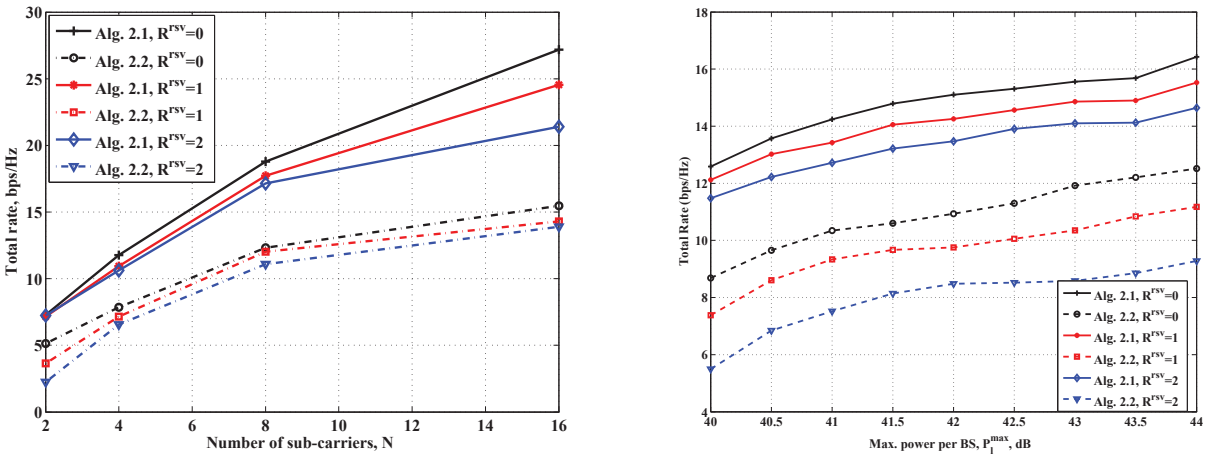
Until $||\boldsymbol{\alpha}'^{*}(t_3) - \boldsymbol{\alpha}'^{*}(t_3-1)|| \leq \varepsilon_1$ and $||\mathbf{P}'^{*}(t_3) - \mathbf{P}'^{*}(t_3-1)|| \leq \varepsilon_2$.

2.4.3 Evaluation of Algorithm 2.1 and Algorithm 2.2

Primarily, we evaluate and compare the total rates achieved by Algorithm 2.1 and Algorithm 2.2 versus the number of sub-carriers and maximum transmit power in Figs. 2.2a and 2.2b, respectively. We set $K_s = 4$ and the 8 users in 2 slices are randomly located in the 2×2 square

area according to a uniform distribution. The results in both Figs. 2.2a and 2.2b indicate that Algorithm 2.1 considerably outperforms Algorithm 2.2 for different values of R_s^{rsv}, N and $P_l^{\max}$.

From Fig. 2.2a, it can be observed that the total rate is increased by increasing the total number of sub-carriers, N, due to the opportunistic nature of fading channels in wireless networks. As expected, with increasing $P_l^{\max}$, the total achievable rate is also increased as shown in Fig. 2.2b. Both figures indicate that by increasing the value of R_s^{rsv}, the total rate decreases because the feasibility region of resource allocation in (2.24) is reduced leading to less total average achieved rate. However, from Fig. 2.2b, increasing R_s^{rsv} has considerable effect on the performance of Algorithm 2.2 as compared to Algorithm 2.1. It can be interpreted as Algorithm 2.1 can efficiently control interference between different cells compared to Algorithm 2.2. Therefore, the chance of feasible power allocation for larger values of R_s^{rsv} is increased by Algorithm 2.1.



(a) (b)

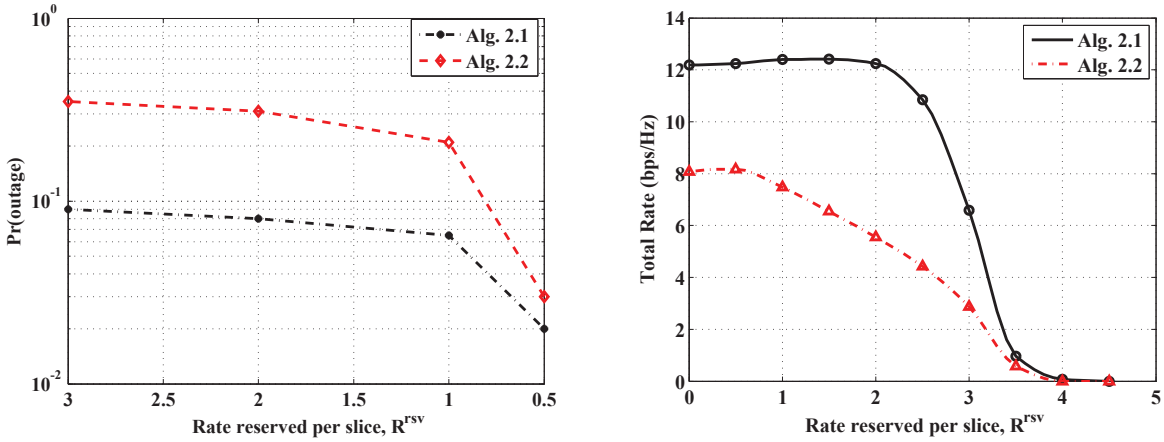
Fig. 2.2 Total rate versus (a) number of sub-carriers, K, and (b) maximum transmit power per BS, $P^{\max}$ (in dB)

To study this point further, we consider the rate-outage probability of C2.1, expressed as

$$\text{Pr(rate-outage)} = \text{Pr}\{\sum_{l\in\mathcal{L}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}R_{l,k_s,n}(\mathbf{P},\boldsymbol{\alpha}) \leq R_s^{\text{rsv}}\}, \quad \forall s\in\mathcal{S}.$$

Via Monte Carlo simulation, we compute Pr(rate-outage) of both Algorithm 2.1 and Algorithm 2.2 for the above-mentioned simulation setting, as depicted in Fig. 2.3a with $N = 8$ and $P_l^{\max} = 40$ dB for all $l \in \mathcal{L}$. The results demonstrate that as the rate reservation per slice

R_s^{rsv} increases, the rate-outage probability of both Algorithm 2.1 and Algorithm 2.2 increases. However, Algorithm 2.2 has larger rate-outage probability compared to the outage probability of Algorithm 2.1, implying that the feasibility region of Algorithm 2.2 is smaller than that for Algorithm 2.1. On the other hand, Algorithm 2.1 can efficiently manage interference in the specific region between different cells as compared to Algorithm 2.2. It is mainly because Algorithm 2.1 has more degrees of freedom to choose the BS and allocate the sub-carriers among users of different slices while the BS assignment is predetermined in Algorithm 2.2. Therefore, the achieved rate of Algorithm 2.2 is less than that of Algorithm 2.1. With increasing R_s^{rsv}, the rate reduction of Algorithm 2.2 is greater than that of Algorithm 2.1, since Algorithm 2.2 cannot manage the interference between different BSs. Hence, Algorithm 2.2 cannot satisfy the minimum rate requirements of slices, leading to reduced VWN efficiency.



(a) (b)

Fig. 2.3 (a) Outage probability, and (b) total rate versus R^{rsv}

For the same setup, in Fig. 2.3b, the total rate of Algorithms 2.1 and 2.2 are plotted for different values of R_s^{rsv}. Fig. 2.3b clearly shows that Algorithm 2.1 yields higher rate than Algorithm 2.2. Note that in all the simulation results, when the problem is infeasible, i.e., there is no power and sub-carrier vectors that can meet the constraint C2.1 for all $s \in \mathcal{S}$, the achieved total rate is set to zero. These simulations highlight the importance of introducing UAF as the joint BS assignment and sub-carrier allocation in the multi-cell wireless networks to manage and control the interference between different cells.

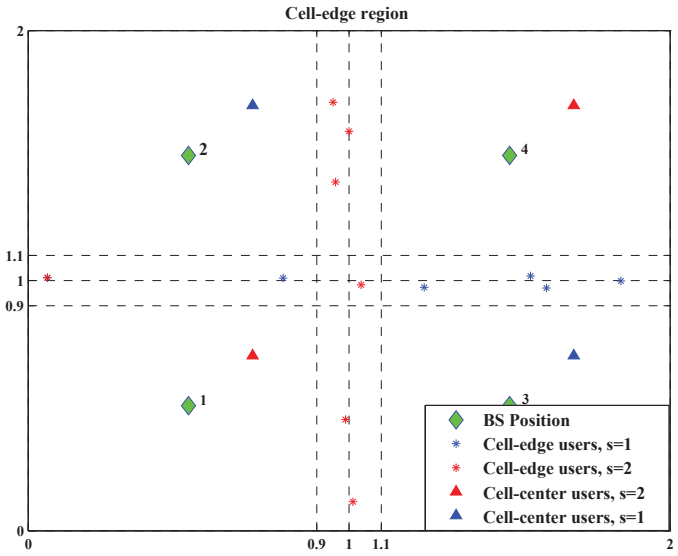
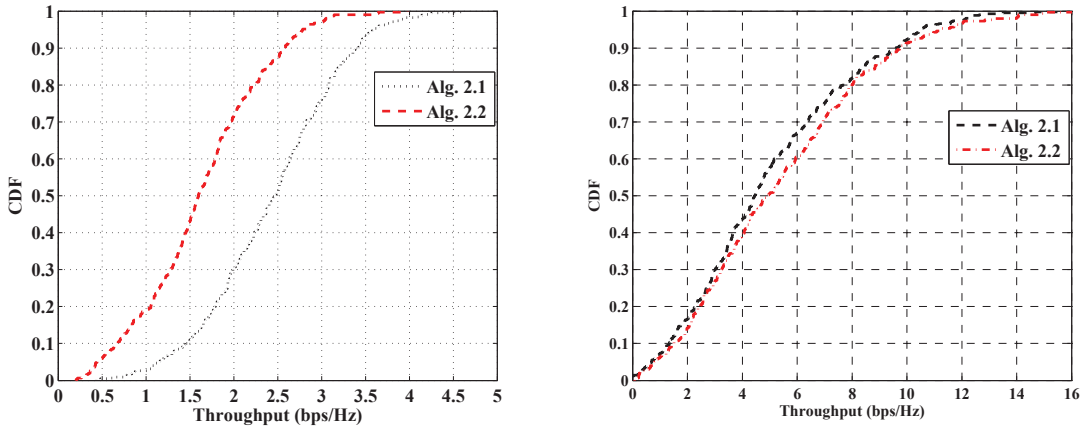


Fig. 2.4 Illustration of network setup to investigate the coverage of multi-cell VWN

2.4.4 Coverage Analysis

In any cellular network, the coverage can be measured by SINR or achieved total rate of users at the cell boundaries. To study the performance of Algorithm 2.1 in terms of coverage, we consider the simulation setup similar to Fig. 2.4 where the majority of users are located in the cell-edge region, consequently, these users experience high interference from other BSs. Therefore, the achieved rate of each user is decreased, which can be considered as the worst-case scenario of coverage analysis.

The cumulative distribution function (CDF) of the total throughput of cell-edge users and cell-center users are depicted in Figs. 2.5a and 2.5b, respectively, for both Algorithms 2.1 and 2.2. It can be seen that Algorithm 2.1 outperforms Algorithm 2.2 for the cell-edge users where 50% of users in the cell-edge achieve a rate of 2.5 bps/Hz by Algorithm 2.1, and around 1.5 bps/Hz via Algorithm 2.2. However, the performance of both algorithms are similar for the cell-center users. It is because via user-association in Algorithm 2.1, the interference among different cells can be controlled while Algorithm 2.2 cannot control the interference through the connectivity of users to different BS and it is pre-determined by the received SINR of reference signal. In other words, Algorithm 2.1 can provide better coverage for cell-edge users for multi-cellular VWN which is desirable from implementation perspective.



(a) (b)

Fig. 2.5 Rate distribution for (a) cell-edge, and (b) cell-center users

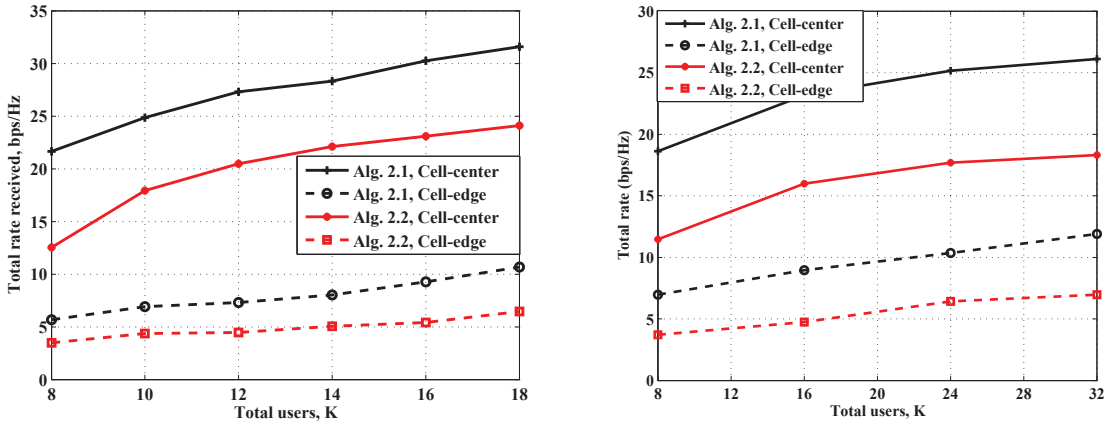
The performance is further investigated with respect to the number of users in the cell-edge in Figs. 2.6a and 2.6b. Algorithm 2.1 can consistently improve the performance of cell-edge users and maintain desirable rate of each slice regardless of the user deployment density as compared to the Algorithm 2.2. For instance, with $K = 18$, for the uniform user distribution, the total rate is increased by 57% from 7 bps/Hz (by Algorithm 2.2) to 11 bps/Hz (by Algorithm 2.1) for cell-edge users and by 33% from 24 bps/Hz (by Algorithm 2.2) to 32 bps/Hz (by Algorithm 2.1) for cell-center users. For non-uniform user distribution, when $K = 32$, the rate is increased by 71% from 7 bps/Hz (by Algorithm 2.2) to 12 bps/Hz (by Algorithm 2.1) for cell-edge users and by 50% from 18 bps/Hz (by Algorithm 2.2) to 27 bps/Hz (by Algorithm 2.1) for cell-center users. These results show the efficiency of applying Algorithm 2.1 in increasing the coverage over the whole network.

2.4.5 Optimality Gap Study

We investigate the performance gap between the optimum solution (by exhaustive search) and the proposed Algorithm 2.1 for $N = 2$ sub-carriers and $K = 4$ users. Fig. 2.7 plots the total rate versus $P^{\max}$ for both Algorithm 2.1 and the exhaustive search. As seen in the figure, the performance of Algorithm 2.1 approaches the performance of exhaustive search as $P^{\max}$ increases because the AGMA approach to convexify the rate is the best fit approximation for the high SINR scenario.

2.4.6 Computational Complexity and Convergence Analysis of Algorithm 2.1

In this section, we investigate the computational complexity and the convergence of Algorithm 2.1. First, we focus on deriving the computational complexity of Algorithm 2.1 analytically. Since CVX is used to solve GP sub-problems with the interior point method in Step 1.A (User-Association problem) and 1.B (Power-Allocation problem), the number of required iterations is $\frac{\log(c/(t^0\varrho))}{\log(\xi)}$ [42], where c is the total number of constraints in (2.14), t^0 is the initial point to approximate the accuracy of interior point method, $0 < \varrho \ll 1$ is the stopping criterion for interior point method, and ξ is used for updating the accuracy of interior point method [42]. As previously discussed, the numbers of constraints in (2.14) are $c_1 = S + LN + 4LK + 1$ for Step 1.A and $c_2 = S + L$ for Step 1.B.



(a) (b)

Fig. 2.6 Total rate for (a) uniform user distribution, and (b) non-uniform user distribution

Moreover, in Steps 1.A and 1.B, for each iteration, the number of computations required to convert the non-convex problems using AGMA into (2.14) and (2.21) is $i_1 = NL^2K + 6NLK + LNSK$ and $i_2 = SLNK + 2LNK$, respectively. Therefore, the order of computational complexity for each step is

- $i_1 \times \frac{\log(c_1/(t_1^0\varrho_1))}{\log(\xi_1)}$ for Step 1. A,
- $i_2 \times \frac{\log(c_2/(t_2^0\varrho_2))}{\log(\xi_2)}$ for Step 1. B.

Based on this analysis, the computational complexity of Step 1.A is significantly higher than that of Step 1.B. Moreover, Step 1.A is more sensitive to N and K than Step 1.B. Since Algorithm 2.1 is a type of block SCA algorithm [46], when (2.2) is feasible, the outer loop of Algorithm 2.1 is converged (Proposition 6 in [47] and Theorem 2 in [46]). For further investigation by simulation, in Fig. 2.8a, the number of iterations required for convergence for Steps 1.A and 1.B versus the total number of sub-carriers, N, is plotted for $K = 8$ and $R^{\text{rsv}} = 2$ bps/Hz. Similarly, in Fig. 2.8b, the number of iterations required for convergence versus the total number of users, i.e., K, for $N = 4$ sub-carriers is plotted in the case of Steps 1.A and 1.B. As N and K increase, the number of iterations required for convergence also increases. The computational complexity of Step 1.A is higher than that of Step 1.B because the total number of constraints for Step 1.A is higher than that for Step 1.B.

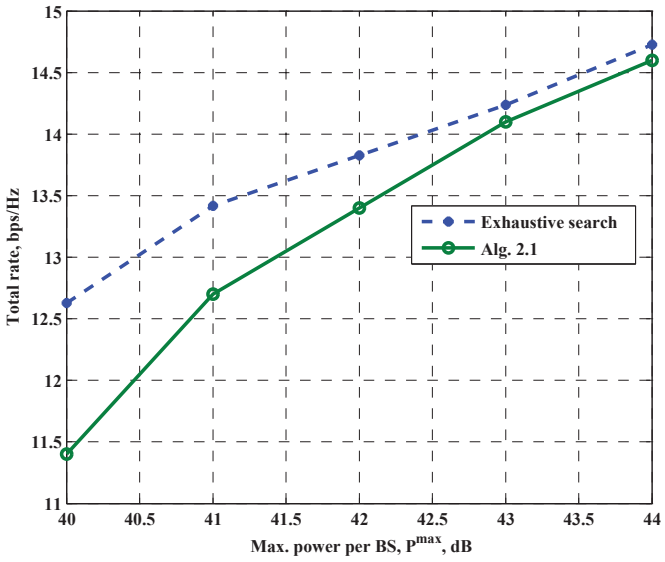
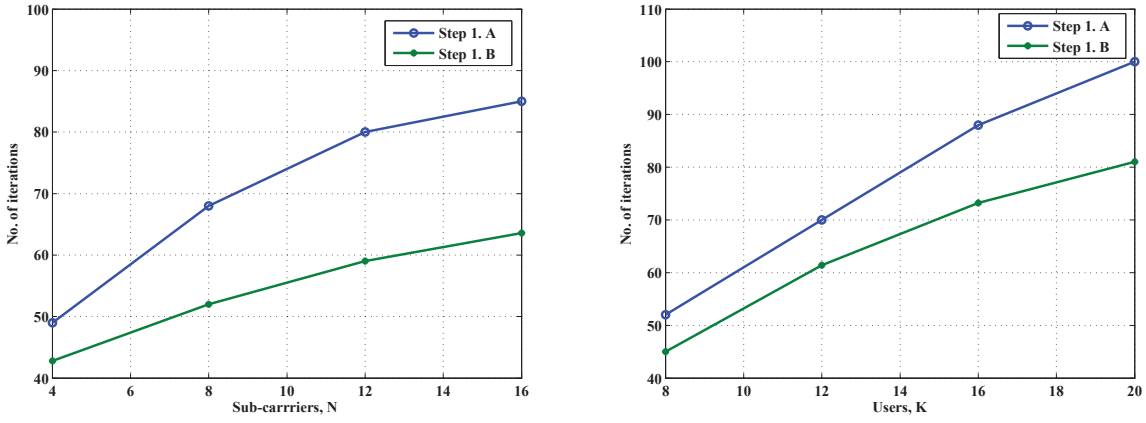


Fig. 2.7 Total rate versus $P^{\max}$ for both exhaustive search and Algorithm 2.1

The major issue of Algorithm 2.1 is that (2.2) can be infeasible, e.g., due to deep fades and/or high interference and C2.1 cannot be met. Therefore, a potential extension work of this chapter could be to consider the admission control policy to adjust R_s^{rsv}.



(a) (b)

Fig. 2.8 Number of required iterations for lower-level iterative algorithms versus (a) number of sub-carriers, N, and (b) total number of users, K

2.5 Concluding Remarks

Joint user association and resource allocation problem for multi-cell OFDMA based VWNs was considered in this chapter. In the proposed setup, we have considered a set of slices (service providers), each has a set of their own users and require a minimum reserved rate. We formulated the related optimization problem based on the new defined optimization variable, called user association factor (UAF), indicating the joint sub-carrier and BS assignment. To solve the proposed non-convex and NP-hard optimization problem, we followed an iterative approach where in each step, one set of optimization variables is derived. However, in each step, the optimization problem is non-convex and NP-hard. To derive the efficient algorithm to solve them, we apply the framework of iterative successive convex approximation via complementary geometric programming (CGP) to transform the non-convex optimization problem into the convex one. Simulation results reveal that, via the proposed approach, the throughput and coverage of VWN, especially for the cell-edge users, are considerably improved as compared to the traditional scenario where the BS is assigned based on the maximum value of SINR.

Chapter 3

Power-Efficient Joint User-Association and Resource-Allocation in Multi-cell VWN

This chapter addresses the power-efficient user-association and resource-allocation problem in a multi-cell VWN. With the aim of minimizing the total transmit power from the BS, we formulate a joint BS assignment, sub-carrier and power allocation problem subject to the slice isolation constraints. This problem is inherently a non-convex optimization problem. To tackle its computational complexity, we apply successive convex approximation (SCA) and complementary geometric programming (CGP) to convert the problem into a computationally efficient formulation and propose an iterative algorithm for solving the problem. Simulation results illustrate the performance enhancement of the VWN achieved through our formulation for different network scenarios.[1]

3.1 Introduction

Maximizing the power efficiency is one of the main goals of any communication network since it directly relates to the operating expenses of the network. As such, there have been a lot of research works done in order to maximize the power efficiency of wireless networks. As mentioned in Chapter 2, while user association and resource allocation is a challenging issue for the practical realization of a VWN, the power efficiency problem adds another level of complexity in the resource allocation problem.

[1] Part of this chapter was presented in [48].

As VWN is still an evolving structure, the existing research works have mainly focused on the power efficient user association problem in the traditional wireless network. For instance, [49] considers the user association problem in a relay network with individual user data requirements. Specifically, the proposed algorithm tries to minimize the total energy consumption from the network, including the physical circuit energy and dynamically associates the users to the macro BS or the relay station depending on the user locations and the received interference. Moreover, [50] considers another the user association problem in a HetNet by using the game theory approach. Specifically, the users are modeled as players and adjust and choose their BS selection strategy while considering their own fairness and long-term payoffs. Similarly, [51] proposes an energy-efficient user association algorithm in massive MIMO enabled HetNet based on the Lagrangian dual analysis that maintains the user throughput requirements as well as the fairness among users. [52] considers the user association problem in a network with coexistence among human and machine-type communication devices and proposes a power efficient algorithm that tries to maintain the QoS of the machine-type devices based on the dual decomposition theory.

The aforementioned works have focused on the user association problem with the objective of maximizing the transmit power efficiency in HetNet. The proposed approaches are not directly applicable to a VWN due to a number of reasons. First, due to the isolation constraint among slices in a VWN, the requirements of the slices need to be met in order to optimally associate users to BSs. Moreover, unlike the problems considered in the above mentioned works, the power efficient user association problem in an OFDMA network is significantly different due to the non-convexity of the problem and the OFDMA exclusive sub-carrier assignment constraint.

We consider a power-efficient user association and resource allocation problem in a multi-cell OFDMA VWN in this chapter. Specifically, the objective of the proposed optimization problem is to minimize the total transmit power of the VWN subject to the constraints. Due to the downlink OFDMA limitation, one sub-carrier in each BS can only be assigned to a single user and each user can only be allocated to a single BS. We use the variable user association factor (UAF) to jointly determine the BS assignment and the sub-carrier allocation to a given user. From the constraints on the UAF and the inter-cell interference, the resource allocation problem is non-convex and NP-hard, which suffers from high computational complexity. To encounter this challenge, we apply CGP and SCA to propose an iterative algorithm with two steps: Step 1 derives the UAF for a given power allocation, and subsequently, for each set of UAF obtained from Step 1, Step 2 derives the optimal power allocation. In each step, we use different transformation techniques like the AGMA to obtain lower-bound GP formulation, which can efficiently be solved by softwares,

like CVX [41].

3.2 System Model with Problem Formulation

We consider the downlink transmission of a VWN, where the coverage of a specific area is provided by a set of BSs, i.e., $\mathcal{L} = \{1, \ldots, L\}$. The total bandwidth of W Hz is shared between the BSs through orthogonal frequency division multiple access (OFDMA) within a set of sub-carriers $\mathcal{N} = \{1, \ldots, N\}$. It is assumed that the bandwidth of each sub-carrier, $\frac{W}{N}$, is much smaller than the coherent bandwidth of the wireless channel, therefore, the CSI in each sub-carrier is flat. The set of BSs serves a set of slices, i.e., $\mathcal{S} = \{1, \ldots, S\}$, where the slice s has its own set of users (denoted by $\mathcal{K}_s = \{1, \ldots, K_s\}$) and requests for a minimum reserved rate of R_s^{rsv}.

Let $h_{l,k_s,n}$ and $P_{l,k_s,n}$ be the CSI of the link from BS $l \in \mathcal{L}$ to user k_s of slice s on sub-carrier n and the allocated power to user k_s of slice s on sub-carrier n, respectively. Due to OFDMA limitation, each user is assigned to only one BS, and to avoid intra-cell interference, orthogonal sub-carrier assignment is assumed among users in a cell. Furthermore, we assume that there is no pre-allocated BS for users. The binary-valued user association factor (UAF), $\alpha_{l,k_s,n} \in \{0, 1\}$ is defined for user k_s of slice s on sub-carrier n of BS l as

$$\alpha_{l,k_s,n} = \begin{cases} 1, & \text{if BS } l \text{ allocates sub-carrier } n \text{ to user } k_s \text{ ,} \\ 0, & \text{otherwise.} \end{cases}$$

From the OFDMA exclusive sub-carrier allocation within each cell l, we have

$$\text{C3.1}: \qquad \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_s} \alpha_{l,k_s,n} \leq 1, \ \forall l \in \mathcal{L}, \ \forall n \in \mathcal{N}.$$

Also, from multi-cell OFDMA limitation, we restrict the access of each user only to one BS as,

$$\text{C3.2}: \left[\sum\nolimits_{n\in\mathcal{N}} \alpha_{l,k_s,n}\right] \left[\sum\nolimits_{\forall l'\neq l} \sum\nolimits_{n\in\mathcal{N}} \alpha_{l',k_s,n}\right] = 0,$$

$\forall k_s \in \mathcal{K}_s, \ \forall s \in \mathcal{S}, \ \forall l \in \mathcal{L}$. Let $\mathbf{P} = [P_{l,k_s,n}]_{\forall l,s,k_s,n}$ and $\boldsymbol{\alpha} = [\alpha_{l,k_s,n}]_{\forall l,s,k_s,n}$ denote the vector of all transmit powers and UAFs of users, respectively. The rate of user k_s of BS l in sub-carrier n

is

$$R_{l,k_s,n}(\mathbf{P}) = \log_2\left[1 + \frac{P_{l,k_s,n}h_{l,k_s,n}}{\sigma^2 + I_{l,k_s,n}}\right], \tag{3.1}$$

where

$$I_{l,k_s,n} = \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in \mathcal{S}} \sum_{\forall k'_s \in \mathcal{K}_s, k'_s \neq k_s} P_{l',k'_s,n} h_{l,k_s,n}$$

and σ^2 is the noise power assumed to be the same for all users in all sub-carriers and BSs. Now, from (3.1), the required minimum rate of slice $s \in \mathcal{S}$ is

$$\text{C3.3}: \quad \sum_{l \in \mathcal{L}} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}) \geq R_s^{\text{rsv}}, \quad \forall s \in \mathcal{S}.$$

Considering C3.1-C3.3, an optimization problem to jointly allocate the BS, sub-carrier and power with the aim of maximizing the energy efficiency can be written as

$$\min_{\boldsymbol{\alpha}, \mathbf{P}} \sum_{l \in \mathcal{L}} \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \alpha_{l,k_s,n} P_{l,k_s,n}, \tag{3.2}$$

subject to: C3.1 - C3.3.

The optimization problem (3.2) has a non-convex objective function due to inter-cell interference and involves non-linear constraints with the combination of continuous and binary variables, i.e., $\mathbf{P}$ and $\boldsymbol{\alpha}$. Consequently, (3.2) is a non-convex mixed-integer, NP-hard optimization problem [36]. Therefore, an efficient algorithm with reasonable computational complexity is needed to solve the resource allocation problem.

3.3 Proposed Algorithm

To solve (3.2), we propose the iterative Algorithm 3.1, where first, we calculate $\boldsymbol{\alpha}$ based on a given $\mathbf{P}$, and then, use the derived $\boldsymbol{\alpha}$ to calculate $\mathbf{P}$ as follows:

$$\underbrace{\boldsymbol{\alpha}(0) \rightarrow \mathbf{P}(0)}_{\text{Initialization}} \rightarrow \ldots \underbrace{\boldsymbol{\alpha}(t)^* \rightarrow \mathbf{P}(t)^*}_{\text{Iteration } t} \rightarrow \underbrace{\boldsymbol{\alpha}^* \rightarrow \mathbf{P}^*}_{\text{Optimal solution}}, \tag{3.3}$$

where $t > 0$ is the iteration number, $\boldsymbol{\alpha}(t)^*$, $\mathbf{P}(t)^*$ are the optimal values at iteration t. The iterative procedure is stopped when

$$||\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)|| \leq \varepsilon_1 \text{ and } ||\mathbf{P}^*(t) - \mathbf{P}^*(t-1)|| \leq \varepsilon_2$$

where $0 < \varepsilon_1 \ll 1$ and $0 < \varepsilon_2 \ll 1$. Note that the user association and power allocation problems are still non-convex and suffer from high computational complexity. In developing an efficient algorithm, we apply CGP via different transformations and convexification approaches to solve the sequence of lower-bound GP-based approximations as discussed in the following sub-sections.

3.3.1 User-Association Problem

For a given $\mathbf{P}(t)$, (3.2) is transformed into

$$\min_{\boldsymbol{\alpha}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} P_{l,k_s,n}(t), \tag{3.4}$$
$$\text{subject to: C3.1, C3.2, } \widetilde{\text{C}}3.3$$

Here, for iteration t, C3.3 is converted to,

$$\widetilde{\text{C}}3.3 : \sum_{l\in\mathcal{L}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}(t)) \geq R_s^{\text{rsv}}, \quad \forall s \in \mathcal{S}.$$

In (3.4), $\boldsymbol{\alpha}$ are the optimization (binary) variables. Thus, (3.4) has much lower computational complexity than (3.2). Furthermore, we relax $\alpha_{l,k_s,n} \in [0,1]$, and then use the AGMA to convert C3.2 and C3.3 as follows. First, we rewrite C3.3 as, $\forall s \in \mathcal{S}$,

$$\widetilde{\text{C}}3.3 : \frac{R_s^{\text{rsv}}}{\sum_{l\in\mathcal{L}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} R_{l,k_s,n}(\mathbf{P}(t))} \leq 1.$$

Now, since the denominator in $\widetilde{\text{C}}3.3$ is a posynomial, for each iteration t_1 of the user association sub-problem, we approximate it and rewrite C3.3 as,

$$\text{C}3.3 : \prod_{s\in\mathcal{L},k_s\in\mathcal{K}_s,n\in\mathcal{N}} R_s^{\text{rsv}} \left[\frac{\alpha_{l,k_s,n}(t_1) R_{l,k_s,n}(\mathbf{P}(t))}{\varphi_{l,k_s,n}(t_1)}\right]^{-\varphi_{l,k_s,n}(t_1)} \leq 1, \tag{3.5}$$

where, $\forall l \in \mathcal{L}, \forall s \in \mathcal{S}, \forall k_s \in \mathcal{K}_S, \forall n \in \mathcal{N}$,

$$\varphi_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}{\sum_{l\in\mathcal{L}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1-1)R_{l,k_s,n}(\mathbf{P}(t))}. \tag{3.6}$$

Next, assuming $x_{l,k_s}(t_1) = \sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1)$ and $y_{k_s}(t_1) = \sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t_1)$, C3.2 can be written as

$$x_{l,k_s}(t_1)[y_{k_s}(t_1) - x_{l,k_s}(t_1)] = 0, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \forall l \in \mathcal{L}. \tag{3.7}$$

To convert (3.7) into a monomial function, we rewrite it as $x_{l,k_s}(t_1)y_{k_s}(t_1) = x^2_{l,k_s}(t_1)$, and by adding 1 to both left and right hand sides, we get $x_{l,k_s}(t_1)y_{k_s}(t_1)+1 = 1+x^2_{l,k_s}(t_1)$. Considering $z_{l,k_s}(t_1) \geq 0$ as an auxiliary variable, (3.7) can be converted into the posynomial inequalities as, [37]

$$x_{l,k_s}(t_1)y_{k_s}(t_1) + 1 \leq z_{l,k_s}(t_1) \leq 1 + x^2_{l,k_s}(t_1), \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \forall l \in \mathcal{L}. \tag{3.8}$$

The above inequalities can be written as

$$\frac{x_{l,k_s}(t_1)y_{k_s}(t_1)+1}{z_{l,k_s}(t_1)} \leq 1, \text{ and } \frac{z_{l,k_s}(t_1)}{1+x^2_{l,k_s}(t_1)} \leq 1.$$

Now, by using AGMA approximation, the above expressions are transformed into

$$\text{C3.2.1: } z^{-1}_{l,k_s}(t_1) + x_{l,k_s}(t_1)y_{k_s}(t_1)z^{-1}_{l,k_s}(t_1) \leq 1,$$
$$\text{C3.2.2:} \left[\frac{1}{\lambda_{l,k_s}(t_1)}\right]^{-\lambda_{l,k_s}(t_1)} z_{l,k_s}(t_1) \times \left[\frac{x^2_{l,k_s}(t_1)}{\xi_{l,k_s}(t_1)}\right]^{-\xi_{l,k_s}(t_1)} \leq 1,$$

where

$$\lambda_{l,k_s}(t_1) = 1/[x^2_{l,k_s}(t_1-1)+1], \tag{3.9}$$
$$\xi_{l,k_s}(t_1) = \frac{x^2_{l,k_s}(t_1-1)}{x^2_{l,k_s}(t_1-1)+1}. \tag{3.10}$$

Now, C3.2 can be replaced by the following constraints

$$\text{C3.2.1}, \text{C3.2.2},$$

$$\text{C3.2.3}: \quad x_{l,k_s}(t_1) = \sum\nolimits_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1),$$
$$\text{C3.2.4}: \quad y_{k_s}(t_1) = \sum\nolimits_{l\in\mathcal{L},n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1).$$

Note that via (3.8), the positive condition for the constraints of GP is met. However, the equality constraints in C3.2.3 and C3.2.4 are not monomials since we have $x_{l,k_s}(t_1) - \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1) = 0$ and $y_{k_s}(t_1) - \sum_{l\in\mathcal{L},n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1) = 0$, and, they have negative constraints. To convert C3.2.3 and C3.2.4 to the monomial functions, we again apply AGMA approximation as

$$\widetilde{\text{C}}3.2.3: \; x_{l,k_s}(t_1) \prod_{n\in N} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\nu_{l,k_s,n}(t_1)}\right]^{-\nu_{l,k_s,n}(t_1)} = 1,$$

$$\widetilde{\text{C}}3.2.4: \; y_{k_s}(t_1) \prod_{l\in\mathcal{L},n\in\mathcal{N}} \left[\frac{\alpha_{l,k_s,n}(t_1)}{\eta_{l,k_s,n}(t_1)}\right]^{-\eta_{l,k_s,n}(t_1)} = 1,$$

where $\nu_{l,k_s,n}(t_1)$ and $\eta_{l,k_s,n}(t_1)$ are given by,

$$\nu_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)}{\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1-1)}, \quad \eta_{l,k_s,n}(t_1) = \frac{\alpha_{l,k_s,n}(t_1-1)}{\sum_{l\in\mathcal{L}}\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1-1)}. \tag{3.11}$$

Hence, the problem for sub-problem 1 can be written as

$$\min_{\boldsymbol{\alpha(t_1)}} \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t_1) P_{l,k_s,n}(t), \tag{3.12}$$
$$\text{s.t: C3.1, C3.2.1 - C3.2.2}, \widetilde{\text{C}}3.2.3 - \widetilde{\text{C}}3.2.4, \widetilde{\text{C}}3.3.$$

Now, (3.12) belongs to GP and can be solved using CVX.

3.3.2 Power Minimization Problem

For given $\boldsymbol{\alpha}$, the power allocation problem is

$$\min_{\mathbf{P}(t_2)} \sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{l,k_s,n}(t) P_{l,k_s,n}(t_2) \tag{3.13}$$
$$\text{subject to:}$$

$$\widetilde{\text{C}}3.3 : \sum_{l\in\mathcal{L}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t)R_{l,k_s,n}(\mathbf{P}(t_2)) \geq R_s^{\text{rsv}}, \quad \forall s\in\mathcal{S},$$

where t_2 is the iteration index. Due to interference in the objective function of $R_{l,k_s,n}(\mathbf{P}(t_2))$, (3.13) is a non-convex optimization problem. We again follow the AGMA approach to convert (3.13) into the equivalent GP problem. First, we rewrite $\widetilde{\text{C}}3.3$ as

$$\prod_{l\in\mathcal{L},k_s\in\mathcal{K}_s,n\in\mathcal{N}}\gamma_{l,k_s,n}(\mathbf{P}(t_2)) \leq 2^{-R_s^{\text{rsv}}}, \quad \forall s\in\mathcal{S},$$

where

$$\gamma_{l,k_s,n}(\mathbf{P}(t_2)) = \frac{\sigma^2 + I_{l,k_s,n}(t_2) + P_{l,k_s,n}(t_2)h_{l,k_s,n}}{\sigma^2 + I_{l,k_s,n}(t_2)},$$

and

$$I_{l,k_s,n}(t_2) = \sum\nolimits_{\forall l'\neq l}\sum\nolimits_{\forall s}\sum\nolimits_{\forall k'_s\neq k_s} P_{l',k'_s,n}(t_2)h_{l,k_s,n}. \tag{3.14}$$

Now using AGMA approach, $\gamma_{l,k_s,n}(\mathbf{P}(t_2))$ can be approximated as

$$\widehat{\gamma}_{l,k_s,n}(\mathbf{P}(t_2)) = (\sigma^2 + I_{l,k_s,n}(t_2))\left(\frac{\sigma^2}{\kappa_o(t_2)}\right)^{-\kappa_o(t_2)} \times \prod_{l\in\mathcal{L},k_s\in\mathcal{K}_s,n\in\mathcal{N}}\left(\frac{P_{l,k_s,n}(t_2)h_{l,k_s,n}}{\kappa_{l,k_s,n}(t_2)}\right)^{-\kappa_{l,k_s,n}(t_2)},$$

where

$$\kappa_{l,k_s,n}(t_2) = \frac{P_{l,k_s,n}(t_2-1)h_{l,k_s,n}}{\sigma^2 + \sum_{l\in\mathcal{L},k_s\in\mathcal{K}_s,s\in\mathcal{S}} P_{l,k_s,n}(t_2-1)h_{l,k_s,n}}, \tag{3.15}$$

$$\kappa_o(t_2) = \frac{\sigma^2}{\sigma^2 + \sum_{l\in\mathcal{L},k_s\in\mathcal{K}_s,s\in\mathcal{S}} P_{l,k_s,n}(t_2-1)h_{l,k_s,n}}.$$

Consequently, (3.13) is transformed into

$$\min_{\mathbf{P}(t_2)}\sum_{l\in\mathcal{L}}\sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}}\alpha_{l,k_s,n}(t)P_{l,k_s,n}(t_2) \tag{3.16}$$

subject to:

$$\bar{\text{C}}3.3: \prod_{l\in\mathcal{L},s\in\mathcal{S},k_s\in\mathcal{K}_s,n\in\mathcal{N}} \widehat{\gamma}_{l,k_s,n}(\mathbf{P}(t_2)) \leq 2^{-R_s^{\text{rsv}}}, \quad \forall s \in \mathcal{S}.$$

The overall optimization problem is iteratively solved as described in Algorithm 3.1 until the UAF and the power vector converge, i.e., $||\boldsymbol{\alpha}(t_1) - \boldsymbol{\alpha}(t_1-1)|| \leq \varepsilon_1$ and $||\mathbf{P}(t_2) - \mathbf{P}(t_2-1)|| \leq \varepsilon_2$ where $0 < \varepsilon_1, \varepsilon_2 \ll 1$.

Algorithm 3.1 : Power-Efficient Joint UA and RA Algorithm

Initialization: Set $t = t_1 = t_2 = 1$, $\boldsymbol{\alpha}(t) = [\mathbf{1}]$, where $\mathbf{1}$ is a vector $C^{1\times KN}$ and $\mathbf{P}(t)$ with power of each sub-carrier of BS m as $P_m^{\max}/K$.
Repeat:
Step 1: Repeat: Set $\boldsymbol{\alpha}(t_1 = 1) = \boldsymbol{\alpha}(t)$, $\mathbf{P}(t_1 = 1) = \mathbf{P}(t)$ and set arbitrary initial for $z_{l,k_s}(t_1)$,
Step 1a: Update $\lambda_{l,k_s}(t_1), \xi_{l,k_s}(t_1), \nu_{l,k_s,n}(t_1)$,
$\eta_{l,k_s,n}(t_1)$ and $\varphi_{l,k_s,n}(t_1)$ as per (3.6) - (16),
Step 1b: Find optimal UAF in (3.12) using CVX [41],
Until $||\boldsymbol{\alpha}^*(t_1) - \boldsymbol{\alpha}^*(t_1-1)|| \leq \varepsilon_1$.
Step 2: Repeat: Obtain $\boldsymbol{\alpha}(t_2 = 1) = \boldsymbol{\alpha}^*(t_1)$.
Step 2a: Update $\kappa_{l,k_s,n}(t_2)$ and $\kappa_o(t_2)$ from (3.15),
Step 2b: Find optimum power from (3.16) via CVX [41],
Until $||\mathbf{P}(t_2) - \mathbf{P}(t_2-1)|| \ll \varepsilon_2$.
Until: $||\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)|| \leq \varepsilon_1$, and $||\mathbf{P}^*(t) - \mathbf{P}^*(t-1)|| \leq \varepsilon_2$.

3.4 Numerical Results and Discussions

We consider a multi-cell VWN with $L = 4$ BSs within a 2×2 square area with $S = 2, N = 4$ and $K = 8$ users uniformly distributed within the area. The channel gains are derived from the Rayleigh fading model as $h_{l,k_s,n} = \chi_{l,k_s,n} d_{l,k_s}^{-\zeta}$ where $\zeta = 3$ is the path loss exponent, $d_{l,k_s} > 0$ is the distance between the BS l and user k_s normalized to the cell radius and $\chi_{l,k_s,n}$ is the exponential random variable with mean of 1. For all of the simulations, we set $\varepsilon_1 = 10^{-5}$ and $\varepsilon_2 = 10^{-6}$, and $R^{\text{rsv}} = R_s^{\text{rsv}}$ for all $s \in \mathcal{S}$ unless otherwise stated. The results are demonstrated based on the average over 100 channel realizations. To compare the performance of our algorithm, we use the sub-optimal approach where the users are assigned to BS based on the received signal strength. Hence, here, the problem is

$$\min_{\boldsymbol{\alpha},\mathbf{P}} \sum_{l\in\mathcal{L}} \sum_{s\in\mathcal{S}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{l,k_s,n} P_{l,k_s,n}, \tag{3.17}$$

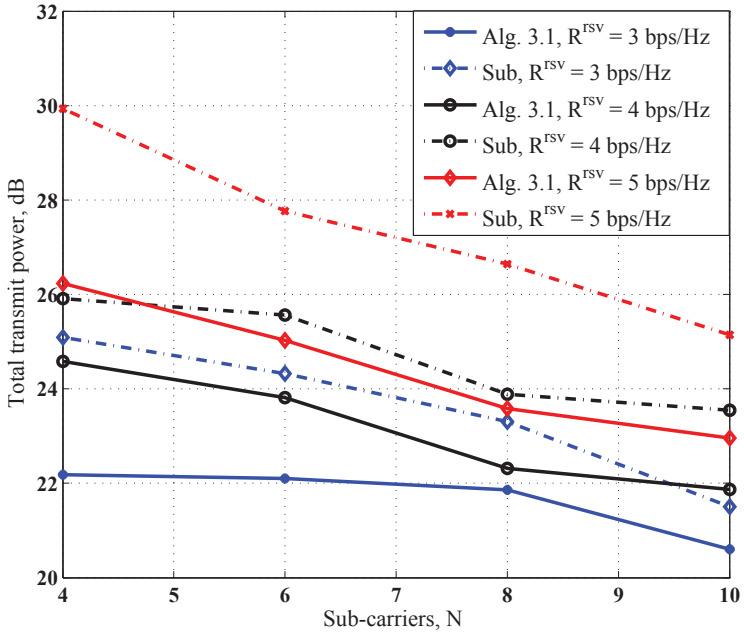


Fig. 3.1 Total transmit power versus N

subject to: C3.1, C3.3.

The proposed Algorithm 3.1 can be applied for solving (3.17) except that all constraints related to C3.2 are removed since the BS assignment is predetermined based on the received signal strength.

Fig. 3.1 shows the total transmit power of all BSs versus R_s^{rsv} and N for both Alg. 3.1 and the sub-optimal approach. As expected, due to the opportunistic nature of fading channels [45], by increasing N, the total transmit power decreases. However, the total transmit power required by the sub-optimal approach is significantly higher than the one required by Alg. 3.1. This is because the BS assignment is based on the highest received signal strength in the sub-optimal approach. Therefore, more transmit power is required to support the rate reservation per slice. Also, with increased rate reserved per slice, the total transmit power increases since the BSs need to transmit at a higher power to support the reserved rate.

Fig. 3.2 plots the total transmit power of all BSs versus K for different R^{rsv}. Clearly, the total transmit power decreases with increasing K for a fixed R^{rsv} because, from the user diversity gain, there is a higher chance of assigning the sub-carriers with less power while fulfilling the minimum

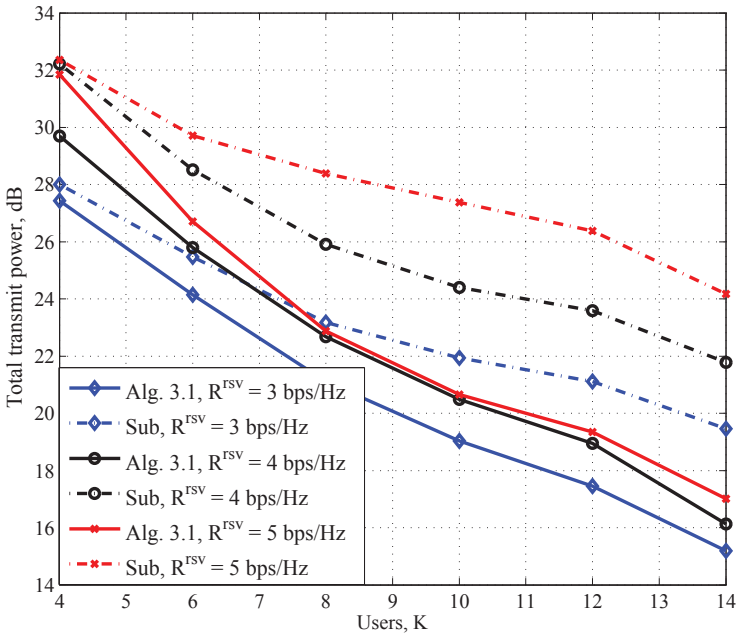


Fig. 3.2 Total transmit power versus K.

rate constraint per slice. Also, similar to Fig. 3.1, the total transmit power of the sub-optimal approach is higher than that of Alg. 3.1 with the difference becoming more distinguishable at higher values of K. Moreover, the total transmit power increases with the increase in R^{rsv}.

In Fig. 3.3, the effects of increasing the number of BSs, i.e., L, is investigated with $K = 32$. It can be observed that increasing L does not have monotonic effect on the transmit power. For instance, for $R^{\text{rsv}} = 3$ and $R^{\text{rsv}} = 4$, the total required power at $L = 4$ is lower than that for $L = 9$ and 16. It is mainly because the inter-cell interference will be increased by increasing the number of BSs, which results in higher power requirement. However, large value of R^{rsv} such as $R^{\text{rsv}} = 5$ bps/Hz, cannot be attained by $L = 4$. Therefore, the number of BSs should be properly designed for a specific region to practically realize a power-efficient and feasible VWN.

3.5 Concluding Remarks

In this chapter, we consider a power-efficient user-association and resource-allocation problem in an OFDMA-based multi-cell VWN where users belonging to different slices require minimum

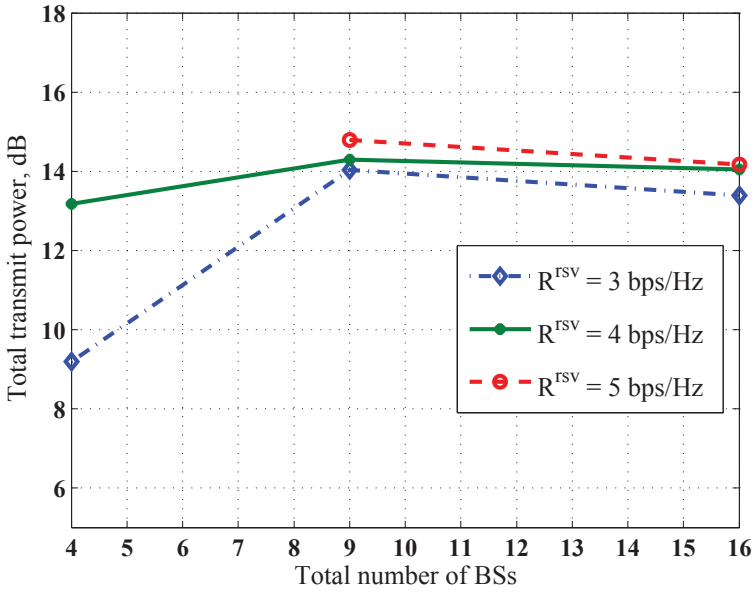


Fig. 3.3 Total transmit power versus L

rate reservation in order to support the QoS of the users. We propose a novel constraint for the joint BS and sub-carrier allocation and convert the highly non-convex optimization problem into more tractable formulation via CGP and SCA. Simulation results reveal the performance gains of the proposed algorithm with respect to different system parameters and indicate the importance of efficient user association in improving the power efficiency.

Chapter 4

Uplink Resource-Allocation in VWN with Massive-MIMO and Dynamic Pilot-Duration

In this chapter, we address the resource allocation problem in a VWN with massive MIMO at the BS. Specifically, we focus on the imperfect CSI estimation scenario in a massive MIMO VWN and formulate the resource allocation problem that jointly assigns sub-carrier, power, antenna and adaptive pilot duration to users with the aim of maximizing the total sum-rate of the system. An iterative algorithm is proposed that tries to dynamically assign these resources to users while maintaining the isolation constraints among slices. The proposed iterative algorithm is compared with the scenario where users have a fixed pilot duration to report their CSI information to the BS. The numerical results demonstrate that there is a significant improvement in the spectral efficiency of the VWN when using adaptive pilot duration compared to the situation with a fixed pilot duration.[1]

4.1 Introduction

Virtualization of wireless networks, as discussed in earlier chapters, is a promising approach to improve network bandwidth- and power-efficiencies wherein the wireless resources, e.g., power, sub-carriers, and antennas are shared between different slices. Unlike the traditional wireless

[1] Part of this chapter was presented in [53].

network, in a VWN, each slice requires a minimum reserved resource (rate or power) from the network in order to support the QoS of its users. Since the performance of a wireless network is affected by the dynamic channel conditions, there may arise situations when the resource requirements of the slices cannot be met giving rise to infeasibility scenario. A number of techniques have been proposed in order to extend the feasibility region of the VWN that could lead to better spectral efficiency.

Application of massive MIMO in the BS of wireless networks has been recently proposed to make extensive use of the degrees of freedom to increase the spectral and energy efficiencies. Channel state information (CSI) estimation in a massive MIMO environment poses a serious challenge and significantly complicates the resource allocation problem. Ideally, the pilot sequences transmitted by the users to assist the BS in estimating the CSI of the users should be mutually orthogonal. However, accommodating a large number of users in the neighboring BSs makes it necessary to reuse the orthogonal sequences among users, which creates an interference and causes imperfect CSI estimation.

There has been a growing interest in the research community to address the issue of mitigating the pilot contamination effects recently. For instance, in [18], various alternatives for precoding and cooperative methods have been presented to alleviate pilot contamination. In [20], the maximum number of admissible users in a downlink pilot-contaminated time division duplexing (TDD) massive MIMO system is derived and an algorithm is proposed to achieve the individual user capacity. In [19], an optimal power and pilot duration allocation algorithm is proposed in a conventional wireless network with massive MIMO by considering different power for data signal and training signal transmission.

In the context of VWNs, [35] has studied the benefits of applying massive MIMO on the achieved network throughput and shown that the feasibility region of optimization problem is considerably expanded and the overall system throughput is improved as a result. Both perfect and imperfect channel estimation have been considered with *fixed* pilot duration in [35]. In this chapter, we aim to investigate whether and to what extent adaptive pilot duration allocation can improve the network performance.

The rest of this chapter is organized as follows: Section 4.2 presents the system model considered in this problem along with the problem formulation. The proposed iterative algorithm is discussed in Section 4.3 followed by Section 4.4 where we present the numerical results with concluding remarks in Section 4.5.

4.2 System Model and Problem Formulation

We consider the uplink transmission of an OFDMA-based VWN with a single BS equipped with an array of $\mathcal{M} = \{1, \ldots, M\}$ antennas. The BS serves a set of slices, $\mathcal{S} = \{1, \cdots, S\}$. Slice $s \in \mathcal{S}$ consists of a set of single-antenna users, $\mathcal{K}_s = \{1, \cdots, K_s\}$ and requires a minimum rate R_s^{rsv}. The total number of users in the system is $K = \sum_{s \in \mathcal{S}} K_s$, and we assume $K \ll M$. An illustration of the system model is presented in Fig. 4.1 along with the uplink pilot model.

The total available bandwidth is divided into a set of sub-carriers, $\mathcal{N} = \{1, \ldots, N\}$ and for OFDMA, each sub-carrier is exclusively assigned to at most a single user at a time. The sub-carrier assignment indicator is denoted as

$$\alpha_{k_s,n} = \begin{cases} 1, & \text{if sub-carrier } n \text{ is allocated to user } k_s, \\ 0, & \text{otherwise.} \end{cases}$$

We denote the sub-carrier assignment vectors for the BS, each slice, and each user as $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_S]$, $\boldsymbol{\alpha}_s = [\boldsymbol{\alpha}_{k_s}]_{k_s=1}^{K_s}$, and $\boldsymbol{\alpha}_{k_s} = [\alpha_{k_s,1}, \cdots, \alpha_{k_s,N}]$, respectively. The uplink pilot duration for user k_s with sub-carrier n in slice s is denoted by $\tau_{k_s,n}$. Correspondingly, the pilot duration vectors of the system, slice s, and user k_s are $\boldsymbol{\tau} = [\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_{\mathcal{S}}]$, $\boldsymbol{\tau}_s = [\boldsymbol{\tau}_{k_s}]_{k_s=1}^{K_s}$, and $\boldsymbol{\tau}_{k_s} = [\tau_{k_s,1}, \cdots, \tau_{k_s,N}]$, respectively.

Let $M_{k_s,n}$ be the number of antennas allocated to user k_s on sub-carrier n. The antenna allocation vector of the system, slice s, and user k_s can be denoted as $\mathbf{M} = [\mathbf{M}_1, \cdots, \mathbf{M}_S]$, $\mathbf{M}_s = [\mathbf{M}_{k_s}]_{k_s=1}^{K_s}$ and $\mathbf{M}_{k_s} = [M_{k_s,1}, \cdots, M_{k_s,N}]$, respectively. Also, let $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_S]$, $\mathbf{P}_s = [\mathbf{P}_{k_s}]_{k_s=1}^{K_s}$ and $\mathbf{P}_{k_s} = [P_{k_s,1}, \cdots, P_{k_s,N}]$ be the allocated power vectors of the system, slice s, and user k_s. respectively, where $P_{k_s,n}$ is the power allocated to user k_s in sub-carrier n. The channel state information vector of user k_s at sub-carrier n is denoted by $\mathbf{h}_{k_s,n} \in \mathbf{C}^{1 \times M_{k_s,n}}$, where the channel amplitude coefficients are given by, [17],

$$h_{k_s,n,m} = \chi_{k_s,n,m} \sqrt{\beta_{k_s}}, \tag{4.1}$$

where $\chi_{k_s,n,m}$ represents the small-scale fading coefficient of the link from the user k_s on the sub-carrier n to the antenna m and β_{k_s} represents the large-scale link power attenuation due to path-loss and shadowing. Note that β_{k_s} is modeled as $\beta_{k_s} = w_{k_s} d_{k_s}^{-\zeta}$ where w_{k_s} is a log-normal random variable representing the shadowing and $d_{k_s} \in [0.1, 1]$ is the normalized distance of the user k_s from the BS with ζ being the path loss exponent.

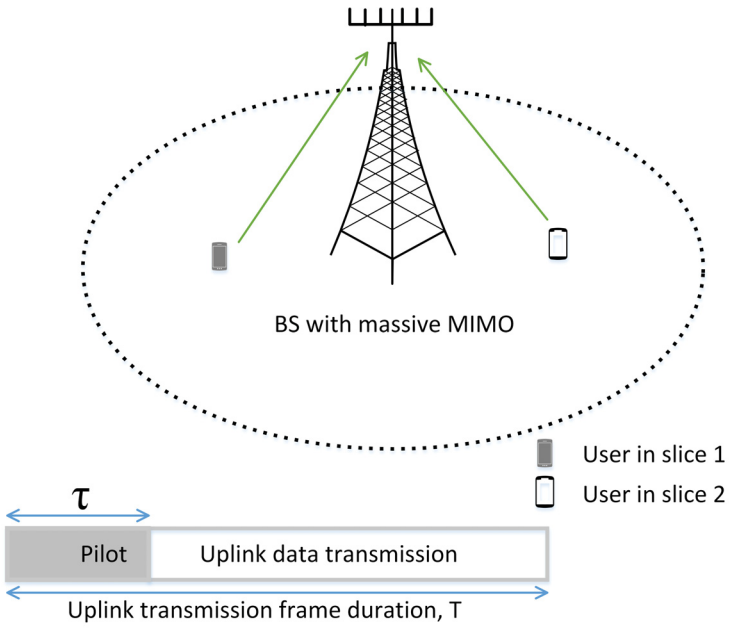


Fig. 4.1 Illustration of massive MIMO VWN and the uplink pilot model

In order to perform sub-carrier and power allocation aiming to maximize the transmission rate, the user-link fading coefficients need to be estimated by using the uplink pilot signals. To this end, all users simultaneously transmit orthogonal pilot sequences at a specific part of the coherence time interval of T. Ideally, the pilot sequences transmitted from the users to their BS and the neighboring BSs should be mutually orthogonal to allow CSI estimation for all the users. However, since the number of orthogonal pilot sequences that can be used within a fixed T and bandwidth is limited, reuse of the orthogonal pilot sequences is unavoidable in neighboring BSs. Therefore, a BS will get the same orthogonal sequences not only from users in its coverage area but also from users in the neighboring areas, causing pilot contamination. In this case, the BS would be unable to accurately estimate the CSI of all users due to interference from users in neighboring BSs.

Consider $\widetilde{\mathbf{h}}_{k_s,n}$ as the estimated channel vector taking into account the pilot contamination error and $\mathbf{f}_{k_s,n} \in \mathrm{C}^{M_{k_s,n}\times 1}$ as the precoding vector for user k_s and on sub-carrier n. Also, let $\mathbf{z}_{k_s,n}$ denote the additive white Gaussian noise (AWGN) vector with zero mean and power spectral density σ^2 assumed to be 1 for simplicity. The received signal at the BS along with the

interference from the pilot contamination can be expressed as,

$$y_{k_s,n} = \sqrt{P_{k_s,n}}\,\widetilde{\mathbf{h}}_{k_s,n}\mathbf{f}_{k_s,n}x_{k_s,n} + \sum\nolimits_{s\in\mathcal{S}}\sum\nolimits_{k'_s\neq k_s}\sqrt{P_{k'_s,n}}\widetilde{\mathbf{h}}_{k'_s,n}\mathbf{f}_{k_s,n}x_{k'_s,n} - \sqrt{P_{k_s,n}}\sum\nolimits_{s\in\mathcal{S}}\sum\nolimits_{k_s\in\mathcal{K}_s}\mathbf{e}_{k_s,n}\mathbf{f}_{k_s,n}x_{k_s,n} + \mathbf{z}_{k_s,n}\mathbf{f}_{k_s,n}, \tag{4.2}$$

Note that the first term in the above expression represents the received signal from user k_s at sub-carrier n after precoding, the second term represents the interference received from other users $k'_s \neq k_s$, the third term represents the error in the CSI estimation from pilot contamination effect with $\mathbf{e}_{k_s,n} = \widetilde{\mathbf{h}}_{k_s,n} - \mathbf{h}_{k_s,n}$. The elements of $\mathbf{e}_{k_s,n}$ are random variables (RVs) with zero mean and variance $\frac{\beta_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1}$ [17], where, $P^{\text{Pilot}}_{k_s,n} = \tau_{k_s,n}P_{k_s,n}$ is the power used for the pilot signal by user k_s at sub-carrier n. Considering the properties of MMSE-based channel estimation, $\mathbf{e}_{k_s,n}$ is independent of $\widetilde{\mathbf{h}}_{k_s,n}, \forall k_s \in \mathcal{K}_s,$ and $\forall n \in \mathcal{N}$. Thus, using MRC precoding where $\mathbf{f}_{k_s,n} = \widetilde{\mathbf{h}}_{k_s,n}$, we get

$$\sqrt{P}_{k_s,n}\sum\nolimits_{s\in\mathcal{S}}\sum\nolimits_{k_s\in\mathcal{K}_s}\mathbf{e}_{k_s,n}\mathbf{f}_{k_s,n}x_{k_s,n} = \sqrt{P}_{k_s,n}\mathbf{e}_{k_s,n}\widetilde{\mathbf{h}}_{k_s,n}x_{k_s,n}$$

Consequently, (4.2) can be written as

$$y_{k_s,n} = \sqrt{P}_{k_s,n}\widetilde{\mathbf{h}}_{k_s,n}\widetilde{\mathbf{h}}^{\text{H}}_{k_s,n}x_{k_s,n} + I_{k_s,n} + \sqrt{P}_{k_s,n}\mathbf{e}_{k_s,n}\widetilde{\mathbf{h}}_{k_s,n}x_{k_s,n} + \mathbf{z}_{k_s,n}\widetilde{\mathbf{h}}^{\text{H}}_{k_s,n} \tag{4.3}$$

where, $I_{k_s,n} = \sum_{\forall s\in\mathcal{S}}\sum_{k'_s\neq k_s}\sqrt{P}_{k_s,n}\widetilde{\mathbf{h}}_{k'_s,n}\widetilde{\mathbf{h}}^{\text{H}}_{k_s,n}x_{k'_s,n}$. Thus, from (4.3), the received SINR at the BS from the user k_s at sub-carrier n can be calculated as

$$\gamma_{k_s,n} = \frac{P_{k_s,n}\|\widetilde{\mathbf{h}}_{k_s,n}\|^4}{||I_{k_s,n}||^2 + P_{k_s,n}\|\mathbf{e}_{k_s,n}\widetilde{\mathbf{h}}^{\text{H}}_{k_s,n}\|^2 + \|\widetilde{\mathbf{h}}_{k_s,n}\|^2}. \tag{4.4}$$

For mutually independent $1\times n$ vectors, $\boldsymbol{p} = [p_1,\ldots,p_n]$ and $\boldsymbol{q} = [q_1,\ldots,q_n]$ whose elements are i.i.d with zero mean, unity variance random variables, it can be shown by the law of large numbers, that $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{p}^{\text{H}}\boldsymbol{q} \xrightarrow{a.s} 0$ where $\xrightarrow{a.s}$ denotes the almost sure convergence. Now, from the law of large numbers for large $M_{k_s,n}$, $||I_{k_s,n}||^2 \to 0$ [17]. As a result, the received SINR becomes,

$$\gamma_{k_s,n} = \frac{P_{k_s,n}\|\widetilde{\mathbf{h}}_{k_s,n}\|^4}{P_{k_s,n}\|\mathbf{e}_{k_s,n}\widetilde{\mathbf{h}}^{\text{H}}_{k_s,n}\|^2 + \|\widetilde{\mathbf{h}}_{k_s,n}\|^2}. \tag{4.5}$$

Substituting the variance of $\widetilde{\mathbf{h}}_{k_s,n}$ to be $\frac{P^{\text{Pilot}}_{k_s,n}\beta^2_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1}$ [17] in the above expression yields the received SINR, $\forall k_s \in \mathcal{K}_s,\ \forall n \in \mathcal{N}$,

$$\begin{aligned}\gamma_{k_s,n} &= \frac{M^2_{k_s,n}P_{k_s,n}(\frac{P^{\text{Pilot}}_{k_s,n}\beta^2_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1})^2}{M_{k_s,n}P_{k_s,n}(\frac{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1})^2(\frac{P^{\text{Pilot}}_{k_s,n}\beta^2_{k_s}}{P^{\text{Pilot}}_{u_s,k}\beta_{k_s}+1})^2 + M_{k_s,n}(\frac{P^{\text{Pilot}}_{k_s,n}\beta^2_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1})^2} \\ &= \frac{M^2_{k_s,n}P_{k_s,n}\zeta_{k_s,n}}{M_{k_s,n}P_{k_s,n}\zeta_{k_s,n}(\frac{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}}{P^{\text{Pilot}}_{u_s,k}\beta_{k_s}+1})^2 + M_{k_s,n}\zeta_{k_s,n}},\end{aligned}$$

where $\zeta_{k_s,n} = (\frac{P^{\text{Pilot}}_{k_s,n}\beta^2_{k_s}}{P^{\text{Pilot}}_{k_s,n}\beta_{k_s}+1})^2$. After some mathematical manipulations, we obtain,

$$\gamma_{k_s,n} = \frac{M_{k_s,n}P^{\text{Pilot}}_{k_s,n}P_{k_s,n}\beta^2_{k_s}}{P_{k_s,n}\beta_{k_s} + P^{\text{Pilot}}_{k_s,n}\beta_{k_s} + 1} \tag{4.6}$$

By substituting $P^{\text{Pilot}}_{k_s,n} = \tau_{k_s,n}P_{k_s,n}$, we obtain, for all $k_s \in \mathcal{K}_s$ and $n \in \mathcal{N}$,

$$\gamma_{k_s,n} = \frac{\tau_{k_s,n}\rho^2_{k_s,n}\beta^2_{k_s}}{1 + (1+\tau_{k_s,n})\rho_{k_s,n}\beta_{k_s}/\sqrt{M_{k_s,n}}}, \tag{4.7}$$

where we substituted $P_{k_s,n} = \rho_{k_s,n}/\sqrt{M_{k_s,n}}$ [17]. As $M_{k_s,n} \to \infty$, we get $\gamma_{k_s,n} = \tau_{k_s,n}\rho^2_{k_s,n}\beta^2_{k_s} = \tau_{k_s,n}P^2_{k_s,n}\beta^2_{k_s}M_{k_s,n}$. Hence, the rate of user k_s on sub-carrier n by considering the pilot duration of $\tau_{k_s,n}$ within the total uplink frame time T is

$$R_{k_s,n} = \frac{T - \tau_{k_s,n}}{T}\log_2\left(1 + \tau_{k_s,n}P^2_{k_s,n}\beta^2_{k_s}M_{k_s,n}\right). \tag{4.8}$$

Thus, the total rate of user k_s is

$$R_{k_s}(\mathbf{P}, \boldsymbol{\alpha}, \mathbf{M}, \boldsymbol{\tau}) = \sum_{n\in\mathcal{N}} \alpha_{k_s,n}R_{k_s,n}$$

which is a function of $\mathbf{P}, \boldsymbol{\alpha},\ \mathbf{M},$ and $\boldsymbol{\tau}$. Now, we define the utility function of a slice s as

$$\begin{aligned}F_s(\mathbf{P}, \boldsymbol{\alpha}, \mathbf{M}, \boldsymbol{\tau}) = &\sum\nolimits_{k_s\in\mathcal{K}_s} R_{k_s}(\mathbf{P}, \boldsymbol{\alpha}, \mathbf{M}, \boldsymbol{\tau}) - c^{\text{M}}_s \sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{k_s,n}M_{k_s,n} \\ &- c^{\text{P}}_s \sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{k_s,n}P_{k_s,n} - c^{\tau}_s \sum_{k_s\in\mathcal{K}_s}\sum_{n\in\mathcal{N}} \alpha_{k_s,n}\tau_{k_s,n},\end{aligned} \tag{4.9}$$

where c_s^{M}, c_s^{P} and c_s^{τ} are pricing factors for the number of allocated antennas, the transmit power and the uplink pilot duration for slice s, respectively. Considering these three pricing factors, $F_s(\mathbf{P}, \boldsymbol{\alpha}, \mathbf{M}, \boldsymbol{\tau})$ is an increasing function of total rate of VWN while it is a decreasing function of the total consumed resource of the VWN for each slice, i.e., power, antenna and pilot duration, which is novel in this context and can be considered as a total revenue minus the costs for each slice. Since in an OFDMA system, a sub-carrier in a BS can be allocated to one user, we have the following constraint:

$$\text{C4.1: } \alpha_{k_s,n} \in \{0,1\}, \text{and} \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_s} \alpha_{k_s,n} \leq 1, \quad \forall n \in \mathcal{N}.$$

The sum of the total transmit power for each user k_s over all sub-carriers n allocated to it poses another constraint, i.e.

$$\text{C4.2: } \sum\nolimits_{n\in\mathcal{N}} \alpha_{k_s,n} P_{k_s,n} \leq P_{k_s}^{\max}, \quad \forall k_s \in \mathcal{K}_s, \quad \forall s \in \mathcal{S},$$

where $P_{k_s}^{\max}$ is the maximum transmit power of user k_s. Since the minimum rate reservation per each slice s is R_s^{rsv}, we have

$$\text{C4.3: } \sum\nolimits_{k_s\in\mathcal{K}_s} R_{k_s} \geq R_s^{\mathrm{rsv}}, \quad \forall s \in \mathcal{S}.$$

If $M_s^{\min}$ and $M_s^{\max}$ are the minimum and maximum numbers of antennas that can be allocated for the slice s, then we have [54]

$$\text{C4.4: } \sum\nolimits_{k_s\in\mathcal{K}_s} \sum\nolimits_{n\in\mathcal{N}} \alpha_{k_s,n} M_{k_s,n} \in \{M_s^{\min}, M_s^{\min+1}, \cdots, M_s^{\max}\},$$

for each slice $s \in \mathcal{S}$. Finally, for each user, the uplink pilot duration $\tau_{k_s,n}$ has a limitation, i.e.,

$$\text{C4.5: } 0 < \tau_{k_s,n} < T, \quad \text{if } \alpha_{k_s,n} = 1.$$

Therefore, the resource allocation problem is

$$\max_{\mathbf{P},\boldsymbol{\alpha},\mathbf{M},\boldsymbol{\tau}} \sum\nolimits_{s\in\mathcal{S}} F_s(\mathbf{P}, \boldsymbol{\alpha}, \mathbf{M}, \boldsymbol{\tau}), \tag{4.10}$$
$$\text{subject to} : \text{C4.1} - \text{C4.5}.$$

Algorithm 4.1 : Iterative Resource-Allocation and Pilot-Duration Allocation Algorithm

Initialization: Set each element of $\boldsymbol{\tau}(l=0)$, $\boldsymbol{\alpha}(l=0)$, $\mathbf{P}_{u_s}(l=0)$, and $\mathbf{N}(l=0)$ to $0.3T$, 1, $P_{u_s}^{\max}/K$ and $N_s^{\max}$, respectively, for all $u_s \in \mathcal{U}_s$ and $s \in \mathcal{S}$. Initialize $l_{\max} \gg 1$, $0 < \varepsilon \ll 1$, $\boldsymbol{\lambda}(l=0)$, $\boldsymbol{\psi}(l=0)$ and $\boldsymbol{\theta}(l=0)$.

Step 1: Obtain the optimum values of $\mathbf{P}_{u_s}^*(l)$, $\boldsymbol{\alpha}^*(l)$, $\mathbf{N}^*(l)$ using Alg. 4.1.A for fixed $\boldsymbol{\tau}(l)$.

Step 2: For fixed $\mathbf{P}_{u_s}^*(l)$, $\boldsymbol{\alpha}^*(l)$, $\mathbf{N}^*(l)$, find the optimal pilot duration $\boldsymbol{\tau}^*(l)$ using Alg. 4.1.B.

Stop: if $||\mathbf{P}_{u_s}^*(l+1) - \mathbf{P}_{u_s}^*(l)|| \leq \varepsilon$, where $||\mathbf{x}||$ is the norm of vector $\mathbf{x}$, otherwise, set $l := l+1$ and go to **Step 1**.

(4.10) is an inherently non-convex optimization problem involving four sets of optimization variables. Thus, finding optimal solution of (4.10) leads to high computational complexity. To tackle this issue, in the next section, we propose a two-step iterative algorithm with a low computational complexity.

4.3 Iterative Algorithm for Joint Resource and Adaptive Pilot Duration Allocation

In the proposed **Algorithm 4.1** to solve problem (4.10), we first apply the framework proposed in [35] to derive the optimum values of $\mathbf{P}$, $\boldsymbol{\alpha}$, and $\mathbf{M}$ for a fixed $\boldsymbol{\tau}$ in Step 1. Then, for the obtained values of $\mathbf{P}$, $\boldsymbol{\alpha}$, and $\mathbf{M}$, we derive the optimal value of $\boldsymbol{\tau}$ in Step 2. The derived solution of iteration l is denoted as $\mathbf{P}^*(l), \boldsymbol{\alpha}^*(l), \mathbf{M}^*(l),$ and $\boldsymbol{\tau}^*(l)$ and the overall solution process can be represented as

$$\underbrace{\boldsymbol{\tau}(0) \rightarrow \mathbf{P}(0), \boldsymbol{\alpha}(0), \mathbf{M}(0)}_{\text{Initialization}} \rightarrow \ldots \underbrace{\boldsymbol{\tau}^*(l) \rightarrow \mathbf{P}^*(l), \boldsymbol{\alpha}^*(l), \mathbf{M}^*(l)}_{\text{Iteration } l} \rightarrow \underbrace{\boldsymbol{\tau}^* \rightarrow \mathbf{P}^*, \boldsymbol{\alpha}^*, \mathbf{M}^*}_{\text{Optimal solution}}$$

Steps 1 and 2 are repeated until the convergence conditions are met. Also, to simplify the algorithm, we just focus on the case that $\tau_{k_s,n} P_{k_s,n}^2 d_{k_s}^2 M_{k_s} \gg 1$ which is a reasonable assumption in massive MIMO context due to large value of M_{k_s}.

4.3.1 Algorithm 4.1.A

For a fixed value of $\boldsymbol{\tau}$, (4.9) involves three sets of variables $\mathbf{P}, \boldsymbol{\alpha}$, and $\mathbf{M}$, which is still a combinatorial function containing both discrete and continuous variables. To simplify this problem, we relax the sub-carrier assignment indicator to be continuous in the interval $[0, 1]$ which will relax

the constraints explained above. Considering the variable transformations $x_{k_s,n} = \alpha_{k_s,n} P_{k_s,n}$ and $y_{k_s,n} = \alpha_{k_s,n} M_{k_s,n}$, the optimization problem of this step can be written as

$$\max_{\boldsymbol{\alpha},\mathbf{x},\mathbf{y}} \sum\nolimits_{s\in\mathcal{S}} \widetilde{F}_s(\boldsymbol{\alpha},\mathbf{x},\mathbf{y}), \tag{4.11}$$

$$\begin{aligned}
&\text{subject to :}\\
&\widetilde{\text{C4.1}} : \alpha_{k_s,n} \in [0,1], \quad \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_s} \alpha_{k_s,n} \le 1, \quad \forall n \in \mathcal{N},\\
&\widetilde{\text{C4.2}} : \sum\nolimits_{n\in\mathcal{N}} x_{k_s,n} \le P_{k_s}^{\max}, \quad \forall k_s \in \mathcal{K}_s, \quad \forall s \in \mathcal{S},\\
&\widetilde{\text{C4.3}} : \sum\nolimits_{k_s\in\mathcal{K}_s} \widetilde{R}_{k_s} \ge R_s^{\text{rsv}}, \quad \forall s \in \mathcal{S},\\
&\widetilde{\text{C4.4}} : M_s^{\min} \le \sum\nolimits_{k_s\in\mathcal{K}_s} \sum\nolimits_{n\in\mathcal{M}} y_{k_s,n} \le M_s^{\max}, \forall s \in \mathcal{S},\\
&\widetilde{\text{C4.5}} : \alpha_{k_s,n}\tau_{k_s,n} < T, \forall n \in \mathcal{N}, \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}
\end{aligned}$$

where $\widetilde{F}_s(\boldsymbol{\alpha},\mathbf{x},\mathbf{y}) = \sum_{k_s\in\mathcal{K}_s} \widetilde{R}_{k_s}(\mathbf{x},\boldsymbol{\alpha},\mathbf{y}) - c_s^{\text{M}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} y_{k_s,n} - c_s^{\text{P}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} x_{k_s,n} - c_s^{\tau} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}\tau_{k_s,n}$.

By assuming $\tau_{k_s,n} P_{k_s,n}^2 \beta_{k_s}^2 M_{k_s} \gg 1$ and new sets of $x_{k_s,n}$, $y_{k_s,n}$ and $\alpha_{k_s,n}$, the total rate of user k_s, R_{k_s}, is a convex function [35]. Consequently, (4.11) is a convex optimization problem which can be solved by Lagrange dual function, defined as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha},\mathbf{x},\mathbf{y},\lambda_{k_s},\phi_s,\theta_s,\psi_s) &= -\sum\nolimits_{s\in\mathcal{S}} \widetilde{F}_s + \sum\nolimits_{k_s\in\mathcal{K}_s} \lambda_{k_s}\Big(\sum_{n\in\mathcal{N}} x_{k_s,n} - P_{\max}\Big)\\
&+ \sum\nolimits_{s\in\mathcal{S}} \phi_s\Big(R_s^{\text{rsv}} - \sum\nolimits_{k_s\in\mathcal{K}_s} \widetilde{R}_{k_s}\Big) + \sum\nolimits_{s\in\mathcal{S}} \theta_s\Big(N_s^{\min} - \sum\nolimits_{k_s\in\mathcal{K}_s} \sum\nolimits_{n\in\mathcal{N}} y_{k_s,n}\Big)\\
&+ \sum\nolimits_{s\in\mathcal{S}} \psi_s\Big(\sum\nolimits_{k_s\in\mathcal{K}_s} \sum\nolimits_{n\in\mathcal{N}} y_{k_s,n} - M_s^{\max}\Big), + \sum\nolimits_{k_s\in\mathcal{K}_s} \eta_{k_s}\Big(\sum_{n\in\mathcal{N}} \alpha_{k_s,n}\tau_{k_s,n} - T\Big),
\end{aligned} \tag{4.12}$$

where the Lagrange multipliers are $\lambda_{k_s}, \phi_s, \theta_s, , \psi_s$ and η_{k_s} for the relaxed constraints $\widetilde{\text{C4.2}}$, $\widetilde{\text{C4.3}}$, $\widetilde{\text{C4.4}}$ and $\widetilde{\text{C4.5}}$, respectively. To solve (4.12), we apply iterative gradient descent method introduced in Alg. 4.1.A where l_1 is the iteration number. In Alg. 4.1.A, the dual variables can be updated as

$$\lambda_{k_s}(l_1+1) = \left[\lambda_{k_s}(l_1) + \delta_{\lambda_{k_s}} \frac{\partial \mathcal{L}}{\partial \lambda_{k_s}}\right]^+, \forall k_s \in \mathcal{K}, \tag{4.13}$$

$$\phi_s(l_1+1) = \left[\phi_s(l_1) + \delta_{\phi_s} \frac{\partial \mathcal{L}}{\partial \phi_s}\right]^+, \quad \forall s \in \mathcal{S}, \tag{4.14}$$

Algorithm 4.1.A: Resource-Allocation Algorithm

Initialization: Set $\boldsymbol{\tau}(l_1 = 0) = 0.3T$, $\boldsymbol{\alpha}(l_1 = 0) = \boldsymbol{\alpha}(l)$, $\mathbf{P}_{k_s}(l_1 = 0) = \mathbf{P}_{k_s}(l)$, and $\mathbf{M}(l_1 = 0) = \mathbf{M}(l)$ for all $k_s \in \mathcal{K}_s$ and $s \in \mathcal{S}$. Initialize $l_1^{\max} \gg 1$, $0 < \varepsilon \ll 1$, $\boldsymbol{\lambda}(l_1 = 0)$, $\boldsymbol{\psi}(l_1 = 0)$ and $\boldsymbol{\theta}(l_1 = 0)$.

1: Update dual variables, λ_{k_s}, ϕ_s, θ_s and ψ_s, by gradient descent method for all $s \in \mathcal{S}$ from (4.13)-(4.17).

2: Using the above updated parameters for iteration $(l_1 + 1)$, compute $P^*_{k_s,n}(l_1 + 1)$ and $M^*_{k_s,n}(l_1 + 1)$ for all $k_s \in \mathcal{K}_s$, and $n \in \mathcal{N}$ using (4.18) and (4.19), respectively.

3: Perform sub-carrier allocation for all $k_s \in \mathcal{K}_s$ and for all $n \in \mathcal{N}$ from (4.21).

Stop if $||\mathbf{P}(l_1 + 1) - \mathbf{P}(l_1)|| \leq \varepsilon$ or $l_1 \geq l_1^{\max}$, otherwise, set $l_1 := l_1 + 1$ and go to **1**.

$$\theta_s(l_1 + 1) = \left[\theta_s(l_1) + \delta_{\theta_s} \frac{\partial \mathcal{L}}{\partial \theta_s}\right]^+, \ \forall s \in \mathcal{S}, \tag{4.15}$$

$$\psi_s(l_1 + 1) = \left[\psi_s(l_1) + \delta_{\psi_s} \frac{\partial \mathcal{L}}{\partial \psi_s}\right]^+, \ \forall s \in \mathcal{S}, \tag{4.16}$$

$$\eta_{k_s}(l_1 + 1) = \left[\eta_{k_s}(l_1) + \delta_{\eta_{k_s}} \frac{\partial \mathcal{L}}{\partial \eta_{k_s}}\right]^+, \forall k_s \in \mathcal{K}_s. \tag{4.17}$$

where $\delta_{\lambda_{k_s}}, \delta_{\phi_s}, \delta_{\theta_s}, \delta_{\psi_s}$ and $\delta_{\eta_{k_s}}$ are the small positive step sizes for their dual variables.

Now, to update the primal variables, $P_{k_s,n}$, $M_{k_s,n}$ and $\alpha_{k_s,n}$, differentiating (4.12) with respect to each of the variables and setting them to zero, we get following expression for the updated value in the iteration $l + 1$ as

$$P_{k_s,n}(l_1 + 1) = \frac{(T - \tau_{k_s,n}(l))}{T} \left[\frac{2(1 + \phi_s)}{\ln(2)(\lambda_{k_s} + c_s^P)}\right]_0^{P_{k_s}^{\max}}, \tag{4.18}$$

and,

$$M_{k_s,n}(l_1 + 1) = \frac{(T - \tau_{k_s,n}(l))}{T} \times \left[\frac{1 + \phi_s}{\ln(2)(\psi_s - \theta_s + c_s^M)}\right]_0^{M_s^{\max}}, \tag{4.19}$$

where $[x]_a^b = \min\{b, \max\{x, a\}\}$. Differentiating (4.12) with respect to $\alpha_{k_s,n}$, we have

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha_{k_s,n}} = \ & c_s^\tau \sum\nolimits_{k_s \in \mathcal{K}_s} \tau_{k_s,n}(l) + \eta_{k_s} \tau_{k_s,n}(l) + \frac{(1 + \phi_s)(T - \tau_{k_s,n}(l))}{T} \times \\ & \left(\log_2(\tau_{k_s,n}(l) P_{k_s,n}(l_1)^2 \beta_{k_s}^2 M_{k_s,n}(l_1) - \frac{3}{\ln(2)}\right),\end{aligned} \tag{4.20}$$

Algorithm 4.1.B: Pilot-Duration Allocation Algorithm

Initialization: Set $\boldsymbol{\alpha}(l_2 = 0) = \boldsymbol{\alpha}^*(l_1)$, $\mathbf{P}_{k_s}(l_2 = 0) = \mathbf{P}^*_{k_s}(l_1)$, and $\mathbf{M}(l_2 = 0) = \mathbf{M}^*(l_1)$ for all $k_s \in \mathcal{K}_s$ and $s \in \mathcal{S}$, and $l_2^{\max} \gg 1$, $a_{k_s,n}(l_2 = 0) = 0$, $b_{k_s,n}(l_2 = 0) = T$, $0 < \varepsilon \ll 1$, $c_{k_s,n}(l_2 = 0) = T/2$, and compute $f_{k_s,n}(l_2 = 0) = f_{k_s,n}(c_{k_s,n}(l_2 = 0))$

Iterative Bisection Method for all k_s and n:

1) For $c_{k_s,n}(l_2) = (a_{k_s,n}(l_2) + b_{k_s,n}(l_2))/2$, calculate $f_{k_s,n}(c_{k_s,n}(l_2))$ and $f_{k_s,n}(a_{k_s,n}(l_2))$

- If $f_{k_s,n}(a_{k_s,n}(l_2)) \times f_{k_s,n}(c_{k_s,n}(l_2)) < 0$, $b_{k_s,n}(l_2) = c_{k_s,n}(l_2)$
- Else, $a_{k_s,n}(l_2) = c_{k_s,n}(l_2)$

2) Consider $\mathbf{c}(l_2) = [c_{k_s,n}(l_2)]_{\forall k_s,\forall n}$ and $\mathbf{f}(l_2) = [f_{k_s,n}(l_2)]_{\forall k_s,\forall n}$,

Stop if:

- $\|\mathbf{c}(l_2) - \mathbf{c}(l_2 - 1)\| < \varepsilon$ or
- $\|\mathbf{f}(l_2) - \mathbf{f}(l_2 - 1)\| < \varepsilon$ or $l_2 > l_2^{\max}$
- Otherwise $l_2 = l_2 + 1$, go to **1**.

and hence, as shown in [54],

$$\alpha^*_{k_s,n}(l_1 + 1) = \begin{cases} 1, & \text{if } \frac{\partial \mathcal{L}}{\partial \alpha_{k_s,n}} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.21}$$

The iteration is repeated until $||\mathbf{P}(l_1 + 1) - \mathbf{P}(l_1)|| \leq \varepsilon$ or $l_1 \geq l_1^{\max}$, where $l_1^{\max}$ is the maximum number of iterations for Algorithm 4.1.A.

4.3.2 Algorithm 4.1.B

For the derived values of $\mathbf{P}^*(l), \mathbf{M}^*(l)$and, $\boldsymbol{\alpha}^*(l)$, the optimization problem for finding $\boldsymbol{\tau}^*(l)$ is

$$\max_{\boldsymbol{\tau}, 0<\tau_{k_s,n}<T} \sum\nolimits_{s\in\mathcal{S}} \widehat{F}_s(\boldsymbol{\tau}), \tag{4.22}$$

where

$$\widehat{F}_s(\boldsymbol{\tau}) = \sum\nolimits_{k_s\in\mathcal{K}_s} R_{k_s}(\boldsymbol{\tau}) - c_s^{\mathrm{M}} \sum_{k_s\in\mathcal{K}_s} \sum_{n\in\mathcal{N}} \alpha_{k_s,n}(l) M_{k_s,n}(l) - \tag{4.23}$$

$$- c_s^{\mathrm{P}} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(l) P_{k_s,n}(l) - c_s^{\tau} \sum_{k_s \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \alpha_{k_s,n}(l) \tau_{k_s,n}.$$

The optimum value of $\boldsymbol{\tau}$ from (4.22) can be obtained by setting $\frac{\partial \widehat{F}_s(\boldsymbol{\tau})}{\partial \tau_{k_s,n}} = 0$ where $0 < \tau_{k_s,n} < T$. To derive this optimal value, we apply the iterative bisection method where we consider

$$f_{k_s,n}(\tau_{k_s,n}) = \frac{T}{\tau_{k_s,n}} - \log_2(\tau_{k_s,n} P_{k_s,n}^{*2}(l) \beta_{k_s,n}^2 M_{k_s,n}^*(l) - 1 - c_s^{\tau} T,$$

and the iteration number l_2 as summarized in Alg. 4.1.B.

4.4 Numerical Results and Discussions

To study the performance of Algorithm 4.1, we simulate a scenario of a VWN with a single BS serving two slices each with $K_s = 4$ users per slice where $P_{k_s}^{\max} = 0$ dB and $R^{\mathrm{rsv}} = R_s^{\mathrm{rsv}} = 2$ bps/Hz. The users are distributed uniformly in the coverage area of the BS. The total number of sub-carriers is $N = 4$ and the total transmission frame duration is set to $T = 1$ s. To study the effects of changing pricing factors, we consider 3 scenarios: (1) Set 1 where $c_s^{\tau} = c_s^M = c_s^P = 0$, (2) Set 2 where $c_s^{\tau} = 0.5$, $c_s^P = 1$, and $c_s^M = 0.07$ and (3) Set 3 where $c_s^{\tau} = 1$, $c_g^P = 2$ and $c_s^M = 0.09$. Obviously, Set 3 has more restricted price parameters than Set 2 and Set 1, while Set 2 has moderate pricing factors compared to other sets. We compare Alg. 4.1 with the approach using *fixed* $\boldsymbol{\tau}$ where $\tau_{k_s,n}/T = 0.3$ for all k_s and n. The simulation results are averaged over 100 different channel realizations and we set the total rate to zero when there is an infeasibility in the solution.

Fig. 4.2 plots the total rate versus R^{rsv} for the three different sets by applying Algorithm 4.1 as well as for a fixed pilot duration. The results indicate that the total rate achieved decreases with increasing R^{rsv} due to the fact that at higher R^{rsv}, C4.3 cannot be fulfilled all the time. Since we set the total rate to zero when there is an infeasibility, the average rate decreases with increasing R^{rsv}. The overall system throughput is improved by using adaptive pilot duration in Alg. 4.1. Moreover, the total rate decreases as the values of c_s^{τ}, c_s^P and c_s^M increase. This is obvious since the utility function (4.9) is defined as a non-increasing function of pricing factors.

From Fig. 4.3, the total rate increases with increasing $M_s^{\max}$ as expected due to the multiplexing gain of massive MIMO. Again, the overall system performance is improved by adaptive pilot duration via Alg. 4.1 as compared to the approach using *fixed* pilot duration. Similarly, as the

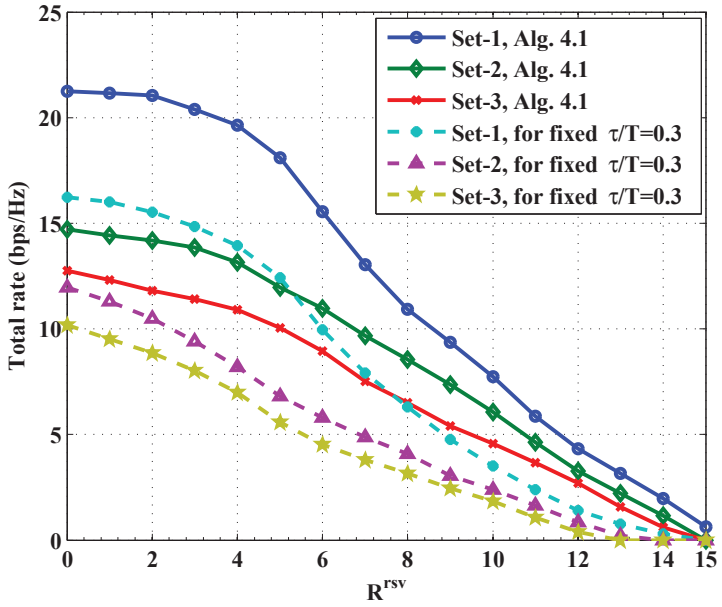


Fig. 4.2 Total rate versus R^{rsv}

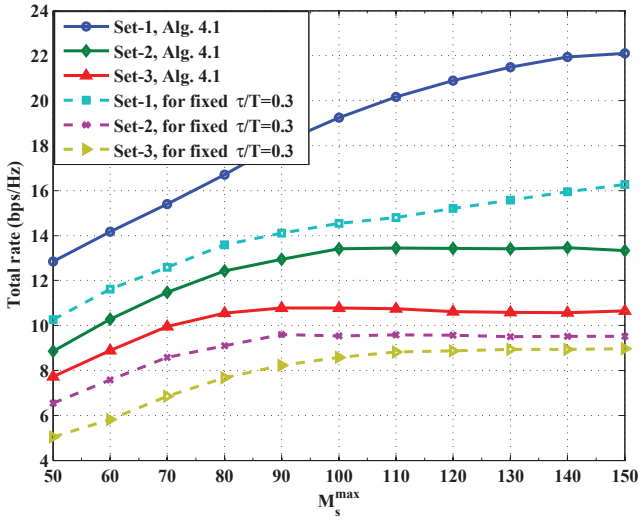


Fig. 4.3 Total rate versus $M_s^{\max}$

costs increase across Sets 1, 2, and 3, the total achieved rate decreases as expected by (4.9).

To get more insight about the effects of the pricing factors on the VWN performance, in Fig. 4.4 and 4.5, we plot the total rate versus c_s^τ and c_s^M with c_s^P fixed, and c_s^τ and c_s^P with c_s^M fixed, respectively. As seen in these two figures, the total rate is not a convex function with respect to

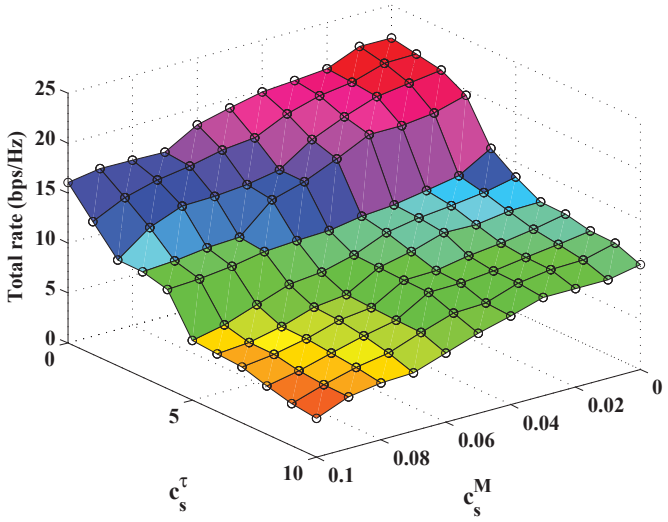


Fig. 4.4 Total rate versus c_s^τ and c_s^M for fixed c_s^P

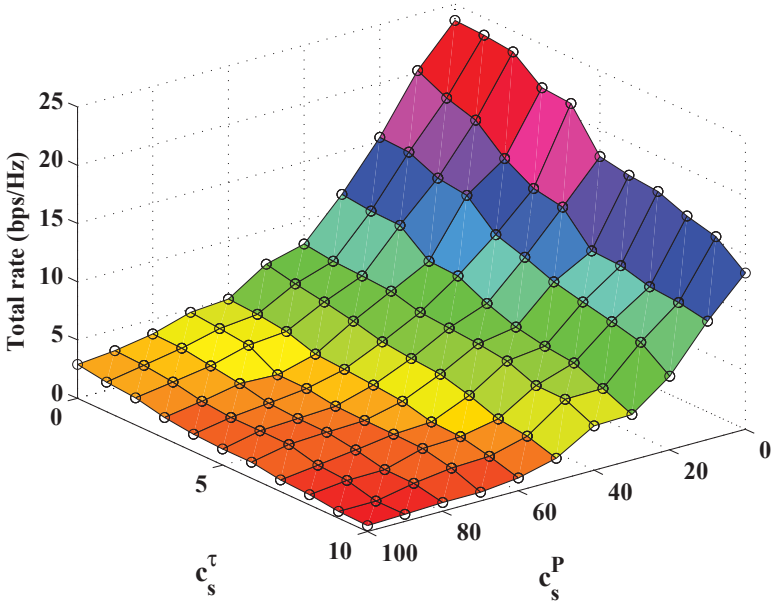


Fig. 4.5 Total rate versus c_s^τ and c_s^P for fixed c_s^M

the pricing factors. Specifically, the values of c_s^τ and c_s^M have significant effects on decreasing the total rate of the VWN.

4.5 Concluding Remarks

In this chapter, we examined the resource allocation problem in a massive MIMO-based VWN. In consideration of possible pilot contamination errors, we formulated an optimization problem to adaptively assign uplink pilot duration, power, antennas and sub-carriers to users in order to maximize the overall system throughput. We developed a low-complexity two-step iterative algorithm wherein the first step finds the optimal power, antenna and sub-carrier allocation to be used in the second step to optimize the uplink pilot duration. Simulation results indicate a significant system performance improvement offered by the proposed scheme using adaptive pilot duration as compared to the scenario with fixed pilot duration.

Chapter 5

User-Association and Resource-Allocation in a C-RAN-based VWN

Although user association and dynamic resource allocation are the prime requirements in a VWN, the practical realization of these algorithms require co-ordination among BSs in order to obtain the knowledge about the channel conditions of users in all slice. To that end, C-RAN has been proposed to facilitate the collaboration among BSs by shifting the baseband processing of all BSs into the cloud. In this chapter, we address the issue of user association and resource allocation in a C-RAN-based VWN. Specifically, we formulate a resource allocation problem that jointly associates users to RRHs/BBUs and allocates the resources (power, antenna) in order to maximize the total sum-rate of the system. We evaluate the performance of the proposed scheme in both the imperfect and perfect CSI scenarios.

5.1 Introduction

In order to practically implement the user association and dynamic resource allocation algorithms in a VWN, the scheduler needs to know the channel state information of all users in all slices. This is possible only through co-ordination among the BSs. Recently, a new concept called C-RAN has been proposed where the signal processing functionalities of the BSs are moved to a pool/cloud of BBUs while RRH units are deployed to provide the radio signals over the mounted antennas. The transmission link between these two separate units is called front-haul link requiring a high capacity media [55].The motivation for this is to create a very low latency

link between the BSs which will enable collaboration among the BSs as well as to increase energy and infrastructure efficiency.

As C-RAN is a recently introduced concept, the research works have focused more on the implementation issues and the challenges in integrating it into the current network. For instance, [7, 56] have addressed the issue of limited capacity front-haul links that poses a severe constraint on the maximum number of served users over the coverage of interest. Specifically, [7] proposes the design of an OFDMA-based C-RAN architecture which allows the RRHs to dynamically associate with the BBUs instead of performing a one-one mapping between RRHs and BBUs. The simulation results performed there demonstrate that the one-one mapping is clearly sub-optimal compared to the dynamic association. Similarly, [57] discusses the various architectural considerations in a front-haul limited C-RAN including partial centralization where the RRH not only implements the RF functionalities but also some layer 2 functions related to baseband processing. Moreover, the challenges in implementing the resource allocation schemes in a C-RAN is discussed and various techniques including Markov decision process (MDP) are proposed. Similarly, [58] proposes a joint power control and fronthaul rate allocation algorithm in a C-RAN network by considering orthogonal sub-carrier allocation to users. Moreover, [59] considers the downlink power minimization problem in a small network with C-RAN. Specifically, the authors use the approach of SCA to propose a resource allocation algorithm that tries to minimize the total transmit power subject to the individual user admission control requirements.

Our proposed setup considers the uplink transmission of users of different slices in the specific region. The coverage will be provided by a set of RRHs with massive MIMO, and the cloud of BBUs which are connected to the RRHs through a limited front-haul capacity link. To provide the isolation between slices, we consider minimum reserved rate and antennas for each slice. The hardware limitation of C-RAN in our setup include maximum transmit power and antenna of each RRH, front-haul capacity limitation and maximum load of each BBU. With the objective to maximize the total throughput of network subject to the isolation and hardware based constraints, we formulate the resource allocation problem which involves joint allocation of BBU, RRH, front-haul, power, and antenna parameters. In this chapter, we propose a modified expression for the throughput over C-RAN to establish the relation between the parameters related to RRHs and BBUs.

The optimization problem is inherently non-convex and NP hard due to intertwined sets of optimization variables called user association parameters (UAP), and a variety of constraints in this setup. To develop an efficient algorithm to solve the problem, we propose a two-step iterative

algorithm where 2 in the first step, joint BBU, front-haul and RRH allocation parameters are assigned. Then, in the second step, the transmit power and the number of allocated antennas to each user are determined by considering all the slice based constraints. The problem for each step is still non-convex and NP-hard. To solve them efficiently, we apply the techniques of CGP and SCA where via different relaxation and transformation techniques, in each step, the problems in each step are transformed into the GP counterparts [37], [34] which can be solved very efficiently by optimization packages like CVX [41]. Via numerical results, we demonstrate the effect of pilot contamination error and the fronthaul limitation on the performance of the VWN.

The rest of this chapter is organized as follows: Section 5.2 presents the system model considered in this problem along with the problem formulation. The proposed iterative algorithm is discussed in Section 5.3 followed by Section 5.4 where we present the numerical results with concluding remarks in Section 5.5.

5.2 System Model

We consider the uplink transmission of a massive MIMO-based VWN with a C-RAN architecture consisting of a set of $\mathcal{S} = \{1, \cdots, S\}$ slices. Each slice $s \in \mathcal{S}$ has a set of $\mathcal{K}_s = \{1, \cdots, K_s\}$ single-antenna users where the QoS of each slice $s \in \mathcal{S}$ is presented by its own minimum required rate R_s^{rsv}. The coverage for all users, $K = \sum_{s \in \mathcal{S}} K_s$, in the VWN, are provided by a set of RRHs, $\mathcal{L} = \{1, \cdots, L\}$ and each RRH is equipped with $M \gg 1$ antennas. Each RRH is connected to the cloud of B BBUs to process the baseband signals for each of the RRH via front-haul link. The system model is illustrated in Fig. 5.1.

Let $f_{k_s,b}$ represent the association between the BBU b and the user k_s such that

$$f_{k_s,b} = \begin{cases} 1, & \text{if user } k \text{ in slice } s \text{ is supported by BBU b,} \\ 0, & \text{otherwise.} \end{cases}$$

We also assume that each BBU b can support at most $o_b^{\max}$ users [6]. This limitation can be mathematically represented as

$$\text{C5.1: } \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} w_{k_s,b} f_{k_s,b} \leq o_b^{\max}, \quad \forall b \in \mathcal{B},$$

where $w_{k_s,b}$ is the load balancing factor for the BBU b to the user k_s and is a system parameter

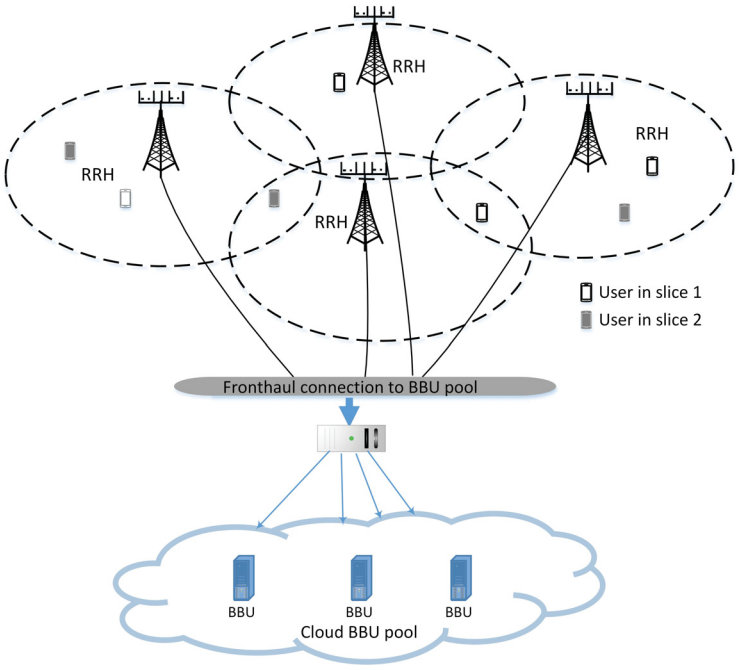


Fig. 5.1 A multi-cell massive MIMO-based VWN with fronthaul limited C-RAN

assigned by the VWN to control the traffic and load of each BBU-user pair. We define F_{k_s} as the summation of the BBU allocation for each user, i.e.

$$\text{C5.2: } F_{k_s} = \sum_{b \in \mathcal{B}} f_{k_s,b}, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}.$$

We denote the user-RRH association factor as the user association factor (UAF) and denote it as α_{l,k_s}, where

$$\alpha_{l,k_s} = \begin{cases} 1, & \text{if user } k_s \text{ in slice } s \text{ is attached to RRH } l, \\ 0, & \text{otherwise.} \end{cases}$$

Since each user can be only attached to at most one RRH at a time, we have the constraint,

$$\text{C5.3: } \sum_{l \in \mathcal{L}} \alpha_{l,k_s} \leq 1, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}. \tag{5.1}$$

To further control the load of C-RAN, we assume that each user is supported by only one

BBU at each transmission instance and each user is assigned to the cloud of BBUs, if and only if, it is assigned to at least one RRH. These two practical implementation concerns can be mathematically represented as

$$\text{C5.4: } F_{k_s} = z_{k_s} \sum_{\forall l \in \mathcal{L}} \alpha_{l,k_s}, \ \forall k_s \in \mathcal{K}_s, \ \ \forall s \in \mathcal{S}.$$

where $z_{k_s} = 1$ is chosen for the RRH load balancing.

Due to front-haul link limitation between the RRHs and the BBU pool [60], we consider the case that each front-haul link between RRH l and BBU b has a maximum capacity of $c_{l,b}^{\max}$, represented as

$$\text{C5.5: } \sum_{s \in \mathcal{S}} \sum_{k_s \in \mathcal{K}_s} f_{k_s,b} \alpha_{l,k_s} \leq c_{l,b}^{\max}, \quad \forall b \in \mathcal{B}, \ \forall l \in \mathcal{L}.$$

In the up-link transmission with massive MIMO based RRHs, we denote $h_{l,k_s,m}$ as the amplitude of the channel coefficient of user k_s and antenna m at RRH l and $\mathbf{h}_{l,k_s} \in \mathbb{C}^{1 \times M_{l,k_s}}$ as the channel vector of user k_s where M_{l,k_s} is the total allocated antennas by RRH l to user k_s. More specifically, $h_{l,k_s,m}$ is

$$h_{l,k_s,m} = \chi_{l,k_s,m} \sqrt{\beta_{l,k_s}}$$

where $\chi_{l,k_s,m}$ represents the fast-fading coefficients from the user k_s to the antenna m of RRH $l \in \mathcal{L}$ and β_{l,k_s} denotes the large-scale fading coefficient of user k_s to RRH $l \in \mathcal{L}$ [18]. Note that β_{l,k_s} includes both path loss and shadowing [18]. We model β_{l,k_s} as $\beta_{l,k_s} = w_{l,k_s} d_{l,k_s}^{-\zeta}$ where w_{l,k_s} is a log-normal random variable representing the shadowing and $d_{l,k_s} \in [0.1, 1]$ is the normalized distance of the user k_s from the BS l with ζ being the path loss exponent. Here, it is assumed that β_{l,k_s} has the same value for all antenna m of RRH l, since the distance between the antennas in a RRH can be assumed to be negligible compared to the distance between the RRH l and user k_s.

Practically, the CSIs are estimated by the RRHs based on the up-link pilots with duration τ at the specific part of the coherence interval of T [17]. In this context, there exist two scenarios in the network for estimating CSI values:

- *Perfect CSI* estimation where it is assumed that the orthogonality of pilot signals in the multi-cell scenario of massive MIMO based networks are preserved. In this scenario the

SINR of user k_s in slice s and in RRH l is

$$\gamma_{l,k_s}^{\text{Perf}} = \beta_{l,k_s} P_{l,k_s} M_{l,k_s}, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \tag{5.2}$$

where p_{l,k_s} is the transmit power of user k_s to RRH l and the noise for all users and RRHs is normalized to 1.

Proof: Without pilot contamination error, the uplink received signal from user $k_s \in \mathcal{K}_s$ on sub-carrier n at RRH l is [17]

$$y_{l,k_s}^{\text{Perf}} = \sqrt{p_{l,k_s}} \mathbf{a}_{l,k_s} \mathbf{h}_{l,k_s} x_{l,k_s} + \sum_{l' \in \mathcal{L}, l' \neq l} \sum_{s \in S} \sqrt{p_{l',k_s}} \mathbf{f}_{l',k_s} \mathbf{h}_{l',k_s} x_{l',k_s} + \mathbf{f}_{l,k_s} \mathbf{n}_{l,k_s},$$

where x_{l,k_s} represents the transmit symbol, $\mathbf{f}_{l,k_s} \in \mathbb{C}^{1 \times M_{l,k_s}}$ is the precoding vector, and $\mathbf{n}_{l,k_s}$ is the corresponding noise coefficient vector. For maximum ratio combining (MRC) detector, we have $\mathbf{f}_{l,k_s} = (\mathbf{h}_{l,k_s})^\dagger$, where $\mathbf{v}\dagger$ is the conjugate transpose operation of vector $\mathbf{v}$. Thus, the received signal vector is [17]

$$y_{l,k_s}^{\text{Perf}} = \sqrt{p_{l,k_s}} (\mathbf{h}_{l,k_s})^\dagger \mathbf{h}_{l,k_s} x_{l,k_s} + \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in S} \sqrt{p_{l',k_s}} (\mathbf{h}_{l,k_s})^\dagger \mathbf{h}_{l',k_s} x_{l',k_s} + \mathbf{h}_{l,k_s}^\dagger \mathbf{n}_{l,k_s},$$

When $M_{l,k_s} \gg 1$, from the law of large numbers, the term $\sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in S}$ $\sqrt{p_{l',k_s}} (\mathbf{h}_{l,k_s})^\dagger \mathbf{h}_{l',k_s} x_{l',k_s}$ in (5.3) becomes zero [17]. By considering $p_{l,k_s} = E_{l,k_s}/M_{l,k_s}$, and assuming $\chi_{l,k_s,m}, \forall l, \forall k_s, \forall m$, to have zero mean and unit variance, (5.3) can be transformed into

$$\frac{1}{\sqrt{M_{l,k_s}}} y_{l,k_s}^{\text{Perf}} = \sqrt{E}_{l,k_s} \beta_{l,k_s} x_{l,k_s} + \sqrt{\beta}_{l,k_s} \tilde{n}_{l,k_s}, \tag{5.3}$$

where $\tilde{n}_{l,k_s}$ represents the AWGN with power 1. Hence, for $M_{l,k_s} \gg 1$, the SINR from user k_s in slice s to RRH l converges to (5.2).

□

- *Imperfect CSI* estimation where there exists the pilot contamination error and the achieved

SINR of user k_s in slice s to RRH $l \in \mathcal{L}$ is

$$\gamma_{l,k_s}^{\text{Imperf}} = \frac{\tau \beta_{l,k_s}^2 P_{l,k_s}^2 M_{l,k_s}}{\tau \sum_{\forall s \in \mathcal{S}} \sum_{\forall l' \in \mathcal{L}} \sum_{\forall k'_s \in \mathcal{K}_s} \beta_{l',k'_s}^2 P_{l',k'_s}^2 M_{l',k'_s} + 1}, \tag{5.4}$$

where the noise power for all users and RRHs is assumed to be to 1 for simplicity.

Proof: Let $\mathbf{\Phi}$ represent the $\tau \times K$ orthonormal pilot sequence matrix used by the users where $\mathbf{\Phi}^H \mathbf{\Phi} = \mathbf{I}_K$ and $\mathbf{I}_K$ is the unit matrix. With pilot contamination, the estimated CSIs are [17]

$$\tilde{h}_{l,k_s,m} = \left(h_{l,k_s,m} + \frac{1}{\sqrt{p_{l,k_s}^{\text{Pilot}}}} w_l \right) \tilde{\beta}_{l,k_s},$$

where $\tilde{\beta}_{l,k_s} = \beta_{l,k_s} (\sum_{l \in \mathcal{L}} \sum_{k_s \in \mathcal{K}_s} \sum_{s \in \mathcal{S}} \beta_{l,k_s} + \frac{1}{p_{l,k_s}^{\text{Pilot}}})^{-1}$ and p_{l,k_s}^{Pilot} is the transmit power for the pilot sequence for the user k_s and w_l represents the contaminated pilot sequence received at the RRH l. Now, the received signal after using MRC at the RRH l from user k_s is

$$y_{l,k_s}^{\text{Imperf}} = \mathbf{f}_{l,k_s} (\sqrt{p_{l,k_s}} \mathbf{h}_{l,k_s} x_{l,k_s} + \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in S} \sum_{\forall k'_s \in \mathcal{K}_s, k'_s \neq k_s} \sqrt{p_{l',k'_s}} \mathbf{h}_{l',k'_s} x_{l',k'_s} + \mathbf{n}_{l,k_s})$$

Now, for MRC, the precoding vector is given by $\mathbf{f}_{l,k_s} = \tilde{\mathbf{h}}_{l,k_s}^{\dagger}$, where $\tilde{\mathbf{h}}_{l,k_s}^{\dagger}$ is the estimate channel vector taking into account the pilot contamination error. Hence,

$$y_{l,k_s}^{\text{Imperf}} = \tilde{\mathbf{h}}_{l,k_s}^{\dagger} (\sqrt{p_{l,k_s}} \mathbf{h}_{l,k_s} x_{l,k_s} + \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in S} \sum_{\forall k'_s \in \mathcal{K}_s, k'_s \neq k_s} \sqrt{p_{l',k'_s}} \mathbf{h}_{l',k'_s} x_{l',k'_s} + \mathbf{n}_{l,k_s})$$

Substituting $p_{l,k_s} = E_{l,k_s} / \sqrt{M}_{l,k_s}$ and $p_{l,k_s}^{\text{Pilot}} = \tau p_{l,k_s}$ in the above expression, we get,

$$y_{l,k_s}^{\text{Imperf}} = \tilde{\mathbf{h}}_{l,k_s}^{\dagger} (\frac{\sqrt{E_{l,k_s}}}{M_{l,k_s}^{3/4}} \mathbf{h}_{l,k_s} x_{l,k_s} + \sum_{\forall l' \in \mathcal{L}, l' \neq l} \sum_{\forall s \in S} \sum_{\forall k'_s \in \mathcal{K}_s, k'_s \neq k_s} \frac{\sqrt{E_{l',k'_s}}}{M_{l',k'_s}^{3/4}} \mathbf{h}_{l',k'_s} x_{l',k'_s} + \mathbf{n}_{l,k_s}) \tag{5.5}$$

By substituting the variance of $\mathbf{h}_{l,k_s}$ [17], and by using the law of large numbers for large M_{l,k_s}, we get the simplified expression for SINR of the user k_s as given by (5.4).

□

Based on the expression for $\gamma_{l,k_s}^{\text{Perf}}$ and $\gamma_{l,k_s}^{\text{Imperf}}$, the rate of each user k_s at the RRH l is

$$R_{l,k_s} = \begin{cases} F_{k_s}\alpha_{l,k_s}\log_2(1+\gamma_{l,k_s}^{\text{Perf}}), & \text{for perfect CSI,} \\ \frac{T-\tau}{T}F_{k_s}\alpha_{l,k_s}\log_2(1+\gamma_{l,k_s}^{\text{Imperf}}), & \text{for imperfect CSI.} \end{cases} \tag{5.6}$$

To provide the isolation between different slices, in the proposed VWN setup, both rate and resource reservation strategy will be developed here where the minimum required rate, i.e., R_s^{rsv}, and the minimum number of antennas for specific region, i.e., M_s^{rsv}, for each slice s are preserved, respectively, [1]. These two isolation based constraints in VWN can be written as

$$\text{C5.6: } \sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}\alpha_{l,k_s}R_{l,k_s} \geq R_s^{\text{rsv}}, \quad \forall s\in\mathcal{S},$$
$$\text{C5.7: } \sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} M_{l,k_s} \geq M_s^{\text{rsv}}, \quad \forall s\in\mathcal{S}.$$

From the hardware limitation, each user's transmit power and RRH's antennas are limited as

$$\text{C5.8: } P_{l,k_s} \leq P_{k_s}^{\max}, \quad \forall k_s\in\mathcal{K}_s, \forall s\in\mathcal{S}, \forall l\in\mathcal{L},$$
$$\text{C5.9: } \sum_{s\in\mathcal{S}}\sum_{k_s\in\mathcal{K}_s} M_{l,k_s} \leq M_l^{\max}, \quad \forall l\in\mathcal{L},$$

where $P_{k_s}^{\max}$ and $M_l^{\max}$ are the maximum transmit power of user k_s and maximum number of antennas mounted on the RRH l, respectively, where we assume $\sum_{s\in\mathcal{S}} M_s^{\text{rsv}} \leq M_l^{\max}$ to eliminate the redundancy between C5.7 and C5.9 and feasibility issue of optimization problem.

With the objective of maximizing the total throughput of VWN, subject to all above implementation constraints, the resource allocation problem of this setup can be written as

$$\max_{\boldsymbol{\alpha},\mathbf{F},\mathbf{P},\mathbf{M}} \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}\alpha_{l,k_s}R_{l,k_s}(\mathbf{P},\mathbf{M}), \tag{5.7}$$
$$\text{subject to: C5.1 - C5.9,}$$

where $\boldsymbol{\alpha}$, $\mathbf{F}$, $\mathbf{P}$, and $\mathbf{M}$ are the vectors of all α_{l,k_s}, $f_{k_s,b}$, P_{l,k_s}, and M_{l,k_s}, respectively, for all $k_s \in \mathcal{K}_s$, $s \in \mathcal{S}$, and $l \in \mathcal{L}$. (5.7) suffers from a high computational complexity due to its highly non-convex and combinatorial structure. To overcome this issue, we first relax the integer

variables to be continuous. Then, we solve the problem with a two-stem iterative algorithm by resorting to SCA and CGP to transform the non-convex optimization problem into the GP-based approximation.

5.3 Proposed Two-level Iterative Algorithm

In order to solve the above problem, we propose a two-level iterative algorithm which involves two major steps:

- *Association algorithm (Step 1)* where UAF parameters, i.e., $\boldsymbol{\alpha}$, and BBUs association factors, i.e., $\mathbf{F}$, are derived by assuming the fixed values of the power and antenna vectors, $\mathbf{P}$, and $\mathbf{M}$,
- *RRH adjusting algorithm (Step 2)* where based on the optimal values of $\boldsymbol{\alpha}^*$ and $\mathbf{F}^*$ from the previous step, the RRH parameters, i.e., $\mathbf{P}$, and $\mathbf{M}$ are derived.

The whole process is demonstrated as

$$\underbrace{\boldsymbol{\alpha}(0), \mathbf{F}(0) \rightarrow \mathbf{P}(0), \mathbf{M}(0)}_{\text{Initialization}} \rightarrow \ldots \underbrace{\boldsymbol{\alpha}(t)^*, \mathbf{F}(t)^* \rightarrow \mathbf{P}(t)^*, \mathbf{M}(t)^*}_{\text{Iteration } t} \rightarrow \underbrace{\boldsymbol{\alpha}^*, \mathbf{F}^* \rightarrow \mathbf{P}^*, \mathbf{M}^*}_{\text{Optimal solution}},$$

where $t > 0$ is the iteration number. Also, $\mathbf{P}(t)^*$ and $\mathbf{M}(t)^*$ are optimal values at the iteration t from convex transformation of the related optimization problems in each step. The iterative procedure is stopped when

$$||\mathbf{P}^*(t) - \mathbf{P}^*(t-1)|| \leq \varepsilon_1 \text{ and } ||\mathbf{M}^*(t) - \mathbf{M}^*(t-1)|| \leq \varepsilon_2$$

where $0 < \varepsilon_1 \ll 1$ and $0 < \varepsilon_2 \ll 1$.

Notably, both problems in each step for finding optimal values are still non-convex optimization problems and suffer from high computational complexity. To solve them efficiently, by applying CGP along with various transformations and convexification approaches, the sequence of lower bound geometric programming based approximation is derived. The sub-algorithms are described in detail in the following sub-sections.

Algorithm 5.1 Iterative Joint User-Association and Resource-Allocation Algorithm

Initialization: Set $t := 1$ and initialize all power of each user by $P_{k_s}^{\max}$, and all antennas by $[M_l^{\max}/K]$.

Repeat

Step 1: Derive $\boldsymbol{\alpha}^*(t)$ and $\mathbf{F}^*(t)$ from (5.9) by considering fixed values of $\mathbf{P}^*(t-1)$ and $\mathbf{M}^*(t-1)$;

Step 2: For fixed values of $\boldsymbol{\alpha}^*(t)$ and $\mathbf{F}^*(t)$, find $\mathbf{P}^*(t)$ and $\mathbf{M}^*(t)$ from (5.24) or (5.26);

Step 3: **Stop** if $||\boldsymbol{\alpha}^*(t) - \boldsymbol{\alpha}^*(t-1)|| \leq \varepsilon_1$, and $||\mathbf{P}^*(t) - \mathbf{P}^*(t-1)|| \leq \varepsilon_2$. Otherwise, set $t := t+1$ and go to **Step 1**.

5.3.1 Association Algorithm

For fixed values of $\mathbf{P}(t-1)$ and $\mathbf{M}(t-1)$, at iteration t, the resource allocation problem is simplified into

$$\max_{\boldsymbol{\alpha},\mathbf{F}} \sum_{s\in\mathcal{S}} \sum_{\forall l\in\mathcal{L}} \sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t)\alpha_{l,k_s}(t)R_{l,k_s}(\mathbf{P}(t-1),\mathbf{M}(t-1)), \tag{5.8}$$

subject to: C5.1 $-$ C5.6,

where the only optimization vectors are $\boldsymbol{\alpha}$ and $\mathbf{F}$. Notably, (5.8) has a less computational complexity compared to (5.7) while it is still non-convex and NP hard problem. To reduce the computational complexity, we first relax α_{l,k_s} from integer variable to the continuous variable as $\alpha_{l,k_s} = [0,1]$. To convert the resulting problem, we consider t_1 as the index of iteration for sub-algorithm 1.

Proposition 1: In each iteration, the GP based approximation of (5.8) is

$$\min_{\boldsymbol{\alpha}(t_1),\mathbf{F}(t_1),s_{l,k_s}(t_1),x_0(t_1)} x_0(t_1) \tag{5.9}$$

subject to:

$$x_1\big[\frac{x_0(t_1)}{c_0(t_1)}\big]^{-c_0(t_1)} \times \prod_{l,k_s}\big[\frac{F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t)}{c_{l,k_s}(t_1)}\big]^{-c_{l,k_s}(t_1)} \leq 1$$

$$\text{C5.1: } \sum_{\forall s\in\mathcal{S}} \sum_{k_s=\mathcal{K}_s} f_{k_s,b}(t_1) \leq o_b^{\max}, \quad \forall b\in\mathcal{B},$$

$$\widetilde{\text{C}}5.2: F_{k_s}(t_1) \times \prod_{b\in\mathcal{B}}\big[\frac{w_{k_s,b}f_{k_s,b}(t_1)}{d_{k_s,b}(t_1)}\big]^{-d_{k_s,b}(t_1)} = 1, \quad \forall k_s\in\mathcal{K}_s, \forall s\in\mathcal{S},$$

$$\text{C5.3:} \sum_{\forall l} \alpha_{l,k_s} \leq 1, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S},$$

$$\widetilde{\text{C}}5.4 : F_{k_s}(t_1) \times \frac{1}{z_{k_s}} \prod_{\forall l, \forall n} \left[\frac{\alpha_{l,k_s}(t_1)}{e_{l,k_s}(t_1)} \right]^{-e_{l,k_s}(t_1)} = 1, \qquad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S},$$

$$\text{C5.5:} \sum_{s \in \mathcal{S}} \sum_{\forall k_s \in \mathcal{K}_s} f_{k_s,b}(t_1) \alpha_{l,k_s}(t_1) \leq c_{l,b}^{\max}, \quad \forall b \in \mathcal{B},$$

$$\widetilde{\text{C}}5.6 : R_s^{\text{rsv}} \prod_{l,k_s} \left[\frac{F_{k_s}(t_1) \alpha_{l,k_s}(t_1) R_{l,k_s}(t)}{\varphi_{l,k_s}(t_1)} \right]^{-\varphi_{l,k_s}(t_1)} \leq 1, \qquad \forall s \in \mathcal{S}.$$

where,

$$c_0(t_1) = \frac{x_0(t_1 - 1)}{x_0(t_1 - 1) + \sum_{s \in \mathcal{S}} \sum_{\forall l \in \mathcal{L}} \sum_{\forall k_s \in \mathcal{K}_s} F_{k_s}(t_1 - 1) \alpha_{l,k_s}(t_1 - 1) R_{l,k_s}(t)}, \tag{5.10}$$

$$c_{l,k_s}(t_1) = \frac{F_{k_s}(t_1 - 1) \alpha_{l,k_s}(t_1 - 1) R_{l,k_s}(t)}{x_0(t_1 - 1) + \sum_{s \in \mathcal{S}} \sum_{\forall l \in \mathcal{L}} \sum_{\forall k_s \in \mathcal{K}_s} F_{k_s}(t_1 - 1) \alpha_{l,k_s}(t_1 - 1) R_{l,k_s}(t)}, \tag{5.11}$$

$$d_{k_s,b}(t_1) = \frac{w_{k_s,b} f_{k_s,b}(t_1 - 1)}{\sum_{\forall b \in \mathcal{B}} w_{k_s,b} f_{k_s,b}(t_1 - 1)}, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \tag{5.12}$$

$$\varphi_{l,k_s}(t_1) = \frac{F_{k_s}(t_1 - 1) \alpha_{l,k_s}(t_1 - 1) R_{l,k_s}(t)}{\sum_{\forall l \in \mathcal{L}} \sum_{\forall k_s \in \mathcal{K}_s} F_{k_s}(t_1 - 1) \alpha_{l,k_s}(t_1 - 1) R_{l,k_s}(t)}, \tag{5.13}$$

$$e_{l,k_s}(t_1) = \frac{\alpha_{l,k_s}(t_1 - 1)}{\sum_{\forall l \in \mathcal{L}} \alpha_{l,k_s}(t_1 - 1)}, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \forall l \in \mathcal{L}. \tag{5.14}$$

Proof: The objective function in (5.7) can be expressed as:

$$\min_{\boldsymbol{\alpha}, \mathbf{F}} - \sum_{s \in \mathcal{S}} \sum_{\forall l \in \mathcal{L}} \sum_{\forall k_s \in \mathcal{K}_s} F_{k_s}(t_1) \alpha_{l,k_s}(t_1) R_{l,k_s}(t) \tag{5.15}$$

Now, similar to (A.1.11), we apply $x_1 \gg 1$ as

$$x_1 - \sum_{s \in \mathcal{S}} \sum_{\forall l \in \mathcal{L}} \sum_{\forall k_s \in \mathcal{K}_s} F_{k_s}(t_1) \alpha_{l,k_s}(t_1) R_{l,k_s}(t)$$

to satisfy the positive condition of objective function of GP. By considering t_1, we use $x_0(t_1) > 0$

to transform (5.15) into

$$\min_{\boldsymbol{\alpha}(t_1),\mathbf{F}(t_1),x_0(t_1)} x_0(t_1) \tag{5.16}$$

$$x_1 - \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t) \leq x_0(t_1). \tag{5.17}$$

Now (5.16) can be rewritten as

$$\frac{x_1}{x_0(t_1) + \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t)} \leq 1,$$

and by using AGMA, we get

$$x_1\big[\frac{x_0(t_1)}{c_0(t_1)}\big]^{-c_0(t_1)} \prod_{l,k_s}\big[\frac{F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t)}{c_{l,k_s}(t_1)}\big]^{-c_{l,k_s}(t_1)} \leq 1, \tag{5.18}$$

where,

$$c_0(t_1) = \frac{x_0(t_1-1)}{x_0(t_1-1) + \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1-1)\alpha_{l,k_s}(t_1-1)R_{l,k_s}(t)}, \tag{5.19}$$

$$c_{l,k_s}(t_1) = \frac{F_{k_s}(t_1-1)\alpha_{l,k_s}(t_1-1)R_{l,k_s}(t)}{x_0(t_1-1) + \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1-1)\alpha_{l,k_s}(t_1-1)R_{l,k_s}(t)}. \tag{5.20}$$

By applying AGMA for C5.2 to reach monomial function, we get

$$\widetilde{\text{C}}5.2 : F_{k_s}(t_1) \times \prod_{b\in\mathcal{B}}\big[\frac{w_{k_s,b}f_{k_s,b}(t_1)}{d_{k_s,b}(t_1)}\big]^{-d_{k_s,b}(t_1)} = 1, \quad \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \tag{5.21}$$

where, $d_{k_s,b}(t_1) = \frac{w_{k_s,b}f_{k_s,b}(t_1-1)}{\sum_{\forall b\in\mathcal{B}} w_{k_s,b}f_{k_s,b}(t_1-1)}, \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}$. Similarly, C5.6 can be expressed as

$$\frac{R_s^{\text{rsv}}}{\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t)} \leq 1, \quad \forall s \in \mathcal{S}, \tag{5.22}$$

wihch can be approximated as

$$\widetilde{\mathrm{C}}5.6: R_s^{\mathrm{rsv}} \prod_{l,k_s} \Big[\frac{F_{k_s}(t_1)\alpha_{l,k_s}(t_1)R_{l,k_s}(t)}{\varphi_{l,k_s}(t_1)}\Big]^{-\varphi_{l,k_s}(t_1)} \leq 1,$$

where

$$\varphi_{l,k_s}(t_1) = \frac{F_{k_s}(t_1-1)\alpha_{l,k_s}(t_1-1)R_{l,k_s}(t)}{\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t_1-1)\alpha_{l,k_s}(t_1-1)R_{l,k_s}(t)}.$$

Also, for the constraint C5.4, we can approximate the equality constraint as follows:

$$\widetilde{\mathrm{C}}5.4: F_{k_s}(t_1) \times \frac{1}{z_{l,k_s}} \prod_{\forall l,\forall n} \left[\frac{\alpha_{l,k_s}(t_1)}{e_{l,k_s}(t_1)}\right]^{-e_{l,k_s}(t_1)} = 1,\ \forall k_s \in \mathcal{K}_s,\ \ \forall s \in \mathcal{S},$$

where, $e_{l,k_s}(t_1) = \frac{\alpha_{l,k_s}(t_1-1)}{\sum_{\forall l\in\mathcal{L}}\alpha_{l,k_s}(t_1-1)}$. So, the overall problem for sub-algorithm 1 can be written as (5.9). □

Consequently, (5.9) can be solved very efficiently via CVX at each iteration t_1. The iterative algorithm is initiated with an arbitrary initial value for $s_{l,k_s}(t_1)$. At each iteration, $\boldsymbol{\alpha}(t_1)$, $\mathbf{F}(t_1)$, $s_{l,k_s}(t_1)$, $x_0(t_1)$, are derived and then, (5.10)-(5.14) are updated. The iterative algorithm will be stopped if $||\boldsymbol{\alpha}^*(t_1) - \boldsymbol{\alpha}^*(t_1-1)|| \leq \varepsilon_1$, and $||\mathbf{F}^*(t_1) - \mathbf{F}^*(t_1-1)|| \leq \varepsilon_2$ where $0 < \varepsilon_1 \ll 1$ and $0 < \varepsilon_2 \ll 1$.

5.3.2 RRH Adjusting Algorithm

For fixed value of $\boldsymbol{\alpha}$ and $\mathbf{F}$ obtained from sub-algorithm 1, the resource allocation problem is simplified into

$$\max_{\mathbf{P},\mathbf{M}} \sum_{s\in\mathcal{S}}\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s} F_{k_s}(t)\alpha_{l,k_s}(t)R_{l,k_s}(\mathbf{P},\mathbf{M}), \tag{5.23}$$
$$\text{subject to: C5.6} - \text{C5.9}.$$

Similar to (5.8), (5.23) has a less computational complexity compared to (5.7) due to the fact that it just involves $\mathbf{P}$ and $\mathbf{M}$. However, it is again non-convex optimization problem and NP hard. To reduce the computational complexity, we first relax M_{l,k_s} from integer variable to the continuous variable between $[0,\ M_l^{\max}]$, then we apply CGP framework to convert (5.23) into its GP-based

approximation as follows in the Proposition 2 where t_2 is the index of iteration for sub-algorithm 2.

Proposition 2: GP approximation of (5.23) is

- Perfect CSI

$$\min_{\mathbf{P,M}} \prod_{l,k_s,s} \left[\left[\frac{1}{\omega_{l,k_s}(t_2)}\right]^{-\omega_{l,k_s}(t_2)} \times \left[\frac{\beta_{l,k_s} P_{l,k_s}(t_2) M_{l,k_s}(t_2)}{j_{l,k_s}(t_2)}\right]^{-j_{l,k_s}(t_2)} \right], \tag{5.24}$$

subject to: C5.8, C5.9

$$\overline{\text{C}}5.6: \prod_{l,k_s} \left[\left[\frac{1}{\omega_{l,k_s}(t_2)}\right]^{-\omega_{l,k_s}(t_2)} \times \left[\frac{\beta_{l,k_s} P_{l,k_s}(t_2) M_{l,k_s}(t_2)}{j_{l,k_s}(t_2)}\right]^{-j_{l,k_s}(t_2)} \right]^{F_{k_s}(t)} \leq 2^{-R_s^{\text{rsv}}}, \quad \forall s \in \mathcal{S}.$$

$$\overline{\text{C}}5.7: M_s^{\text{rsv}} \prod_{\forall k_s, l} \left[\frac{M_{l,k_s}(t_2)}{\lambda_{l,k_s}(t_2)}\right]^{-\lambda_{l,k_s}} \leq 1, \quad \forall s \in \mathcal{S}.$$

where

$$\begin{aligned} \omega_{l,k_s}(t_2) &= \frac{1}{1+\beta_{l,k_s} P_{l,k_s}(t_2-1) M_{l,k_s}(t_2-1)}, \\ j_{l,k_s}(t_2) &= \frac{\beta_{l,k_s} P_{l,k_s}(t_2-1) M_{l,k_s}(t_2-1)}{1+\beta_{l,k_s} P_{l,k_s}(t_2-1) M_{l,k_s}(t_2-1)}, \\ \lambda_{l,k_s}(t_2) &= \frac{M_{l,k_s}(t_2-1)}{\sum_{ks\in\mathcal{K}_s}\sum_{l\in\mathcal{L}} M_{l,k_s}(t_2-1)}. \end{aligned} \tag{5.25}$$

- Imperfect CSI

$$\begin{aligned} \min_{\mathbf{P,M}} \prod_{l,k_s,s} \Bigg[&(1+\tau \sum_{\forall l'\neq l} \sum_{\forall k'_s \neq k_s} \sum_{\forall s\in\mathcal{S}} \beta^2_{l',k'_s} P^2_{l',k'_s}(t_2) M_{l',k'_s}(t_2)) \times \left[\frac{1}{g_1(t_2)}\right]^{-g_1(t_2)} \times \\ &\prod_{l,k_s} \left[\frac{\tau \beta^2_{l,k_s} P^2_{l,k_s}(t_2) M_{l,k_s}(t_2)}{g_{l,k_s}(t_2)}\right]^{-g_{l,k_s}(t_2)} \Bigg] \end{aligned} \tag{5.26}$$

subject to: C5.8, C5.9

$$\hat{\text{C}}5.6: \prod_{l,k_s} \Bigg[(1+\tau \sum_{\forall l'\neq l} \sum_{\forall k'_s \neq k_s} \sum_{\forall s\in\mathcal{S}} \beta^2_{l',k'_s} P^2_{l',k'_s}(t_2) M_{l',k'_s}(t_2)) \times \left[\frac{1}{g_1(t_2)}\right]^{-g_1(t_2)} \times$$

$$\prod_{l,k_s}\Big[\frac{\tau\beta_{l,k_s}^2P_{l,k_s}^2(t_2)M_{l,k_s}(t_2)}{g_{l,k_s}(t_2)}\Big]^{-g_{l,k_s}(t_2)}\Bigg]^{F_{k_s}(t)}\leq 2^{-R_s^{\text{rsv}}},\quad \forall s\in\mathcal{S},$$

$$\hat{\text{C}}5.7: M_s^{\text{rsv}}\prod_{\forall k_s,l}\Big[\frac{M_{l,k_s}(t_2)}{\lambda_{l,k_s}(t_2)}\Big]^{-\lambda_{l,k_s}}\leq 1,\quad \forall s\in\mathcal{S}.$$

where

$$g_1(t_2)=\frac{1}{1+\tau\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s}\sum_{\forall s\in\mathcal{S}}\beta_{l,k_s}^2P_{l,k_s}^2(t_2-1)M_{l,k_s}(t_2-1)},\tag{5.27}$$

$$g_{l,k_s}(t_2)=\frac{\tau\beta_{l,k_s}^2P_{l,k_s}^2(t_2-1)M_{l,k_s}(t_2-1)}{1+\tau\sum_{\forall l\in\mathcal{L}}\sum_{\forall k_s\in\mathcal{K}_s}\sum_{\forall s\in\mathcal{S}}\beta_{l,k_s}^2P_{l,k_s}^2(t_2-1)M_{l,k_s}(t_2-1)}.\tag{5.28}$$

Proof: For the perfect scenario, the objective function of (5.23) is

$$\min_{\mathbf{P,M}}\prod_{l,k_s,s}\left(\frac{1}{1+\beta_{l,k_s}P_{l,k_s}M_{l,k_s}}\right),\tag{5.29}$$

which can be expressed as

$$\Big[\frac{1}{\omega_{l,k_s}(t_2)}\Big]^{\omega_{l,k_s}(t_2)}\times\Big[\frac{\beta_{l,k_s}P_{l,k_s}(t_2)M_{l,k_s}(t_2)}{j_{l,k_s}(t_2)}\Big]^{j_{l,k_s}(t_2)},\tag{5.30}$$

where,

$$\omega_{l,k_s}(t_2)=\frac{1}{1+\beta_{l,k_s}P_{l,k_s}(t_2-1)M_{l,k_s}(t_2-1)},\text{ and}\tag{5.31}$$

$$j_{l,k_s}(t_2)=\frac{\beta_{l,k_s}P_{l,k_s}(t_2-1)M_{l,k_s}(t_2-1)}{1+\beta_{l,k_s}P_{l,k_s}(t_2-1)M_{l,k_s}(t_2-1)}.$$

Now, the objective function can be written as

$$\prod_{l,k_ss,n}\left[\Big[\frac{1}{\omega_{l,k_s}(t_2)}\Big]^{-\omega_{l,k_s}(t_2)}\times\Big[\frac{\beta_{l,k_s}P_{l,k_s}(t_2)M_{l,k_s}(t_2)}{j_{l,k_s}(t_2)}\Big]^{-j_{l,k_s}(t_2)}\right].\tag{5.32}$$

For C5.6, the AGMA approximation is

$$\overline{\text{C}}5.6: \prod_{l,k_s,} \left[\left[\frac{1}{\omega_{l,k_s}(t_2)}\right]^{-\omega_{l,k_s}(t_2)} \times \left[\frac{\beta_{l,k_s} P_{l,k_s}(t_2) M_{l,k_s}(t_2)}{j_{l,k_s}(t_2)}\right]^{-j_{l,k_s}(t_2)} \right]^{F_{k_s}(t)} \leq 2^{-R_s^{\text{rsv}}}, \quad \forall s \in \mathcal{S}.$$

Similarly for the constraint C5.7, we can approximate the expression into monomials as follows:

$$\frac{M_s^{\text{rsv}}}{\sum_{ks\in\mathcal{K}_s} \sum_{l\in\mathcal{L}} M_{l,k_s}(t_2)} \leq 1 \tag{5.33}$$

$$\overline{\text{C}}5.7: M_s^{\text{rsv}} \prod_{\forall k_s,n,l} \left[\frac{M_{l,k_s}(t_2)}{\lambda_{l,k_s}(t_2)}\right]^{-\lambda_{l,k_s}} \leq 1, \quad \forall s \in \mathcal{S}.$$

where, $\lambda_{l,k_s}(t_2) = \frac{M_{l,k_s}(t_2-1)}{\sum_{ks\in\mathcal{K}_s} \sum_{l\in\mathcal{L}} M_{l,k_s}(t_2-1)}$. Hence, the overall problem for the perfect CSI estimation scenario for sub-algorithm 2 can be written as (5.24)

The rate of user k_s in RRH l in the imperfect CSI scenario can be rewritten as

$$R_{l,k_s} = \log_2 \left(\frac{1 + \tau \sum_{\forall l\in\mathcal{L}} \sum_{\forall k_s\in\mathcal{K}_s} \sum_{\forall s\in\mathcal{S}} \beta_{l,k_s}^2 P_{l,k_s}^2 M_{l,k_s}}{1 + \tau \sum_{\forall l'\neq l} \sum_{\forall k_s'\neq k_s} \sum_{\forall s\in\mathcal{S}} \beta_{l',k_s'}^2 P_{l',k_s'}^2 M_{l',k_s'}} \right). \tag{5.34}$$

Hence, the objective function of (5.23) can be rewritten as

$$\min_{\mathbf{P,M}} \prod_{l,k_s,s} \left(\frac{1 + \tau \sum_{\forall l'\neq l} \sum_{\forall k_s'\neq k_s} \sum_{\forall s\in\mathcal{S}} \beta_{l',k_s'}^2 P_{l',k_s'}^2 M_{l',k_s'}}{1 + \tau \sum_{\forall l\in\mathcal{L}} \sum_{\forall k_s\in\mathcal{K}_s} \sum_{\forall s\in\mathcal{S}} \beta_{l,k_s}^2 P_{l,k_s}^2 M_{l,k_s}} \right), \tag{5.35}$$

Now, by considering t_2 as the iteration index, the denominator (5.35) can be transformed by AGMA as

$$\left[\frac{1}{g_1(t_2)}\right]^{g_1(t_2)} \times \prod_{l,k_s,s} \left[\frac{\tau \beta_{l,k_s}^2 P_{l,k_s}^2(t_2) M_{l,k_s}(t_2)}{g_{l,k_s}(t_2)}\right]^{g_{l,k_s}(t_2)}, \tag{5.36}$$

where $g_1(t_2)$ and $g_{l,k_s}(t_2)$ are introduced in (5.27) and (5.28). Hence, the GP approximation of objective function is

$$\min_{\mathbf{P,M}} \prod_{l,k_s,s} \Bigg[(1 + \tau \sum_{\forall l'\neq l} \sum_{\forall k_s'\neq k_s} \sum_{\forall s\in\mathcal{S}} \beta_{l',k_s'}^2 P_{l',k_s'}^2(t_2) M_{l',k_s'}(t_2)) \times \left[\frac{1}{g_1(t_2)}\right]^{-g_1(t_2)} \times \tag{5.37}$$

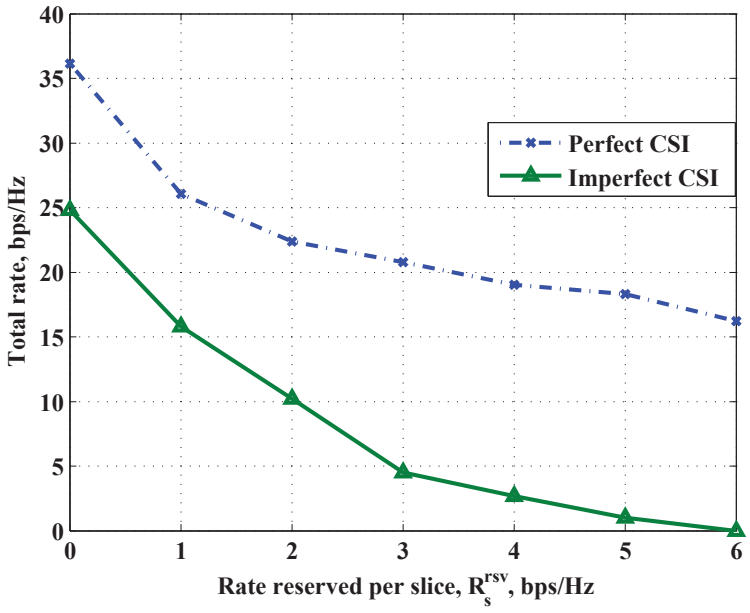


Fig. 5.2 Total rate versus R_s^{rsv}

$$\prod_{l,k_s,s} \Big[\frac{\tau\beta_{l,k_s}^2 P_{l,k_s}^2(t_2) M_{l,k_s}(t_2)}{g_{l,k_s}(t_2)}\Big]^{-g_{l,k_s}(t_2)}\Big].$$

Consequently, (5.23) is approximated to its GP format in (5.26).

□

At each iteration, problem (5.24) or (5.26) can be solved via CVX. The iterative algorithm will be stopped if $||\mathbf{P}(t_2) - \mathbf{P}(t_2-1)|| \ll \varepsilon_3$, $||\mathbf{M}(t_2) - \mathbf{M}(t_2-1)|| \ll \varepsilon_4$, where $0 < \varepsilon_3 \ll 1$ and $0 < \varepsilon_4 \ll 1$.

5.4 Numerical Results and Discussions

5.4.1 System Parameters

We consider a multi-cell VWN with $L = 4$ RRHs connected to $B = 3$ BBUs in a 2×2 square area serving users in $S = 2$ slices. The RRHs are located at coordinates $(0.5, 0.5)$, $(0.5, 1.5)$, $(1.5, 0.5)$, and $(1.5, 1.5)$ and the users are generated randomly based on the uniform distribution within the area of interest. The path loss exponent for modeling the channel coefficients is taken to be

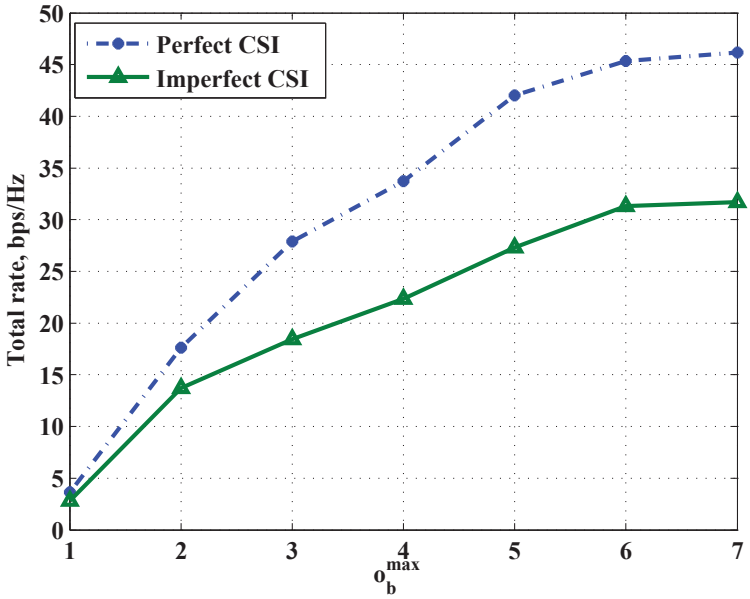


Fig. 5.3 Total rate versus $o_b^{\max}$

$\zeta = 3$ [17]. We set $K = 8$ and $P_{k_s}^{\max} = 0$ dB, $\forall k \in \mathcal{K}$ unless otherwise stated. For all of the simulations, we set $x_1 = 10^7$, $\varepsilon_1 = \varepsilon_3 = 10^{-5}$, $\varepsilon_2 = \varepsilon_4 = 10^{-6}$ and $M_l^{\max} = 150$, for all l. We assume each front-haul link has a maximum capacity of $c_{l,b}^{\max} = 10$ baseband signals. In all of the simulations, when there is no feasible solution for the system, i.e. any of the constraints given by (5.7) does not hold, the total rate is set to be zero.

5.4.2 Performance Analysis

In order to evaluate the performance of the algorithm for the perfect and imperfect CSI scenarios, in Fig. 5.2, the total rate versus R_s^{rsv} is shown. From Fig. 5.2, the total rate decreases with increasing R_s^{rsv} due to the reduction in the feasibility region by increasing R_s^{rsv} as demonstrated in [34]. Moreover, the total rate obtained is more in the case of perfect CSI as compared to the case of imperfect CSI due to the interference from pilot contamination from users in neighboring RRHs in the case of imperfect CSI.

To study the effect of the BBU capacity on the system performance, $o_b^{\max}$, in Fig. 5.3, the total rate is plotted versus $o_b^{\max}$ for $K = 12$. As observed in Fig. 5.3, the total rate is an increasing function with respect to $o_b^{\max}$. Similar to the previous observation, the total rate in the perfect CSI

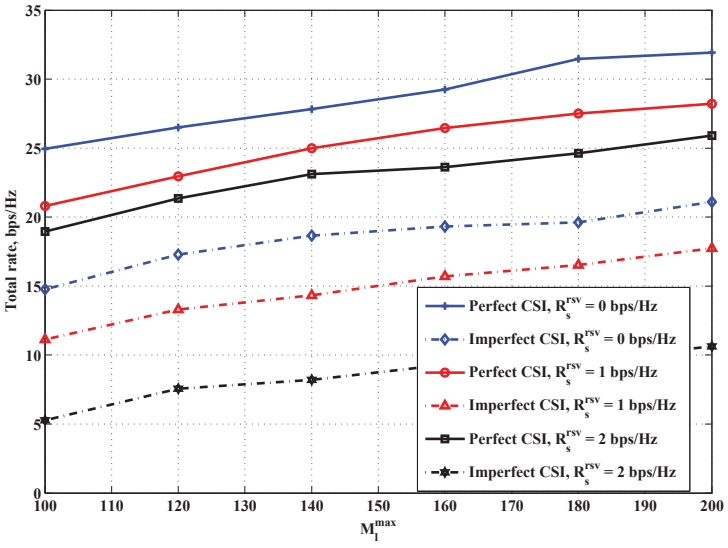


Fig. 5.4 Total rate versus $M_l^{\max}$

estimation is always higher than that of the imperfect CSI scenario which implies the importance of CSI estimation in the massive-MIMO aided C-RAN. For $o_b^{\max} \geq 6$ the total rate does not change even by increasing $o_b^{\max}$ since the system is able to support all the users at that value of $o_b^{\max}$ and any further increase does not have any effect on the system performance. To investigate the performance of Algorithm 5.1 with respect to the massive MIMO parameters, the total rate versus $M_l^{\max}$ is plotted in Fig. 5.4. From Fig. 5.4, the total rate increases with increasing $M_l^{\max}$ which is due to the spatial multiplexing gain obtained from the increase in the number of antennas. Again, the total rate with perfect CSI is higher that that with imperfect CSI estimation, and the total rate decreases with increasing R_s^{rsv} similar to Fig. 5.2.

To get more insight about the effect of pilot duration, in Fig. 5.5, the total rate versus τ/T is shown. From Fig. 5.5, the total rate keeps increasing as τ/T increases up to $\tau/T = 0.3$ and then decreases later. This is because at very low values of τ, the pilot contamination effect is more pronounced. However, as τ keeps increasing, there is less transmission time and more time is allocated in sending the pilot signal. Hence, a proper design of the pilot duration is essential in order to maintain high efficiency of the VWN and based on the simulation results, $\tau/T = 0.3$ is optimal for this setup.

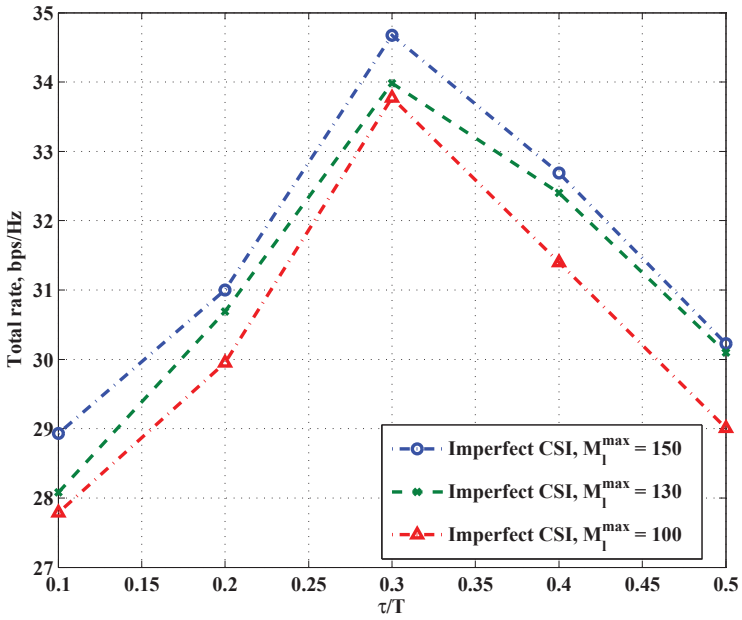


Fig. 5.5 Total rate versus τ/T for different $M_l^{\max}$

5.5 Concluding Remarks

In this chapter, we considered the resource allocation problem in the uplink of a C-RAN-based VWN with fronthaul limitations. Specifically, we formulated the resource allocation problem in a C-RAN VWN where the users are assigned RRHs as well as BBUs and allocated power and antennas with the aim of maximizing the total sum rate of the system. we proposed an iterative algorithm based on CGP and SCA that efficiently allocates these resources to users while maintaining the resource reservation per slices in the VWN. The performance of the proposed algorithm was compared for both perfect and imperfect CSI scenarios.

Chapter 6

Resource Allocation in a NOMA-based VWN

In this chapter, we consider a new multiple access technique and address the resource allocation problem in a NOMA-based VWN. Specifically, we formulate a power-efficient resource allocation problem in a NOMA-based VWN to find the optimal power allocation coefficients to users while maintaining the slice isolation requirements. By using various convexification approaches, we develop a computationally efficient algorithm that tries to minimize the total transmit power from the BS. The simulation results performed demonstrate that the proposed NOMA scheme significantly outperforms the OFDMA scheme in terms of the power efficiency.[1]

6.1 Introduction

Non-orthogonal multiple access (NOMA) has been recently introduced as an effective approach to increase spectrum efficiency and provide massive connectivity [24, 62]. Compared to the existing orthogonal multiple access techniques such as OFDMA, via NOMA, multiple users share the entire spectrum at the same time and frequency (and code), but with different allocated power levels. Since the users share the time and frequency resources, sophisticated techniques for decoding the superimposed signal need to be implemented at the receiver. By implementing successive interference cancellation (SIC), the receiver iteratively subtracts the strongest signal from the superimposed signal and decodes the intended signal [63]. In contrast, in OFDMA, the users are

[1] Part of this chapter is presented in [61].

allocated different sub-carriers, which effectively removes interference among users by exclusive sub-carrier allocation within a cell. The important question in this scenario is whether NOMA can improve the spectrum efficiency as compared to OFDMA.

There has been a significant research interest in this context. For instance, [64] compares the system level performance of the NOMA scheme with different mechanisms for power allocation including the user grouping based on their channel gains and equal power allocation to all users. The authors propose a sub-optimal power allocation scheme called fractional transmit power allocation (FTPA) that is similar to the transmission power control mechanism in LTE. Similarly, [65] analyzes the performance of NOMA compared to OFDMA for the cellular uplink setup. The optimization problem of this work includes the minimum required throughput of each user as a constraint. The performance of the system in the cell-edge has been shown to be significantly improved in the case of NOMA compared to OFDMA. Similarly, [66] proposes an enhanced proportional fairness scheme based on NOMA and shows the improvement of cell throughput by up to 28% compared to OFDMA scheme. In [27], a power allocation problem for the downlink transmission of NOMA system is formulated and solved by applying difference of convex functions (DC) programming. In order to develop the proposed algorithm, the greedy user selection approach is used to assign users to sub-carriers, and then, DC approximation is applied to allocate power for each user.

In this chapter, we investigate the use of NOMA in the VWNs to improve the network performance. The objective is to minimize the total transmit power in a VWN, while maintaining the minimum required throughput of each slice. Since the original problem is non-convex and computationally intractable, we use the approach of CGP and AGMA to convert it into an efficient algorithm. The simulation results demonstrate that NOMA is more power-efficient than OFDMA in various scenarios. Specifically, the power efficiency is improved by up to 45- 54 % with NOMA as compared to OFDMA.

6.2 System Model

Consider a VWN with a single BS that serves a set of slices (i.e., $\mathcal{S}$), in which each slice $s \in \mathcal{S}$ has its own set of users denoted by $\mathcal{K}_s$. The total number of users in the system is given by $K = \sum_{s \in \mathcal{S}} K_s$. To provide the isolation among slices, the VWN should preserve a minimum required rate per each slice s, denoted by R_s^{rsv}. We consider the following two transmission modes for the VWN:

- NOMA where the whole frequency band of interest is shared among users,
- OFDMA where the specific bandwidth is divided into a set of sub-carriers denoted by $\mathcal{N}$ and each sub-carrier can be allocated to a maximum of one user at a time.

In this chapter, our focus is to compare the power efficiency of these two approaches for our system model. We assume that the bandwidth W is divided into a set of sub-carriers $\mathcal{N} = \{1, \cdots N\}$, and the channel gain from the BS to the user k_s in slice s and in sub-carrier n is

$$h_{k_s,n} = \chi_{k_s,n} d_{k_s}^{-\zeta}, \tag{6.1}$$

where $\chi_{k_s,n}$ is the fading coefficient, $d_{k_s} \in [0.1, 1]$ is the distance of the user $k_s \in \mathcal{K}_s$ to the BS normalized to the cell radius and ζ is the path loss exponent.

6.2.1 NOMA

With NOMA, the transmitter encodes the information for each user spread on the entire bandwidth and transmits the superimposed signal. Hence, considering the transmission from the BS to K users, the received superimposed signal plus noise at the receiver of the user k_s in slice s at sub-carrier n is given by

$$y_{k_s,n} = h_{k_s,n} x + w_{k_s,n}, \tag{6.2}$$

where, $h_{k_s,n}$ is the complex channel gain from the BS to the user k_s which is assumed to be known to both the BS transmitter and the user k_s, x denotes the superimposed transmitted signal for all users and $w_{k_s,n}$ is the noise component for the user k_s at sub-carrier n. With the orthogonal scheme, the upper bound on the capacity of the user k_s can be achieved by allocating all the power and degrees of freedom available to user k_s with all the other users getting zero rate and is given by

$$R_{k_s,n} < \log(1 + \frac{\rho_{k_s,n} h_{k_s,n}}{\sigma^2}), \tag{6.3}$$

where $\rho_{k_s,n}$ is the power allocation coefficient for the user k_s at sub-carrier n and σ^2 is the noise power spectral density and is assumed to be equal for all users. Thus, considering the case of only two users, for simplicity, the capacity region consists of two extreme points where only one user gets the maximum rate achievable at any instant while the other user gets zero rate.

However, with SIC implemented at the receiver, the users iteratively subtract the signals from users lying below the current user in the decoding order. More specifically, in the two-user scenario, where $|h_{1,n}| \leq |h_{2,n}|$, user 1 treats the signal for user 2 at sub-carrier n as noise and can hence achieve a rate of

$$\begin{aligned} R_{1,n} &= \log(1 + \frac{\rho_{1,n} h_{1,n}}{\rho_{2,n}|h_{1,n}|^2 + \sigma^2}) \\ &= \log(1 + \frac{(\rho_{1,n} + \rho_{2,n}) h_{1,n}}{\sigma^2}) - \log(1 + \frac{\rho_{2,n} h_{1,n}}{\sigma^2}) \end{aligned} \tag{6.4}$$

Now, user 2 performs successive interference cancellation, it first decodes the signal for user 1, subtracts the determined signal from the superimposed signal and extracts its data. Thus user 2 can achieve a rate of

$$R_{2,n} = \log(1 + \frac{\rho_{2,n} h_{2,n}}{\sigma^2}) \tag{6.5}$$

In general, with the decoding order of $|h_{1,n}| \leq |h_{2,n}| \leq \cdots \leq |h_{K,n}|$, the user k_s, with index i, can successively remove the interference of all users with indices $j < i$ at sub-carrier n. For the rest of the users, i.e., users with indices $j > i$, the interference cannot be removed. Consequently, the received SINR at the user k_s, with index i at the sub-channel n, is given by

$$\gamma_{i,n}^{\text{NOMA}} = \frac{\rho_{i,n} h_{i,n}}{\sigma^2 + h_{i,n} \sum_{s \in \mathcal{S}} \sum_{j=i+1}^{K} \rho_{j,n}}. \tag{6.6}$$

Similarly, the rate of user k_s, with index i, at the sub-carrier n is

$$\begin{aligned} R_{k_s,n}^{\text{NOMA}} = R_{i,n} &= \log_2(1 + \gamma_{i,n}^{\text{NOMA}}) \\ &= \log_2 \left(1 + \frac{\rho_{i,n} h_{i,n}}{\sigma^2 + h_{i,n} \sum_{s \in \mathcal{S}} \sum_{j=i+1}^{K} \rho_{j,n}} \right) \end{aligned} \tag{6.7}$$

Each slice $s \in \mathcal{S}$ in the VWN has a minimum reserved rate of R_s^{rsv} in order to support the QoS requirement of the users, which can be expressed as

$$\text{C6.1:} \sum_{k_s \in \mathcal{K}_S} \sum_{n \in \mathcal{N}} R_{k_s,n}^{\text{NOMA}} \geq R_s^{\text{rsv}}, \quad \forall s \in \mathcal{S}.$$

6.2.2 OFDMA

We consider an OFDMA system where the total available frequency is divided into $n \in \mathcal{N}$ sub-carriers and if $\alpha_{k_s,n}$ is the sub-carrier allocation indicator for the sub-carrier n and user k_s in slice $s \in \mathcal{S}$, then

$$\alpha_{k_s,n} = \begin{cases} 1, & \text{if sub-carrier } n \text{ is allocated to user } k_s, \\ 0, & \text{otherwise.} \end{cases}$$

Due to the OFDMA exclusive sub-carrier assignment, we have a constraint on $\alpha_{k_s,n}$ as

$$\text{C6.2: } \sum_{\forall s} \sum_{\forall k_s} \alpha_{k_s,n} \leq 1, \quad \forall n \in \mathcal{N}.$$

The received SINR at the user k_s at sub-carrier $n \in \mathcal{N}$ and in slice $s \in \mathcal{S}$ is

$$\gamma_{k_s,n}^{\text{OFDMA}} = \frac{P_{k_s,n} h_{k_s,n}}{\sigma^2}, \tag{6.8}$$

Hence, the rate of user k_s at sub-carrier n is

$$R_{k_s,n}^{\text{OFDMA}} = \alpha_{k_s,n} \log_2(1 + \gamma_{k_s,n}^{\text{OFDMA}}). \tag{6.9}$$

In this case, the minimum reserved rate of each slice is represented as

$$\text{C6.3: } \sum_{k_s \in \mathcal{K}_S} \sum_{n \in \mathcal{N}} R_{k_s,n}^{\text{OFDMA}} \geq R_s^{\text{rsv}}, \quad \forall s \in \mathcal{S}.$$

Consider $\boldsymbol{\rho} = [\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_S]$ as the vector of power allocation coefficients of all users in all slices in NOMA, where $\boldsymbol{\rho}_s = [\boldsymbol{\rho}_{k_s}]_{k_s=1}^{K_s}$ and $\boldsymbol{\rho}_{k_s} = [\rho_{k_s,1}, \cdots, \rho_{k_s,N}]$, respectively. Similarly, for the OFDMA case, the power allocation vector of the system can be represented as $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_S]$, where $\mathbf{P}_s = [\mathbf{P}_{k_s}]_{k_s=1}^{K_s}$ and $\mathbf{P}_{k_s} = [P_{k_s,1}, \cdots, P_{k_s,N}]$. Also, the sub-carrier allocation vector of the system can be represented as $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_S]$, where $\boldsymbol{\alpha}_s = [\boldsymbol{\alpha}_{k_s}]_{k_s=1}^{K_s}$ and $\boldsymbol{\alpha}_{k_s} = [\alpha_{k_s,1}, \cdots, \alpha_{k_s,N}]$.

Now, for the case of NOMA, the optimization problem to minimize the total transmit power can be expressed as

$$\min_{\boldsymbol{\rho}} \sum\nolimits_{s \in \mathcal{S}} \sum\nolimits_{k_s \in \mathcal{K}_S} \sum\nolimits_{n \in \mathcal{N}} \rho_{k_s,n}, \tag{6.10}$$

$$\text{subject to : C6.1.}$$

For the case of OFDMA, the corresponding resource allocation problem is

$$\min_{\mathbf{P},\boldsymbol{\alpha}} \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_S} \sum\nolimits_{n\in\mathcal{N}} \alpha_{k_s,n} P_{k_s,n}, \tag{6.11}$$
$$\text{subject to : C6.2} - \text{C6.3}.$$

The proposed algorithm to solve the optimization problem is described in the subsequent section for both NOMA and OFDMA schemes.

6.3 Proposed Algorithm

The formulated optimization problems in (6.10) and (6.11) are non-convex and solving them is challenging. Besides, (6.11) involves binary integer variables. To develop an efficient algorithm to solve (6.11), we deploy an iterative framework of successive convex approximation, in which the non-convex function is transformed into a convex one in each iteration.

For this transformation, we apply the CGP and variable relaxation to convert (6.10) into the GP formulation. For (6.11), we apply the dual approach which has been widely utilized for solving OFDMA-based resource allocation problems [3, 67].

6.3.1 Iterative Algorithm for NOMA-based Resource Allocation

Let us first write $R_{k_s,n}^{\text{NOMA}} = \log_2(1+\gamma_{i,n}^{\text{NOMA}})$ as

$$\widetilde{R}_{k_s,n}^{\text{NOMA}} = \log_2\big(\frac{\sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\rho_{j,n} + \rho_{i,n}h_{i,n}}{\sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\rho_{j,n}}\big). \tag{6.12}$$

From the above, C6.1 can be rewritten as

$$\prod_{i\in\mathcal{K}_s}\prod_{n\in\mathcal{N}}\left(\frac{\sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\beta_{j,n}}{\sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\beta_{j,n} + \beta_{i,n}h_{i,n}}\right) \leq 2^{(-R_s^{\text{rsv}})}, \qquad \forall s\in\mathcal{S}.$$

To apply the CGP, consider t_1 as the iteration number. In each iteration t_1, the non-convex function should be approximated to its convex counterpart. Based on the structure of $\widetilde{R}_{k_s,n}^{\text{NOMA}}$, we

Algorithm 6.1 : Iterative Algorithm Based on CGP for NOMA-based VWN

Initialization: Set $t_1 = 1$, $\boldsymbol{\rho}(t_1) = [\mathbf{1}]$, where $\mathbf{1}$ is a vector $C^{1\times K}$.
Repeat:
Step 1: Update $s_{i,n}(t_1), g_{j,n}(t_1), r_{i,n}(t_1)$, and $z_{i,n}(t_1)$ from (6.14)-(6.15),
Step 2: Find optimal $\boldsymbol{\rho}^*(t_1)$ from (6.16) via CVX [41],
Until: $||\boldsymbol{\rho}^*(t_1) - \boldsymbol{\rho}^*(t_1 - 1)|| \leq \varepsilon_1$.

can apply AGMA approximation to propose the monomial approximation of $\widetilde{R}_{k_s,n}^{\text{NOMA}}$. At iteration t_1, $\widetilde{R}_{k_s,n}^{\text{NOMA}}$ can be approximated to the following convex function, for all i,

$$x_{i,n}(t_1) = (\sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\rho_{j,n})(\frac{\sigma^2}{s_{i,n}(t_1)})^{-s_{i,n}(t_1)}\times \prod_{\forall s,j=i+1}^{K}\left(\frac{h_{i,n}\rho_{j,n}(t_1)}{g_{j,n}(t_1)}\right)^{-g_{j,n}(t_1)}\left(\frac{\rho_{i,n}(t_1)h_{i,n}}{r_{i,n}(t_1)}\right)^{-r_{i,n}(t_1)}, \tag{6.13}$$

where for all i and $n \in \mathcal{N}$,

$$s_{i,n}(t_1) = \frac{\sigma^2}{z_{i,n}(t_1)}, \quad g_{j,n}(t_1) = \frac{\rho_{j,n}(t_1 - 1)h_{j,n}}{z_{i,n}(t_1)}, \quad r_{i,n}(t_1) = \frac{\rho_{i,n}(t_1 - 1)h_{i,n}}{z_{i,n}(t_1)}, \tag{6.14}$$

$$z_{i,n}(t_1) = \sigma^2 + h_{i,n}\sum_{s\in\mathcal{S}}\sum_{j=i+1}^{K}\rho_{j,n}(t_1 - 1) + h_{i,n}\rho_{i,n}(t_1 - 1). \tag{6.15}$$

Considering (6.13)-(6.15), the optimization problem (6.10) at iteration t_1 is approximated to the following convex optimization problem

$$\begin{aligned} &\min_{\boldsymbol{\rho}(t_1)} \sum_{i=1}^{K}\sum_{n=1}^{N}\rho_{i,n}(t_1) \\ &\text{subject to: (6.13) } - \text{ (6.15)}, \\ &\prod_{i\in\mathcal{K}_s}\prod_{n\in\mathcal{N}} x_{i,n}(t_1) \leq 2^{(-R_s^{\text{rsv}})}, \qquad \forall s \in \mathcal{S}. \end{aligned} \tag{6.16}$$

The overall iterative algorithm to solve (6.10) based on the convex function (6.16) is presented in Algorithm 6.1.

6.3.2 Dual Approach for OFDMA-based Resource Allocation

Since (6.11) involves binary variables $\boldsymbol{\alpha}$, we first relax $\alpha_{k_s,n} \in [0,1], \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}$. Now, by considering $y_{k_s,n} = \alpha_{k_s,n} P_{k_s,n}$, the total rate of OFDMA can be rewritten as [3], [67],

$$\widetilde{R}^{\text{OFDMA}}_{k_s,n}(\boldsymbol{\alpha}, \mathbf{y}) = \alpha_{k_s,n} \log_2(1 + \frac{y_{k_s,n} h_{k_s,n}}{\alpha_{k_s,n}\sigma^2}). \tag{6.17}$$

Note that the above expression belongs to a class of convex functions with the format of $f(a,b) = a\log(1+b/a)$ [42]. Therefore, C6.3 can be written as

$$\widetilde{\text{C}}6.3: \sum_{k_s \in \mathcal{K}_S} \sum_{n \in \mathcal{N}} \widetilde{R}^{\text{OFDMA}}_{k_s,n}(\boldsymbol{\alpha}, \mathbf{y}) \geq R^{\text{rsv}}_s, \quad \forall s \in \mathcal{S}. \tag{6.18}$$

Consequently, (6.11) can be written as

$$\begin{aligned} &\min_{\mathbf{y},\boldsymbol{\alpha}} \sum\nolimits_{s\in\mathcal{S}} \sum\nolimits_{k_s\in\mathcal{K}_S} \sum\nolimits_{n\in\mathcal{N}} y_{k_s,n}, \\ &\text{subject to: C}6.2, \widetilde{\text{C}}6.3. \end{aligned} \tag{6.19}$$

Proposition 1: Problem (6.19) is convex and can be solved using the Lagrange dual method. [3]

Proof: In (6.17), $\widetilde{R}_{k_s,n}(\boldsymbol{\alpha}, \mathbf{y})$ is of the form $f(a,b) = a\log(1+b/a)$ which is a convex function and can be solved by the Lagrangian method [42].

The corresponding Lagrange function for (6.19) is

$$\mathcal{L}(\phi_s, \nu_n, \mathbf{y}, \boldsymbol{\alpha}) = \sum_{\forall s, \forall k_s, \forall n} y_{k_s,n} + \sum_{\forall s} \phi_s (R^{\text{rsv}}_s - \sum_{\forall k_s, n} \widetilde{R}^{\text{OFDMA}}_{k_s,n}) + \sum_{\forall n} \nu_n (\sum_{s} \sum_{k_s} \alpha_{k_s,n} - 1), \tag{6.20}$$

where $\phi_s, \forall s \in \mathcal{S}$ and $\nu_n, \forall n \in \mathcal{N}$ are the Lagrange variables associated to $\widetilde{\text{C}}6.1$ and C6.2, respectively. Considering $\boldsymbol{\phi}$ and $\boldsymbol{\nu}$ as the vectors of the Lagrange variables for ϕ_s and $\nu_n, \forall s, \forall n$, respectively, the dual function for (6.20) is, [42]

$$\mathcal{D}(\boldsymbol{\phi}, \boldsymbol{\nu}) = \min_{\mathbf{y},\boldsymbol{\alpha}} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\nu}, \mathbf{y}, \boldsymbol{\alpha}). \tag{6.21}$$

Thus, the dual problem can be written as

$$\max_{\boldsymbol{\phi},\boldsymbol{\nu}} \mathcal{D}(\boldsymbol{\phi},\boldsymbol{\nu}) \quad \text{subject to: } \boldsymbol{\phi} > 0 \ \& \ \boldsymbol{\nu} > 0. \tag{6.22}$$

Since problem (6.19) is convex, the duality gap is zero and hence, the solution of the dual problem is equal to the solution of the primal problem [42]. Hence, by applying KKT conditions, the optimal power allocation for user k_s in slice s and sub-carrier n, i.e., $P^*_{k_s,n}$, is

$$P^*_{k_s,n} = \left[\frac{\phi_s}{\ln(2)} - \frac{\sigma^2}{h_{k_s,n}}\right]_0^{P_{\max}}, \tag{6.23}$$

where, $[x]_a^b = \max\{\min\{x,b\},a\}$. Also, the optimal sub-carrier allocation, $\alpha^*_{k_s,n}$, is

$$\alpha^*_{k_s,n} = \begin{cases} 0, & \frac{\partial(\mathcal{L}(\phi_s,\nu_n,\mathbf{y},\boldsymbol{\alpha}))}{\alpha^*_{k_s,n}} < 0 \\ \in [0,1], & \frac{\partial(\mathcal{L}(\phi_s,\nu_n,\mathbf{y},\boldsymbol{\alpha}))}{\alpha^*_{k_s,n}} = 0 \\ 1, & \frac{\partial(\mathcal{L}(\phi_s,\nu_n,\mathbf{y},\boldsymbol{\alpha}))}{\alpha^*_{k_s,n}} > 0, \end{cases} \tag{6.24}$$

where,

$$\frac{\partial(\mathcal{L}(\phi_s,\nu_n,\mathbf{y},\boldsymbol{\alpha}))}{\partial\alpha^*_{k_s,n}} = \nu_n - \phi_s\left(\log_2(1+\gamma_{k_s,n}) - \frac{\gamma_{k_s,n}}{(1+\gamma_{k_s,n})\ln(2)}\right), \quad \forall s \in \mathcal{S}$$

and $\gamma_{k_s,n} = \frac{y_{k_s,n}h_{k_s,n}}{\alpha_{k_s,n}\sigma}$. Now, from the KKT conditions, we have

$$\zeta_{k_s,n} = \phi_s\left[\log_2(1+\gamma_{k_s,n}) - \frac{\gamma_{k_s,n}}{(1+\gamma_{k_s,n})\ln(2)}\right], \forall s \in \mathcal{S}. \tag{6.25}$$

To satisfy the OFDMA exclusive sub-carrier allocation, $\alpha^*_{k_s,n}$ is chosen such that $\zeta_{k_s,n}$ is maximum [68], mathematically represented as

$$\alpha^*_{k_s,n} = \begin{cases} 1, & k'_s = \max_{\forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}} \frac{\partial(\mathcal{L}(\phi_s,\nu_n,\mathbf{y},\boldsymbol{\alpha}))}{\partial\alpha^*_{k_s,n}} \\ 0, & k_s \neq k'_s. \end{cases} \tag{6.26}$$

Algorithm 6.2 : Resource Allocation for OFDMA-based VWN

Initialization: Set $t_2 = 1$, $\boldsymbol{\alpha}(t_2) = [\mathbf{1}]$, where $\mathbf{1}$ is a vector $C^{1\times KN}$, $p_{k_s,n}(t_2) = 1, \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}, \forall n \in \mathcal{N}$, $t_2^{\max} = 5000$.
Repeat:
Step 1: Update $\phi_s(t_2+1) = \left[\phi_s(t_1) + \delta_{\phi_s}\frac{\partial \mathcal{L}}{\partial \phi_s}\right]^+, \forall s \in \mathcal{S}$.
Step 2: Repeat: Set inner loop iteration index as $t_3 = 1$.
Step 2a: Update $P^*_{k_s,n}(t_3), \forall k_s \in \mathcal{K}_s, \forall n \in \mathcal{N}$, from (6.23),
Step 2b: Find $\zeta_{k_s,n}(t_3)$ from (6.25) and set $\alpha_{k_s,n}(t_3) = 1$, if $\zeta_{k_s,n}(t_3) = \max[\zeta_{k_s,n}], \forall k_s \in \mathcal{K}_s, \forall s \in \mathcal{S}$,
Until $||\mathbf{P}(t_3) - \mathbf{P}(t_3 - 1)|| \ll \varepsilon_2$.
Until: $||\phi_s(t_2) - \phi_s(t_2 - 1)|| \leq \varepsilon_2$, or $t_2 > t_2^{\max}$.

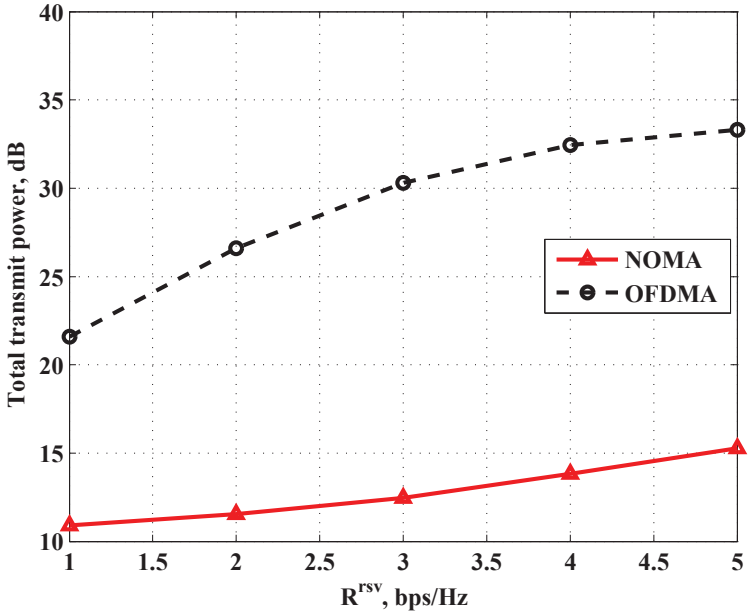


Fig. 6.1 Total transmit power versus R^{rsv}

The overall algorithm is described in Algorithm 6.2. □

To solve the convex problem (6.19), the iterative algorithm based on the dual function can be applied with a low computational complexity as demonstrated in [3], [67] which is summarized in Algorithm 6.2.

6.4 Numerical Results and Discussions

To study the performance of the proposed algorithm for NOMA and compare it with the OFDMA scheme, we simulate a scenario of a VWN with a single BS serving two slices each with $K_s = 8$ users, where $K = \sum_{s\in\mathcal{S}} K_s$ and $R^{\text{rsv}} = R_s^{\text{rsv}}$ for all $s \in \mathcal{S}$. The users are randomly located (from a uniform distribution) within the whole area of interest unless otherwise stated. The total number of sub-carriers is taken to be $N = 64$. The channel gains are derived according to the Rayleigh fading model. More specifically, $h_{k_s,n} = \chi_{k_s,n} d_{k_s}^{-\zeta}$ where $\zeta = 3$ is the path loss exponent, $d_{k_s} \in [0.1, 1]$ is the normalized distance between the BS and user k_s normalized to the cell radius, i.e., cell radius is 1, and χ_{k_s} is the exponential random variable with mean of 1. The results are taken over the average of 100 different channel realizations.

In Fig. 6.1, the total transmit power versus R^{rsv} is depicted for both NOMA and OFDMA schemes. From Fig. 6.1, it is clear that the total transmit power increases with increasing R^{rsv} for both cases. It is because the BS needs to transmit at a higher transmit power to satisfy the minimum reserved rate per slice. However, the total transmit power in the case of OFDMA is higher than that in the case of NOMA, indicating that NOMA is more power efficient that OFDMA. Specifically, the total transmit power has been decreased by almost 45% from 22 dB to almost 12 dB at $R^{\text{rsv}} = 1$ bps/Hz and by 54% from 33 dB to 15 dB at $R^{\text{rsv}} = 5$ bps/Hz, respectively, with NOMA as compared to OFDMA.

Fig. 6.2 plots the total transmit power versus K for different R^{rsv}. From Fig. 6.2 and as expected from the multi-user diversity gain [45], it can be observed that the total transmit power decreases with increasing K for a fixed R^{rsv}. Also, similar to Fig. 6.1, it is clear that the total transmit power is an increasing function of R^{rsv}, while it is higher in the case of OFDMA as compared to NOMA. In Fig. 6.3, we study the effect of NOMA with non-uniform user distribution over the VWN for two scenarios, where in the first scenario, users are located close to the cell-center, i.e., the normalized distance, $d_{k_s} \in [0.1, 0.7]$ and in the second scenario, the users are located close to the cell boundary, i.e. $d_{k_s} \in [0.8, 1]$. Fig. 6.3 shows the total transmit power versus R_s^{rsv} for both OFDMA and NOMA. Based on the results in Fig. 6.3, the total transmit power with OFDMA is more than in the case of NOMA for both scenarios. Also, with increasing R_s^{rsv}, the total transmit power sharply increases for OFDMA as compared to that for NOMA, e.g., for $R_s^{\text{rsv}} > 3$ bps/Hz. More importantly, via OFDMA, for the cell-edge scenario, there is no solution for the resource allocation problem for $R_s^{\text{rsv}} > 3$ bps/Hz. However NOMA can reach the feasible solution by increasing the transmit power. This indicates the effectiveness of NOMA in

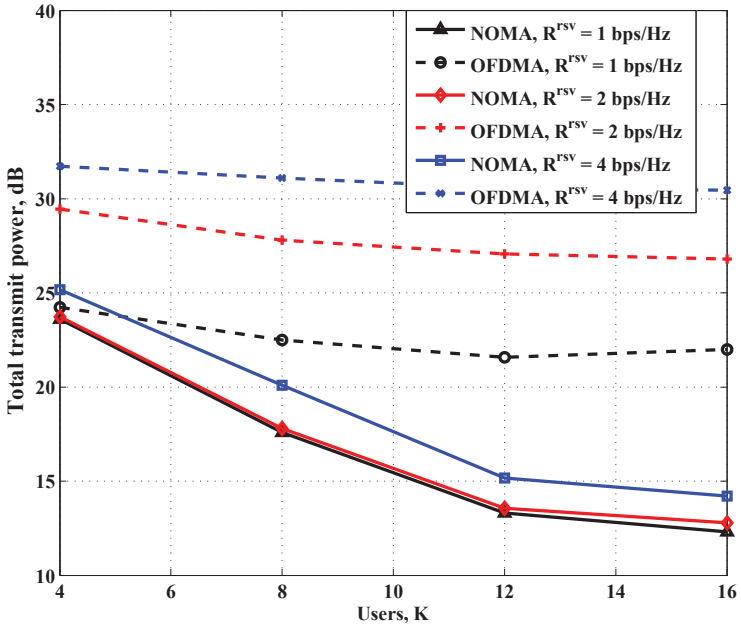


Fig. 6.2 Total transmit power versus K

achieving higher energy efficiency while preserving the isolation based constraints of the slices in the VWN.

6.5 Concluding Remarks

In this chapter, we investigated the power efficiency performance of NOMA compared to OFDMA for a VWN. In particular, we formulated an optimization problem with the objective to minimize the transmit power, while supporting the minimum reserved rate per each slice to ensure effective isolation among users in the slices. Since the resource allocation problem is non-convex and suffers from high computational complexity, we developed the CGP and AGMA approximation to propose the computationally tractable iterative algorithm. Via simulation results, we investigate the performance of the algorithm and compare it with the OFDMA scheme. Simulation results reveal that the proposed algorithm outperforms the OFDMA in terms of the required transmit power, specifically when most of users are located near the cell-edge and there is a diverse variation in the channel conditions.

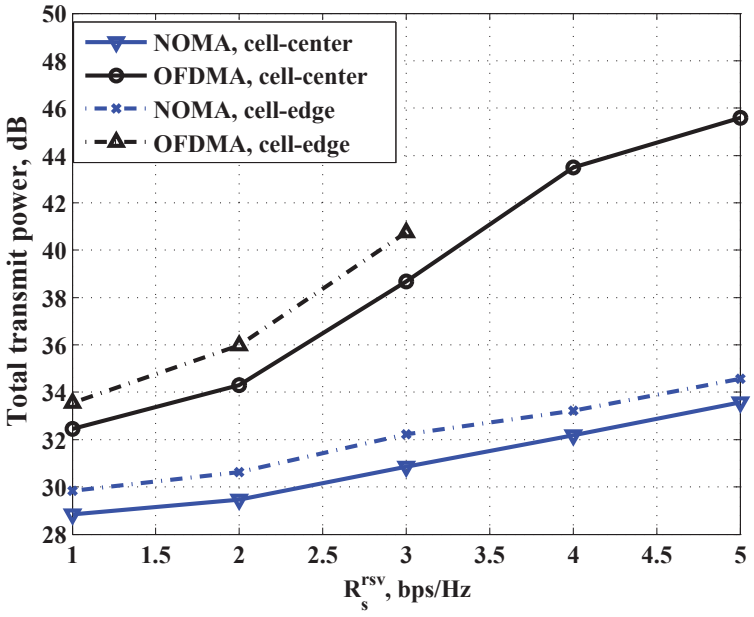


Fig. 6.3 Total transmit power versus R^{rsv} for cell-edge and cell-center users

Chapter 7

Conclusion

7.1 Summary

In this thesis, the issue of user-association and resource-allocation in a multi-cell VWN was addressed. We considered various aspects, including massive MIMO, C-RAN, NOMA in studying the resource allocation problems in both the single-cell and multi-cell VWNs. By applying various optimization techniques and approximations, computationally efficient algorithms were developed to allocate resources to users in slices under various scenarios while preserving the slice isolation constraints of the VWN.

In Chapter 2, we considered a multi-cell OFDMA VWN and formulated a resource allocation problem to jointly associate users to BSs and allocate sub-carriers and power with the aim of maximizing the total sum-rate of the system. The original non-convex problem was converted into the GP form using CGP and SCA was applied to propose a computationally efficient algorithm. The performance of the proposed algorithm was compared with the conventional approach of associating users to BS based on the maximum received SINR. Numerical results performed demonstrated the improvement in spectral efficiency of the VWN with the proposed scheme.

Chapter 3 addressed the power-efficiency issue in a multi-cell VWN. Specifically, we extended the joint user-association and resource-allocation problem to minimize the total transmit power of the VWN subject to the isolation constraints of the slices. Simulation results demonstrated that the proposed algorithm significantly outperforms the conventional approach of associating users based on the maximum received SINR in terms of the power efficiency of the VWN.

In Chapter 4, we addressed the resource allocation problem in a massive MIMO VWN.

Specifically, we extended the resource allocation problem in a VWN with BSs using massive MIMO and considered adaptive pilot duration as an integral part of the resource allocation problem to maximize the total sum-rate of the system subject to the constraints on the VWN as well as the pilot contamination effects. Simulation results have demonstrated that the application of massive MIMO at the BS significantly extends the feasibility region of the VWN and the performance is improved with the adaptive pilot duration as compared to the fixed pilot duration.

The resource allocation problem in a C-RAN-based VWN was addressed in Chapter 5 where we considered the joint user-association and resource-allocation problem in a multi-cell VWN by applying the C-RAN architecture. Specifically, we formulated the resource allocation problem to jointly associate users to RRHs/BBUs and allocate power and antennas to users. In order to maintain the slice isolation requirement, we considered both minimum reserved rate and number of antennas per slice in this setup. With the aim of maximizing the total sum rate of the system, an iterative algorithm based on SCA and CGP was proposed. The simulation results have been performed to compare the performance of the proposed algorithm in perfect and imperfect CSI scenarios and demonstrated the effect of pilot contamination on the system performance.

Chapter 6 considered the power allocation problem in a VWN using NOMA where users are assigned the same time-frequency resources but different power allocations while preserving the isolation constraints of the slices. With the aim of minimizing the total transmit power of the VWN, we proposed a power-efficient algorithm and compared the performance of the VWN against the OFDMA scheme. The simulation results have demonstrated that NOMA is significantly more power efficient than OFDMA in the VWN.

7.2 Potential extensions for future work

As VWN is still an evolving architecture and the requirements of the next generation network keep changing, there are a number of possible extensions based on the works in this thesis. Some of these are listed below:

- We considered resource allocation problems in a VWN in this thesis with the isolation constraints mainly based on the reserved rate and number of antennas per slice, a possible extension would be to consider the provision of admission control in the slices since due to the dynamic channel conditions in the wireless environment, there may arise infeasibility scenarios in the VWN when the constraints cannot be satisfied.

- We considered power-efficient resource allocation problems in Chapters 2 and 6 for OFDMA and NOMA, respectively, where the objective was to minimize the total transmit power from the BS. Further studies can consider an energy-efficient scheme by taking into account the circuit power in the BS as well by using various power consumption models in the BS.

- We considered the adaptive pilot duration allocation problem in a single cell VWN in Chapter 3. Further studies can consider a multi-cell VWN where multiple BSs will serve users in different slices in a location.

- Although the provision of minimum resource reservations per slice ensures that QoS of the users being served is satisfied, due to the dynamic channel conditions, the resource allocation problem might turn out to be infeasible. Future extensions could consider the admission control policy that could limit the number of users and/or slices in the network so that the QoS of users currently served is guaranteed.

- Further studies could consider various other interactions between the slices and the network operator, such as pricing mechanisms and lease policies in the VWN.

- The resource allocation in NOMA VWN could be extended to consider a multi-cell scenario with another sub-algorithm to consider the user association when the coverage of a certain area involves multiple BSs.

- Although the resource allocation problem in the C-RAN VWN is highly non-convex and involves a lot of variables, it could be extended to cover the OFDMA scheme in the multi-cell scenario by relaxing some constraints or by assuming fixed number of antennas and/or power per user.

Appendix: Brief Notes on Geometric Programming and Successive Convex Approximation

A.1 Introduction to Geometric Programming

A.1.1 Monomial and Posynomial function

Let $x = (x_1, \ldots, x_n)$ be a vector with components x_i where the components, x_i, are real positive variables, a monomial function $g(x)$ is a function of the form

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

where $c > 0$ and a_i are real numbers.

A posynomial is a function of the form

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \ldots x_n^{a_{nk}} \tag{A.1.2}$$

where, $c_k > 0$. from the definition above, a posynomial is a sum of one or more monomials.

A.1.2 Arithmetic-Geometric Mean Approximation

The arithmetic-geometric mean approximation (AGMA) is widely used in solving optimization problems by converting the summation of monomials into the approximated product forms. Specifically, if $f(\boldsymbol{x})$ is posynomial function, i.e. $f(\boldsymbol{x}) = \sum_k g_k(\boldsymbol{x})$ where $g_k(\boldsymbol{x})$ are monomials,

then, by AGMA, we have

$$f(\boldsymbol{x}) \geq \hat{f}(\boldsymbol{x}) = \prod_k \left(\frac{g_k(\boldsymbol{x})}{\alpha_k(\boldsymbol{x}_0)} \right)^{\alpha_k(\boldsymbol{x}_0)}, \tag{A.1.3}$$

where the parameters $\alpha_k(\boldsymbol{x}_0)$ can be obtained by computing

$$\alpha_k(\boldsymbol{x}_0) = \frac{g_k(\boldsymbol{x}_0)}{f(\boldsymbol{x}_0)}, \forall k. \tag{A.1.4}$$

In the above approximation, $\boldsymbol{x}_0$ is a fixed point with $\boldsymbol{x}_0 > 0$ and is the optimal solution obtained from the last iteration of the optimization. It has been proved that $\hat{f}(\boldsymbol{x})$ is the best local monomial approximation of $f(\boldsymbol{x})$ near $\boldsymbol{x}_0$ [42].

A.1.3 Geometric Programming

Geometric programming (GP) is a class of non-linear optimization problems, which can be solved very efficiently via numerical methods [38]. Various resource allocation problems have been solved by converting them into GP problems to reach computationally tractable algorithms, e.g., [37, 38, 69–71]. A geometric programming problem is an optimization problem of the form

$$\text{minimize } f_0(x) \tag{A.1.5}$$

$$\text{subject to } f_i(x) \leq 1, \quad i = 1, \ldots, m, \tag{A.1.6}$$

$$g_j(x) = 1, \quad j = 1, \ldots, p, \tag{A.1.7}$$

where $f_i(x)$ are posynomial functions, and $g_j(x)$ are monomial functions and $x_i > 0$ are the optimization variables. The above equation (A.1.5) describes the GP in a standard form where the inequality constraints are posynomials and the equality constraints are monomials.

A.1.4 Complementary Geometric Programming

Although GP has been extensively studied in the literature and there are a number of efficient tools to solve GP problems, the optimization problems in practical scenarios in communication systems are not typically in the standard GP form. There are a lot of restrictions, for example, on the equality and inequality constraints, which cannot be met for many practical problems related to the resource allocation of wireless networks. For example, in some cases, the equality con-

straints contain posynomial functions, inequality constraints present lower bound of posynomial function or the posynomial functions contain negative coefficients. Depending on the nature of the optimization problem, these types of problems belong to either one of classes of optimization problems such as generalized GP, signomial programming or complementary geometric programming (CGP).

A CGP problem can be presented as

$$\begin{aligned} &\min_{\mathbf{x}} F_0(\mathbf{x}), \\ \text{subject to:} \quad &F_i(\mathbf{x}) \leq 1, \quad i = 1, .., I, \\ &G_j(\mathbf{x}) = 1, \quad j = 1, .., J, \end{aligned} \tag{A.1.8}$$

where $F_0(\mathbf{x}) = f_0^+(\mathbf{x}) - f_0^-(\mathbf{x})$, $F_i(\mathbf{x}) = \frac{f_i^+(\mathbf{x})}{f_i^-(\mathbf{x})}$, $i = 1, \cdots, I$ and $G_j(\mathbf{x}) = \frac{g_j(\mathbf{x})}{f_j(\mathbf{x})}$ in which $f_0^+(\mathbf{x})$, $f_0^-(\mathbf{x})$, $j = 0, 1, \cdots, J$, are posynomial functions, while $g_j(\mathbf{x})$ and $f_j(\mathbf{x})$ are monomial and posynomial functions [43], respectively.

One approach to solve (A.1.8) is to convert it into a sequence of standard GP problems [39] that can be solved to reach a global solution. In other words, successive convex approximation (SCA) can be applied, where the non-convex optimization problem is approximated as a convex problem in each iteration. Specifically, AGMA can be applied to transform the non-posynomial functions to posynomial form, i.e., $F_i(\mathbf{x})$, and $G_j(\mathbf{x})$ to their monomial functions, respectively.

Using AGMA, at the iteration l, the approximated forms of $f_i^-(\mathbf{x}) = \sum_{k=1}^{K^{i^-}} g_k^{i^-}(\mathbf{x})$ and $f_j(\mathbf{x}) = \sum_{k=1}^{K_j} g_k^j(\mathbf{x})$ are

$$\widetilde{f}_i^-(\mathbf{x}(l)) = \prod_{k=1}^{K_{i^-}} \left(\frac{g_k^{i^-}(\mathbf{x}(l))}{\xi_k^{i^-}(l)} \right)^{\xi_k^{i^-}(l)}, \tag{A.1.9}$$

$$\text{and, } \widetilde{f}_j(\mathbf{x}(l)) = \prod_{k=1}^{K_j} \left(\frac{g_k^j(\mathbf{x}(l))}{\zeta_k^j(l)} \right)^{\zeta_k^j(l)}, \tag{A.1.10}$$

where $\xi_k^{i^-}(l) = \frac{g_k^{i^-}(\mathbf{x}(l-1))}{f_i^-(\mathbf{x}(l-1))}$ and $\zeta_k^j(l) = \frac{g_k^j(\mathbf{x}(l-1))}{f_j(\mathbf{x}(l-1))}$. Subsequently, $\widetilde{F}_i(\mathbf{x}(l)) = \frac{f_i^+(\mathbf{x}(l))}{\widetilde{f}_i^-(\mathbf{x}(l))(\mathbf{x}(l))}$ and $\widetilde{G}_j(\mathbf{x}(t)) = \frac{g_j(\mathbf{x}(t))}{\widetilde{f}_j(\mathbf{x}(t))}$ are posynomial and monomial functions, respectively [39], and the optimization problem related to each iteration l of (A.1.8) becomes

$$\min_{\mathbf{x}(l)} \quad \Xi + f_0^+(\mathbf{x}(l)) - f_0^-(\mathbf{x}(l)), \tag{A.1.11}$$
$$\text{subject to:}$$
$$\widetilde{F}_i(\mathbf{x}(l)) \leq 1, \widetilde{G}_j(\mathbf{x}(l)) = 1, \quad i = 1, .., I, j = 1, .., J,$$

where $\Xi \gg 1$ is a sufficiently large constant added to the objective function in (A.1.11) to keep it always positive [39]. However, the objective function of (A.1.11) still cannot satisfy the posynomial condition of (A.1.5). To reach the GP-based formulation for each iteration, we introduce the auxiliary variable $x_0 > 0$ for a linear objective function and use it to transform (A.1.11) into

$$\min_{\mathbf{x}_0(t)} x_0(l), \tag{A.1.12}$$
$$\text{subject to: } \frac{\Xi + f_0^+(\mathbf{x}(l))}{f_0^-(\mathbf{x}(l)) + x_0} \leq 1,$$
$$\widetilde{F}_i(\mathbf{x}(t)) \leq 1, \widetilde{G}_j(\mathbf{x}(t)) = 1, \tag{A.1.13}$$
$$i = 0, 1, \ldots, I, j = 1, \ldots, J,$$

where $\mathbf{x}_0(t) = [x_0(l), x_n(l), \cdots, x_0(l)]$. Similar to $F_i(\mathbf{x})$, term $\frac{\Xi + f_0^+(\mathbf{x}(l))}{f_0^-(\mathbf{x}(l)) + x_0}$ can be converted into posynomial function via AGMA, and finally, the resulting optimization problem has a GP-based structure and can be solved by efficient numerical algorithms, [39].

It has been shown that the solution obtained by the iterative algorithm based on the GP-based approximation of problem (A.1.8) can offer a performance very close to that of the optimal solution [39].

A.2 Multi-block Successive Convex Approximation

Successive convex approximation (SCA) is an iterative approach to solve general non-convex optimization problems where at each iteration of the algorithm, the non-convex problem is approximated by its corresponding convex form by considering a set of variables while fixing others. In other words, at each iteration, a locally tight approximation of the original optimization problem is solved subject to the tight convex restriction of the constraint sets.

Consider an optimization problem of the form

$$\min \quad f(x) \tag{A.2.1}$$

$$\text{s. t.} \quad x \in \mathcal{X}, \tag{A.2.2}$$

where the feasible set $\mathcal{X}$ is the Cartesian product of n closed convex sets: $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ where $\mathcal{X}_i \subseteq \mathbb{R}^{m_i}$ and $\sum_i m_i = m$. Here, the variables, $x \in \mathbb{R}^m$, in the original optimization problem can be decomposed into subsets as $x = (x_1, x_2, \ldots, x_n)$ where $x_i \in \mathcal{X}_i, i = 1, \ldots, n$.

In the multi-block SCA approach, the algorithm updates a single block of variables in each iteration, i.e. at iteration t, if the selected block is block i, then the global upper-bound approximation of the original objective function $f(.)$ is formed at the point x^{t-1}. Specifically, if $u_i(x_i, x^{t-1})$ is the convex approximation of the original objective function $f(x)$ at the point x^{t-1}, then, at the iteration t, the following sub-problem is solved,

$$\min_{x_i} \quad u_i(x_i, x^{t-1}) \tag{A.2.3}$$

$$\text{s. t.} \quad x_i \in \mathcal{X}_i. \tag{A.2.4}$$

The convergence and the complexity analysis of the SCA-based approach has been studied in various works, for instance [46]. Moreover, with the arithmetic-geometric mean approximation, the SCA approach converges to a locally optimal solution that satisfies the KKT conditions of the original problem [47].

References

[1] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: A substrate for virtualizing wireless resources in cellular networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, Oct. 2012.

[2] H. Wen, P. K. Tiwary, and T. Le-Ngoc, "Current trends and perspectives in wireless virtualization," in *Intl. Conf. on Sel. Topics in Mobile and Wireless Netw. (MoWNeT)*, Aug 2013, pp. 62–67.

[3] S. Parsaeefard, V. Jumba, M. Derakhshani, and T. Le-Ngoc, "Joint Resource Provisioning and Admission Control in Wireless Virtualized Networks," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 2020 – 2025.

[4] C. Liang and F. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 99, pp. 358 – 380, Aug. 2014.

[5] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 3, no. 99, pp. 67–79, May 2016.

[6] V. N. Ha, L. B. Le, and N. D. Dao, "Cooperative transmission in C-RAN considering fronthaul capacity and cloud processing constraints," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014.

[7] C. Liu, K. Sundaresan, M. Jiang, S. Rangarajan, and G.-K. Chang, "The case for reconfigurable backhaul in C-RAN based small cell networks," in *IEEE Intl. Conf. on Computer Commun. (INFOCOM)*, Apr. 2013, pp. 1124–1132.

[8] I. Guvenc, "Capacity and fairness analysis of heterogeneous networks with range expansion and interference coordination," *IEEE Wireless Commun. Lett.*, vol. 15, no. 10, pp. 1084–1087, Oct. 2011.

[9] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Outage probability for heterogeneous cellular networks with biased cell association," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.

[10] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *IEEE Intl. Conf. Commun. (ICC)*, June 2012, pp. 2457–2461.

[11] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[12] X. Tang, P. Ren, Y. Wang, Q. Du, and L. Sun, "User association as a stochastic game for enhanced performance in heterogeneous networks," in *IEEE Intl. Conf. Commun. (ICC)*, June 2015, pp. 3417–3422.

[13] M. Fallgren, "An optimization approach to joint cell, channel and power allocation in multicell relay networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2868–2875, 2012.

[14] Y. H. Cho, H. Kim, S.-H. Lee, and H. S. Lee, "A QoE-aware proportional fair resource allocation for Multi-Cell OFDMA Networks," *IEEE Wireless Commun. Lett.*, no. 19, pp. 82–85, 2015.

[15] N. Forouzan and S. A. Ghorashi, "New algorithm for joint subchannel and power allocation in multi-cell OFDMA-based cognitive radio networks," *IET Commun.*, vol. 8, no. 4, pp. 508–515, Mar. 2014.

[16] N. Ksairi, P. Bianchi, P. Ciblat, and W. Hachem, "Resource allocation for downlink cellular OFDMA systems - part I: Optimal allocation," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 720–734, 2010.

[17] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 61, no. 4, pp. 1–5, Apr. 2013.

[18] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics in Signal Processing*, vol. 8, pp. 742 – 758, Oct. 2014.

[19] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, pp. 605 – 608, 2014.

[20] J.-C. Shen, J. Zhang, and K. B. Letaief, "User capacity of pilot contaminated TDD massive MIMO systems," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 3713 – 3718.

[21] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Commun.*, vol. 15, no. 3, pp. 1835–1850, Mar. 2016.

[22] L. Chen, H. Jin, H. Li, J. B. Seo, Q. Guo, and V. Leung, "An energy efficient implementation of C-RAN in HetNet," in *IEEE Veh. Tech. Conf. (VTC)*, Sep. 2014, pp. 1–5.

[23] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.

[24] NTT Docomo Inc., "5G radio access: Requirements, concept and technologies," July 2014.

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

[26] S. Liu, C. Zhang, and G. Lyu, "User selection and power schedule for downlink non-orthogonal multiple access (NOMA) system," in *IEEE Intl. Conf. Commun. (ICC)*, June 2015, pp. 2561–2565.

[27] P. Parida and S. S. Das, "Power allocation in OFDM-based NOMA systems: A DC programming approach," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1026–1031.

[28] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, "LTE wireless virtualization and spectrum management," in *Intl. Conf. on Sel. Topics in Mobile and Wireless Netw. (MoWNeT)*, Oct. 2010, pp. 1 – 6.

[29] K. Tan, H. Shen, J. Zhang, and Y. Zhang, "Enabling flexible spectrum access with spectrum virtualization," in *IEEE Intl. Symp. on New Frontiers in Dynamic Spectrum Access Netw. (DySPAN)*, Oct. 2012, pp. 47–58.

[30] F. Fu and U. Kozat, "Stochastic game for wireless network virtualization," *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 84–97, Feb. 2013.

[31] M. Yang, Y. Li, D. Jin, J. Yuan, I. You, and L. Zeng, "Opportunistic sharing scheme for spectrum allocation in Wireless Virtualization," in *Intl. Conf. on Innovative Mobile and Internet Services in Ubiq. Comp. (IMIS)*, vol. 18, no. 9, July 2014, pp. 1685 – 1696.

[32] G. Liu, F. Yu, H. Ji, and V. Leung, "Distributed resource allocation in full-duplex relaying networks with wireless virtualization," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4959–4964.

[33] X. Zhang, Y. Li, D. Jin, L. Su, L. Zeng, and P. Hui, "Efficient resource allocation for wireless virtualization using time-space division," in *IEEE Intl. Conf. on Wireless Commun. and Mobile Comp. (IWCMC)*, Aug. 2012, pp. 59–64.

[34] 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, Apr. 2016.

[35] V. Jumba, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Resource provisioning in wireless virtualized networks via massive-MIMO," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 237 – 240, Feb. 2015.

[36] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[37] M. Chiang, "Geometric programming for communication systems," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1-2, pp. 1–154, 2005.

[38] 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, July 2007.

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

[40] M. Derakhshani, X. Wang, T. Le-Ngoc, and A. Leon-Garcia, "Airtime usage control in virtualized multi-cell 802.11 networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, 2014.

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

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

[44] B. Marks and G. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Journal of Operations Research*, vol. 26, no. 8, pp. 681–683, 1978.

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

[46] M. Razaviyayn, "Successive convex approximation: Analysis and Applications," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2014.

[47] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 342–355, Jan. 2014.

[48] R. Dawadi, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Energy-efficient resource allocation in multi-cell virtualized wireless networks," in *IEEE Intl. Conf. on Ubiquitous Wireless Broadband (ICUWB)*, Oct 2015, pp. 1–5.

[49] J. Li and N. Liu, "User association for minimum energy consumption with macro-relay interference," in *IEEE Intl. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Aug 2015, pp. 1547–1552.

[50] K. Han, D. Liu, Y. Chen, and K. K. Chai, "Energy-efficient user association in HetNets: An evolutionary game approach," in *IEEE Intl. Conf. on Big Data and Cloud Computing (BdCloud)*, Dec. 2014, pp. 648–653.

[51] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. Elkashlan, "Distributed energy efficient fair user association in massive mimo enabled hetnets," *IEEE Wireless Commun. Lett.*, vol. 19, no. 10, pp. 1770–1773, Oct. 2015.

[52] H. Tian, Y. Xu, K. Xu, J. Jing, and K. Wu, "Energy-efficient user association in heterogeneous networks with m2m/h2h coexistence under qos guarantees," *China Communications*, vol. 12, no. 7, pp. 93–103, Dec. 2015.

[53] R. Dawadi, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Adaptive pilot-duration and resource allocation in virtualized wireless networks with massive MIMO," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, vol. 61, no. 4, Apr. 2016.

[54] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of Base Station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, Jun. 2012.

[55] China Mobile Research Institute, "C-RAN: The road towards green RAN," July 2013.

[56] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[57] R. Wang, H. Hu, and X. Yang, "Potentials and Challenges of C-RAN supporting Multi-RATs toward 5G Mobile Networks," *IEEE Access*, vol. 2, pp. 1187–1195, Oct. 2014.

[58] L. Liu, S. Bi, and R. Zhang, "Joint power control and fronthaul rate allocation for throughput maximization in OFDMA-based cloud RAN," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4097–4110, Nov. 2015.

[59] A. Abdelnasser and E. Hossain, "On resource allocation for downlink power minimization in OFDMA small cells in a cloud-RAN," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[60] A. Checko, H. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. Berger, and L. Dittmann, "Cloud RAN for mobile networks- a technology overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, First quarter 2015.

[61] R. Dawadi, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Power-efficient resource allocation in NOMA VWN," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2016.

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

[63] S. Sen, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Successive interference cancellation: A back-of-the-envelope perspective," in *ACM SIGCOMM Workshop on Hot Topics in Networks*, Oct. 2010, pp. 171–176.

[64] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 66–70.

[65] T. Takeda and K. Higuchi, "Enhanced user fairness using non-orthogonal access with SIC in cellular uplink," in *IEEE Veh. Tech. Conf. (VTC)*, Sep. 2011, pp. 1–5.

[66] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," in *IEEE Veh. Tech. Conf. (VTC)*, May 2014, pp. 1–5.

[67] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," July 2006, pp. 1394–1398.

[68] N. Hassan and M. Assaad, "Resource allocation in multiuser OFDMA system: Feasibility and optimization study," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2009, pp. 1–6.

[69] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization and Engineering*, vol. 8, no. 1, p. 67–127, 2007.

[70] M. Chiang, D. Oneill, D. Julian, and S. Boyd, "Resource allocation for QoS provisioning in wireless ad hoc networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, vol. 5, Nov. 2001, pp. 2911–2915.

[71] M. Chiang and A. Sutivong, "Efficient optimization of constrained nonlinear resource allocation," in *IEEE Global Commun. Conf. (GLOBECOM)*, vol. 7, Dec. 2003, pp. 67–74.